THE CAMBRIDGE HISTORY OF

SCIENCE

VOLUME I
ANCIENT SCIENCE

EDITED BY
ALEXANDER JONES
AND LIBA TAUB

CONTENTS

ILLUSTRATIONS

NOTES ON CONTRIBUTORS

Andrew Barker is Emeritus Professor of Classics at the University of Birmingham and a Fellow of the British Academy. He has published 8 books and over 100 articles on ancient Greek music and musical theory.

Karine Chemla, Researcher (CNRS, laboratory SPHERE) has edited: *The History of Mathematical Proof in Ancient Traditions* (2012); *Texts, Textual Acts and the History of Science* (with Virbel, 2015); *The Oxford Handbook of Generality in Mathematics and the Sciences* (with Chorlay and Rabouin, 2016); and *Cultures without Culturalism* (with Fox Keller, 2017).

Christopher Cullen read Engineering Science at Oxford, and later did a PhD in Classical Chinese at SOAS, University of London, under the supervision of D. C. Lau. He has published widely on the history of science in China, and served as editor of the *Science and Civilisation in China* series. He is Emeritus Director of the Needham Research Institute and an Emeritus Fellow of Darwin College, Cambridge.

Serafina Cuomo is Professor of Ancient History at Durham University. She works on the history of science and technology in ancient Greece and Rome, and has published on ancient mathematics, land-surveying, and military technology. She is completing a book on ancient numeracy.

Andrea Falcon is Associate Professor in the Department of Philosophy at Concordia University, Montreal. He is the author of *Corpi e Movimenti. La fortuna del "De caelo" nel mondo antico* (2001); *Aristotle and the Science of Nature: Unity without Uniformity* (Cambridge 2005); *Aristotelianism in the First Century BCE: Xenarchus of Seleucia* (Cambridge 2012); and *Aristotelismo* (2017). He is the editor of *Brill's Companion to the Reception of Aristotle* (2016).

Markham J. Geller is Professor für Wissensgeschichte at the Freie Universität Berlin and Topoi Excellence Cluster since 2010, on secondment from University College London. The PI of the ERC Advanced Grant Project, *BabMed* (2013–2018), he is also co-PI with Philip van der Eijk on a multi-year project, *The Transfer of Medical Episteme in the "Encyclopaedic" Compilations of Late Antiquity,* within the Collaborative Research Center *Episteme in Bewegung* at the Freie Universität Berlin. He has worked extensively on Sumerian, Akkadian, and Aramaic magic and medicine.

Klaus Geus is an ancient historian, philologist, and geographer. In 2009 he was appointed professor of the historical geography of the ancient Mediterranean at the Freie Universität Berlin. Geus works in the areas of ancient geography and astronomy. He has published 20 books and 250 papers and articles.

Daniel W. Graham is Professor of Philosophy at Brigham Young University. He is president of the International Association for Presocratic Studies. His books include *Explaining the Cosmos* (2006); (with co-editor Patricia Curd) *The Oxford Handbook of Presocratic Philosophy* (2008); *The Texts of Early Greek Philosophy* (2010); and *Science before Socrates* (2013).

Jens Høyrup was educated as a physicist at Copenhagen University. From 1973 he taught first in the domain of social, then human sciences, at Roskilde University, Denmark, until he retired in 2005. Much of his research has dealt with the conceptual, cultural, and social history of pre-modern mathematics.

Annette Imhausen studied mathematics, chemistry, and Egyptology at Mainz University and the Freie Universität Berlin. She completed her doctoral dissertation in the history of mathematics and held fellowships at the Dibner-Institute (Cambridge, MA) and Trinity Hall (Cambridge, England). Since 2009 she has been Professor for the History of Early Science at Frankfurt University.

Alexander Jones is Leon Levy Director and Professor of the History of the Exact Sciences in Antiquity at the Institute for the Study of the Ancient World, New York University. His research interests include contacts between Babylonian and Greco-Roman astronomy and astrology, texts and artifacts of Hellenistic and Roman period astronomy, and the scientific work of Claudius Ptolemy.

Rolf Krauss studied Egyptology, history, and geography at the Freie Universität Berlin, earning a doctorate in 1981 with his dissertation "Problems of the Egyptian Calendar." In 1996 his "Astronomical Concepts in the Pyramid Texts" was accepted at Hamburg University for German Habilitation (2nd dissertation). From 1982 until 2007 he was on the academic staff of the State Museums in Berlin.

James G. Lennox is Professor Emeritus of History and Philosophy of Science at the University of Pittsburgh. His research in history and philosophy of biology includes a translation with commentary of Aristotle's *On the Parts of Animals* (Oxford 2002), *Aristotle's Philosophy of Biology* (Cambridge 2001), and many articles and book chapters on Aristotle and Charles Darwin.

Eric Lewis is an associate professor of Philosophy at McGill University. He has published in ancient philosophy, and the philosophy of music and new media art.

Vivienne Lo 羅維前 is the Director of the UCL China Centre for Health and Humanity. Her research focuses on the social and cultural origins of self-care. She translates and analyzes manuscript material from early and medieval China and studies the transmission of scientific knowledge along the so-called Silk Roads through to the modern Chinese medical diaspora.

Philipp A. Maas is an Indologist who currently works at the University of Leipzig, Germany. His main interests are the pre-modern cultural and intellectual history of South Asia and the multiple genres of Sanskrit literature. Maas has published widely, especially on early classical Ayurveda and Yoga.

Clemency Montelle is Associate Professor in the School of Mathematics and Statistics at the University of Canterbury, Christchurch, New Zealand. She has research interests in the mathematical history of early cultures, including Mesopotamia, Greece, India, and the Islamic Near East. Her *Chasing Shadows: Mathematics, Astronomy, and the Early History of Eclipse Reckoning* (Johns Hopkins University Press) focuses on the theoretical treatment of eclipse phenomena in the ancient world. She is, with Kim Plofker, completing a book on the history of Sanskrit astronomical tables in the second millennium.

John F. Nunn retired from his role as Head of the Anaesthesia Division of the Clinical Research Centre, British Medical Research Council. He has published widely on ancient Egyptian medicine, including a book with that title (British Museum Press, 1996).

Vivian Nutton, FBA, is Professor of the History of Medicine at the First Moscow State Medical University. His recent books include *Ancient Medicine* (2nd edn, 2013), a translation of Johann Guinter and Andreas Vesalius' *Principles of Anatomy according to the Opinion of Galen*, 1538 (2017), and *John Caius, An Autobibliography*.

Kim Plofker is Associate Professor in the Department of Mathematics at Union College in Schenectady, New York, USA. Her areas of interest include the history of mathematics and astronomy in India and the medieval Islamic world, and the cross-cultural transmission of scientific models.

Francesca Rochberg is Catherine and William L. Magistretti Distinguished Professor of Near Eastern Studies in the Department of Near Eastern Studies and the Office for the History of Science and Technology at the University of California, Berkeley. She publishes widely on the Babylonian astral sciences, especially celestial divination and astronomy.

Nathan Sidoli received his PhD from the University of Toronto, in the History and Philosophy of Science and Technology, and is currently Associate Professor of the History and Philosophy of Science at Waseda University, Tokyo. His research focuses on the Greek mathematical sciences and their development in medieval Arabic sources.

A. Mark Smith is Curators' Distinguished Professor of History at the University of Missouri. He has published widely in the history of optics and visual theory from antiquity to the seventeenth century, with special focus on the work of Ibn al-Haytham.

John Steele is Professor of the History of the Exact Sciences in Antiquity in the Department of Egyptology and Assyriology at Brown University. A historian of astronomy, much of his work focuses on the development of the astral sciences in Babylonia.

Liba Taub is Professor of History and Philosophy of Science at the University of Cambridge, where she is Director of the Whipple Museum of the History of Science; she is also a Fellow of Newnham College. Having published extensively on Greco-Roman science, as well as scientific instruments and models, her most recent monograph is *Science Writing in Greco-Roman Antiquity* (CUP, 2017).

Laurence Totelin is a senior lecturer in ancient history at Cardiff University. Her research focuses on Greek and Roman botany and pharmacology. She is the author of *Hippocratic Recipes: Oral and Written Transmission of Pharmacological Knowledge in Fifth- and Fourth-Century Greece* (Brill, 2009) and, with Gavin Hardy, *Ancient Botany* (Routledge, 2016).

Miira Tuominen (PhD 2002, University of Helsinki) is tenured at the University of Jyväskylä. She has published on a wide array of topics, including two monographs, *Ancient Philosophers on Starting Points for Knowledge* (2007) and *Ancient Commentators on Plato and Aristotle* (2009). She is now finishing a monograph on Porphyry's ethics in his *On Abstinence*.

Philip van der Eijk is Alexander von Humboldt Professor of Classics and History of Science at the Humboldt-Universität zu Berlin. He has published on ancient medicine, philosophy, and science, and comparative literature and patristics. Among his most recent publications is *Galen, Works on Human Nature, Vol. 1: Mixtures (De temperamentis)* (Cambridge, 2018) (co-authored with P. N. Singer).

Cristina Viano is Senior Researcher (Directeur de Recherche) at the Centre National de la Recherche Scientifique, Paris, Centre Léon Robin. Her main research fields concern the history of philosophy and ancient science, in particular Aristotle (natural philosophy, ethics, and rhetoric) and Alexandrian alchemy. She has published *La matière des choses. Le livre IV des Météorologiques d'Aristote, et son interprétation par Olympiodore* (Paris: Vrin, 2006).

GENERAL EDITORS' PREFACE

The idea for *The Cambridge History of Science* originated with Alex Holzman, former editor for the history of science at Cambridge University Press. In 1993, he invited us to submit a proposal for a multivolume history of science that would join the distinguished series of Cambridge histories, launched over a century ago with the publication of Lord Acton's fourteen-volume *Cambridge Modern History* (1902–12). Convinced of the need for a comprehensive history of science and believing that the time was auspicious, we accepted the invitation.

Although reflections on the development of what we call "science" date back to antiquity, the history of science did not emerge as a distinctive field of scholarship until well into the twentieth century. In 1912, the Belgian scientist-historian George Sarton (1884–1956), who contributed more than any other single person to the institutionalization of the history of science, began publishing *Isis*, an international review devoted to the history of science and its cultural influences. Twelve years later, he helped to create the History of Science Society, which by the end of the century had attracted some 4,000 individual and institutional members. In 1941, the University of Wisconsin established a department of the history of science, the first of dozens of such programs to appear worldwide.

Since the days of Sarton, historians of science have produced a small library of monographs and essays, but they have generally shied away from writing and editing broad surveys. Sarton himself, inspired in part by the Cambridge histories, planned to produce an eight-volume *History of Science*, but he completed only the first two installments (1952, 1959), which ended with the birth of Christianity. His mammoth three-volume *Introduction to the History of Science* (1927–48), more a reference work than a narrative history, never got beyond the Middle Ages. The closest predecessor to *The Cambridge History of Science* is the three-volume (four-book) *Histoire Générale des Sciences* (1957–64), edited by René Taton, which appeared in an English translation under the title *General History of the Sciences* (1963–4).

Edited just before the late twentieth-century boom in the history of science, the Taton set quickly became dated. During the 1990s, Roy Porter began editing the very useful Fontana History of Science (published in the United States as the Norton History of Science), with volumes devoted to a single discipline and written by a single author.

The Cambridge History of Science comprises eight volumes, the first four arranged chronologically from antiquity through the eighteenth century, the latter four organized thematically and covering the nineteenth and twentieth centuries. Eminent scholars from Europe and North America, who together form the editorial board for the series, edit the respective volumes:

Volume 1: *Ancient Science*, edited by Alexander Jones, University of Toronto, and Liba Taub, University of Cambridge
Volume 2: *Medieval Science*, edited by David C. Lindberg and Michael H. Shank, University of Wisconsin–Madison
Volume 3: *Early Modern Science*, edited by Katharine Park, Harvard University, and Lorraine Daston, Max Planck Institute for the History of Science, Berlin
Volume 4: *Eighteenth-Century Science*, edited by Roy Porter, late of Wellcome Trust Centre for the History of Medicine at University College London
Volume 5: *The Modern Physical and Mathematical Sciences*, edited by Mary Jo Nye, Oregon State University
Volume 6: *The Modern Biological and Earth Sciences*, edited by Peter J. Bowler, Queen's University of Belfast, and John V. Pickstone, University of Manchester
Volume 7: *The Modern Social Sciences*, edited by Theodore M. Porter, University of California, Los Angeles, and Dorothy Ross, Johns Hopkins University
Volume 8: *Modern Science in National and International Context*, edited by Ronald L. Numbers, University of Wisconsin–Madison, Hugh Richard Slotten, University of Otago, and David N. Livingstone, Queen's University of Belfast

Our collective goal is to provide an authoritative, up-to-date account of science – from the earliest literate societies in Mesopotamia and Egypt to the end of the twentieth century – that even nonspecialist readers will find engaging. Written by leading experts from every inhabited continent, the essays in *The Cambridge History of Science* explore the systematic investigation of nature and society, whatever it was called. (The term "science" did not acquire its present meaning until early in the nineteenth century.) Reflecting the ever-expanding range of approaches and topics in the history of science, the contributing authors explore non-Western as well as Western science, applied as well as pure science, popular as well as elite science, scientific practice as well as scientific theory, cultural context as well as intellectual content, and the dissemination and reception as well as the production of scientific knowledge. George Sarton would scarcely recognize this collaborative effort as the history of science, but we hope we have realized his vision.

David C. Lindberg
Ronald L. Numbers

ACKNOWLEDGMENTS

We thank David C. Lindberg and Ronald L. Numbers for the invitation to edit Volume 1 on *Ancient Science* in *The Cambridge History of Science*. David Lindberg was committed to this volume having broad coverage; with this in mind, we are pleased to have persuaded authors to contribute chapters covering such a wide geographical and chronological span. We are very sad that David did not live to see the final product, and we record our gratitude to him for his encouragement and thoughtful guidance. Ron Numbers was also extremely helpful and encouraging; for his continued enthusiasm, we are grateful.

Editorial staff at Cambridge University Press have been supportive throughout, providing skillful oversight and suggestions. In particular, we thank Alex Holzman, Eric Crahan, Deborah Gershenowitz, Dana Bricken, Kristina Deusch, Bethany Thomas, and Cassi Roberts, as well as Joshua Hey and Auriol Griffith-Jones. In Cambridge, Frances Willmoth, Emma Perkins, and Arthur Harris gave valuable assistance at various stages as did Michael Coxhead (London). Finally, we thank the authors for their contributions, patience, and good humor in bringing this volume to publication.

Alexander Jones
Liba Taub

INTRODUCTION

Alexander Jones and Liba Taub

Volume One of *The Cambridge History of Science* traces the principal scientific traditions of the Old World in antiquity that have left substantial textual evidence. Some of these traditions are also represented by other sorts of evidence – archaeological, visual, and material. However, our understanding of what these traditions were is in every instance grounded primarily in texts. In fact, it is not merely the case that written sources are a primary source of knowledge of these traditions, but these traditions were in themselves functions of literate scholarly cultures.

These traditions eventually became part of the interconnected intellectual world of the Middle Ages, especially through the wide circulation of knowledge that was facilitated by the spread of Islam. As a general phenomenon this interconnectedness became apparent only retrospectively; in many cases, however, there was much interaction already in antiquity. Historians of ancient science have long recognized the significance of transmissions and transformations crossing geographical and linguistic boundaries; more recently we are increasingly conscious that knowledge transfers occurred within individual cultures, between distinct communities of people separated by education, self-identification, and other differentiating factors. For example, we can investigate not just transmission of astronomical knowledge between Mesopotamia and Greece, but between specialists and non-specialists or between distinct groups of specialists (for example, mathematicians, astrologers, philosophers, even physicians).

We are not always able to construct detailed narratives, because we are always dealing with fragmentary evidence, which may be densely abundant for one particular context and chronological period but otherwise almost nonexistent. We are seldom offered a direct glimpse of moments of discovery or innovation; biographical information about key figures in the processes of change is usually scanty or unreliable. What we can best describe and investigate are the practices that subsisted between and as a consequence of such moments.

Our understanding of ancient scientific traditions has changed for a number of reasons, including the discovery of new evidence (through archaeology, say, or rediscovered manuscripts) as well as the emergence of new historiographical questions and methods. A particularly important example of such a change is the shift in emphasis in current scholarship from preoccupation with scientific concepts and methods to the people who engaged in scientific work, their education, their motivations, and their professional status. Who counted as a 'professional' varies from culture to culture. For example, in China and Mesopotamia – but not, apparently, in Greece – astronomers were professionals holding appointments and carrying out set duties. Many cultures had various types of health professionals, including physicians, midwives, and root-cutters. Where there were professions, this implied professional training, but not necessarily formal accreditation.

While we recognize that applying the name 'Science' – as if there ever existed in antiquity a unifying conception even approximately coextensive with the modern one – is an anachronism, nevertheless it is a convenient and useful anachronism. Rather than applying a single criterion for what constituted a scientific *tradition*, we take into consideration three elements that did not all have to be present in a particular tradition: the collection and organization of information and knowledge; prediction; and causal explanation. While thus refusing to define science in a reductive manner, we regard some combination of these activities as characterizing scientific endeavors.

The ancient scientific traditions dealt with in this volume were not exclusively theoretical, pursued purely or even primarily as knowledge for knowledge's sake. Even traditions such as Greek mathematics, which were caricatured even in antiquity as 'ivory-tower' pursuits, had practical applications and social roles. Conversely, it is sometimes impossible to tell from the surface level of many of the texts whether they were truly concerned with real-world problems, because many of these problems are cast as practical, but were actually artificially constructed intellectual or didactic exercises.

We have not felt it necessary to attempt to cover every ancient scientific tradition, either geographically or culturally, nor is the aim of this volume to discuss every single tradition that might qualify as scientific according to the principles we have given. We believe that we have included the most important and well-documented scientific traditions of antiquity, and that, broadly speaking, the chapters reflect the variety of such traditions in each major culture to the extent that this is possible in the present state of historical scholarship. Our aim is to be representative, not comprehensive.

We resist the temptation to project the categories of modern scientific culture backwards; thus it is not meaningful to write a history of 'chemistry' or 'physics' in any ancient context. We recognize that even a term such as 'astronomy' did not mean the same thing in any ancient context that it would in a modern university, but there did at least exist more or less coherent intellectual traditions in which the heavenly bodies were the primary objects of study. Both the areas of ancient science and the ways in which they have come to be categorized have to a large extent been determined by the trajectories of past scholarship, and by the readiness of present-day scholars to study what they regard as scientific or technical fields. Nevertheless, we have attempted to be inclusive when considering what counts as science. For example, chapters treat botany, understood in antiquity primarily as being about classification and *materia medica*; music theory as an explanatory and mathematical science largely concerned with the pitch systems of ancient music; and astrology and astral divination as complex systems with close ties to astronomy and cosmology.

Each chapter's author has determined the appropriate chronological range to cover: generally, each one starts from the earliest documented period, but where 'antiquity' ends is a matter of convention that varies from culture to culture. So, for example, as we see in the chapter on Indian mathematics, some styles of teaching and learning persist down to the present. Labels like 'antiquity', 'classical', and 'medieval' are thus terms of convenience.

In some of the cultures considered here, notably Egypt, Mesopotamia, and China, institutional and administrative settings established the frameworks for some scientific traditions. In Mesopotamia, for example, literacy, scholarship, and scientific activities were closely related, and largely overlapped. Institutions and governmental structures were less significant in other cultures, while other social and cultural factors had a greater impact in shaping scientific and technical work. In Greece and Rome, certain scientific pursuits were associated with specific philosophical sects or 'schools' (e.g. zoology and botany with the Peripatetics), but institutions or patronage supporting science – when they existed at all – tended to be short-lived.

In some cultural contexts scientific practices were embedded within hieratic institutions: for example, observational and mathematical astronomy in the temples of Babylonia. Elsewhere we find the coexistence of religion-based and science-based practices, such as physicians operating within the confines of temples of Asclepius where divine dream-based healing was practiced. Even in the most apparently secular approaches to scientific questions, terms like 'divine' frequently occur. One is hard-pressed to find any instance in which a scientific author attacks institutionalized religion; indeed, within the cultures under consideration here, there is almost no evidence of any adversarial relationship between religion and science.

Each chapter of this volume is intended to be self-standing while contributing to the larger project of *The Cambridge History of Science*. As editors, we have not imposed a single approach on the authors of individual chapters. Some are presented chronologically, some thematically; contributors have adopted whatever approach they regard as most illuminating. They were not discouraged from giving an informed, personal interpretation of material; hence, the chapters do not always present a totally neutral 'take' on the subject. For the most part, individual authors have not presented their subject as the precursor of something that happens later. Rather, each chapter considers the science of that culture as something worth understanding in its own right and in its own context. While the contributors have aimed to make their subject comprehensible to non-specialists, there is much that will be of interest even to specialists. From our own experience as editors, we know that we have each learned much through reading the chapters presented here.

Part I

MESOPOTAMIA

1

SCIENCE AND ANCIENT MESOPOTAMIA

Francesca Rochberg[1]

How the study of physical phenomena in ancient Mesopotamia relates to the history of science is a question as important for the study of ancient Mesopotamia as it is for the history of science. It addresses both the nature of knowledge in the oldest literate culture as well as the historical reach of what we call science. If the essence of science is to be found in its systematization of knowledge about phenomena and in the various practices associated with such knowledge systems – practices such as celestial observation, prediction, and explanation – then science was a central part of cuneiform intellectual culture.

Divination, magic, and medicine were integral parts of what the scribes termed "scholarship" (*ṭupsarrūtu*, literally "the art of the scribe") as well as "wisdom" (*nēmequ*). Scholarship and wisdom were classified as a "secret of the great gods" (*pirišti ilāni rabûti*), referring to a conception of the origins of knowledge with the divine. Cuneiform knowledge was thus reserved for initiates, and injunctions against scribes who were not among the privileged few with access to texts classified as "secret" (*pirištu*) or "guarded" (*niṣirtu*) are known from the Middle Babylonian (ca. sixteenth to eleventh centuries BCE) to the Late Babylonian (ca. fourth to first centuries BCE) periods.[2] The classification of knowledge as secret applied to divinatory texts, incantations, apotropaic rituals against ominous signs, medical texts, scholarly commentaries on divinatory texts, and astronomical texts, and by the late first millennium the interrelations among these forms of knowledge become more apparent. A Late Babylonian astronomical text giving rules for calculating month lengths and intervals of lunar visibility around the full moon,

[1] There are various abbreviations that are standard within Assyriology. Those unfamiliar with these may consult http://cdli.ox.ac.uk/wiki/abbreviations_for_assyriology.

[2] See A. Lenzi, *Secrecy and the Gods: Secret Knowledge in Ancient Mesopotamia and Biblical Israel* (State Archives of Assyria Studies 19; Helsinki: University of Helsinki Press, 2008), pp. 64–6, and cf. P.-A. Beaulieu, "New Light on Secret Knowledge in Late Babylonian Culture," *ZA* 82 (1992), 98–111.

for example, begins with the statement: "Tablet of the guarded secret of heaven, secret knowledge of the great gods."[3]

The sources for cuneiform scholarship span two millennia, beginning in the Old Babylonian Period (ca. 1800–1600 BCE) and continuing until the early centuries of the Common Era. This chapter focuses first on the content of cuneiform scholarship and wisdom, follows with aspects of the methods of the scholar-scribes – observation, prediction, and explanation – particularly with respect to celestial divination and astronomy, and closes with a note on the modern nomenclature and classification of cuneiform astronomical/astrological texts.

CUNEIFORM SCHOLARSHIP AND WISDOM

Assyro-Babylonian scholarly divination originated in Babylonia in the second millennium BCE, where collections of texts for the reading of signs, particularly those from the heavens and from the exta of sacrificed sheep, were typically formulated in the casuistic, or case form "If P then Q," style, as in the following:

> If water secretes inside the gall bladder: The flood will come.[4]
>
> If the gall bladder is turned and has wrapped around the "finger": The king will seize the enemy country.[5]

The tradition was both systematic and authoritative, and tablet series containing celestial and terrestrial signs (Akkadian *ittātu*) became part of the spread of cuneiform writing to the west of Babylonia during the second millennium, to Emar, Harādum, Alalakh, and Qatna, as well as to the Hittite capital of Hattusas, as important components of an international cuneiform scribal tradition.[6] Development of scholarly divination in the Middle Babylonian (ca. 1600–1100 BCE) and Middle Assyrian periods (ca. 1400–1050 BCE) indicates the formation at that time of authoritative series, which later, especially in the seventh century BCE, assumed a prominent place in the state libraries of Nineveh, Nimrud, and

[3] BM 42282+42294 obv. 1 [*ṭu*]*ppi niṣirtu šamê pirištu ilāni rabûti*; see L. Brack-Bernsen and H. Hunger, "BM 42282+42294 and the Goal-Year Method," *SCIAMVS* 9 (2008), 6.

[4] A. Goetze, *Old Babylonian Omen Texts* (Yale Oriental Series 10; New Haven, CT and London: Yale University Press, 1947), no. 31, col. ii, ll. 38–41.

[5] Ibid., no. 31, col. ii, ll. 24–30.

[6] The 13th century BCE Emar omens are found in D. Arnaud, *Recherches au pays d'Astata Emar, vol. 6* (Paris: Recherche sur les civilisations, 1987); Nos. 650–65 are celestial omens. The Har~dum text is published in F. Joannès, "Un Précurseur Paléo-Babylonien de la Série Šumma Ālu," in H. Gasche, M. Tanret, C. Janssen, and A. Degraeve (eds.), *Cinquante-deux réflexions sur le Proche-Orient ancien offertes en hommage à Léon de Meyer* (Mesopotamian History and Environment, Occasional Publications 2; Leuven: Peeters, 1994), pp.305–12.

Assur.[7] Cuneiform scribal culture continued in the Babylonia of the Neo-Babylonian, Hellenistic, and Parthian periods, preserving as well as expanding upon the traditional knowledge of omens, rituals, prayers, hemerologies, commentaries, and medical, magical, and astronomical/astrological texts, until the end of cuneiform writing itself.

Compilations of omens in lists represent the result of scholarly systematization and theorization about the meaning of signs, thus establishing in our minds their connection to science. To the divinatory sciences, therefore, belong all the cuneiform scholarly texts formulated in the casuistic manner, which associated a protasis (if-clause) with an apodosis (then-clause) such that a phenomenon was systematically "explained." Explanation in this context is meant in the sense used by David Pingree when he defined science as "a systematic explanation of perceived or imaginary phenomena or else [it] is based on such an explanation."[8] In Pingree's view, Babylonian divination was "a systematic explanation of phenomena based on the theory that certain of them are signs sent by the gods to warn those expert in their interpretation of future events."[9] While this statement only opens up for debate what the nature of explanation is in the divinatory sciences, one way in which divination was explanatory has to do with the relation of an omen apodosis to its protasis and how events were thought to be connected to one another. The establishment of connections, referred to in the texts as divine decisions or judgments, further manifests the Babylonian notion of divine causality and the view of an intimate involvement of the gods in physical phenomena.[10]

The divine judgments came in the form of socially relevant events such as attack by enemies, fall of market prices, hunger and want, devastation by flood, pestilence, or plagues of locusts. Fortune or misfortune for the ruling elite (king, prince, lord) was the main concern, as in the following:

> If Venus stands behind the Moon: the king will have no rival.
> If Venus stands in the Moon's position: the king's land will revolt against him.

[7] In addition to the text series, see the correspondence between scholars and the kings Esarhaddon and Assurbanipal in S. Parpola, *Letters from Assyrian Scholars to the Kings Esarhaddon and Assurbanipal*, 2 vols. (Neukirchen-Vluyn: Butzon and Bercker Kevelaer, 1970–83), vol. 1: *Texts*; vol. 2: *Commentary and Appendices*; H. Hunger, *Astrological Reports to Assyrian Kings* (State Archives of Assyria (= SAA) 8; Helsinki: University of Helsinki Press, 1992); and S. Parpola, *Letters from Assyrian and Babylonian Scholars* (State Archives of Assyria (= SAA) 10; Helsinki: University of Helsinki Press, 1993).

[8] D. Pingree, "Hellenophilia Versus the History of Science," *Isis* 83 (1992), 554–63, quotes from 559–60.

[9] Ibid.

[10] F. Rochberg, *In the Path of the Moon: Babylonian Celestial Divination and Its Legacy* (Leiden and Boston, MA: Brill, 2010), pp. 411–24.

If Venus reaches the Moon and enters into the Moon: the king's son will seize his father's throne.[11]

Such public apodoses were generally found in celestial divination, malformed birth omens (of the series *Šumma izbu*), and extispicy. Other omen series (as in the physiognomic omens of *Alamdimmû* or the birth omens of *Iqqur īpuš*) focused on the stability of a man's household, personal health, wealth, happiness, and lifespan. Private apodoses would later be integrated within natal astrological omens and horoscopes.[12]

As most clearly represented in the surviving texts of the library at Nineveh, the corpora of five distinct scholarly professions represent the scholars' repertoire of knowledge, namely, those of the "scribe of *Enūma Anu Enlil*" (*ṭupšar Enūma Anu Enlil*), who was expert in astral phenomena, the "one who inspects (the liver and exta)" (*bārû*), i.e., the diviner expert in extispicy; the "exorcist" (*āšipu*), who treated human beings afflicted by divine disfavor via incantations and rituals aimed at re-establishment of the right relationship between human and divine; the "physician" (*asû*), who treated the body in the grip of demonic or divine influence (what we call disease); and the "lamentation priest" (*kalû*), who was responsible for religious ritual performance (songs of lamentation, also the playing of the kettledrum for the ritual against the evil of a lunar eclipse).[13] Rigid distinctions did not obtain between these scribal professions and the texts they wrote, copied, and utilized. Omens (including astral, abnormal birth, and human physiognomic) and astronomical texts are, for example, within the professional domain of *āšipu*s and *kalû*s.

Astral omens begin to appear in the Old Babylonian period with particular attention to lunar eclipses. Eventually the canonical *Enūma Anu Enlil* encompassed a range of phenomena of the moon, sun, planets, fixed stars, and weather. Of particular though not exclusive interest to the scholars were periodic phenomena, and the understanding of astronomical periodicities was therefore increasingly of importance. The letters to the Assyrian kings Esarhaddon and Assurbanipal in the seventh century reflect some ability to predict astronomical phenomena such as planetary appearances and even lunar eclipses, at least in the short term. Also attested in the seventh century, in a tablet that gives celestial omens in a numerical cryptography, are periods

[11] E. Reiner and D. Pingree, *Babylonian Planetary Omens, Part 3* (Groningen: Styx, 1998), p. 45, lines 38–9 and 46.

[12] F. Rochberg, *The Heavenly Writing: Divination, Horoscopy, and Astronomy in Mesopotamian Culture* (Cambridge: Cambridge University Press, 2004), pp. 202–6, and passim. See also F. Rochberg, *Babylonian Horoscopes* (Transactions of the American Philosophical Society 88; Philadelphia, PA: American Philosophical Society, 1998).

[13] P.-A. Beaulieu and J. P. Britton, "Rituals for an Eclipse Possibility in the 8th Year of Cyrus," *Journal of Cuneiform Studies* (henceforth "*JCS*") 46 (1994), 73–86; D. Brown and M. Linssen, "BM 134701 = 1965–10–14,1 and the Hellenistic Period Eclipse Ritual from Uruk," *RA* 91 (1997), 147–66; and M. Linssen, *The Cults of Uruk and Babylon: The Temple Ritual Texts as Evidence for Hellenistic Cult Practice* (Leiden and Boston, MA: Brill, 2004), pp. 306–20.

for the planetary synodic cycles, some consistent with later so-called goal-year periods (see, pp. 20 and 24, below).[14] Prediction of lunar and planetary phenomena utilizing the parameters of so-called goal-year periods is also attested in the sixth-century Strassmaier Cambyses 400.[15] By the fifth century and later, the small group of cuneiform horoscopes (or proto-horoscopes[16]) required the calculation of planetary positions, and these could either be made with goal-year methods, or perhaps by means of interpolations from the mathematical schemes characteristic of the ephemerides of Seleucid Babylonia.[17] Predictions for astral phenomena did not diminish the ominous significance of signs. The response to an occurrence of a lunar eclipse, even in the Seleucid period (after 300 BCE) when prediction of eclipses is well attested, was the performance of an apotropaic ritual to dispel its evil.[18]

Other celestial omens are not from periodic phenomena, such as the following planetary omens (note that square brackets indicate breaks in the clay tablet where the text is restored by duplicates):

> [If Venus at her appearance is red: (abundance for the people)], the harvest of the land will succeed, the king of Akkad will experience [joy] – the east wind blows.
>
> If Venus at her appearance is black: Enlil will glare angrily [at the land], in the land business will be poor [. . .], the south wind blows.
>
> [If Venus at her appearance is white:] There will be drought in the land, [. . .] – the north wind blows.[19]

In addition to celestial and terrestrial omens, medical texts systematized both symptoms and therapeutic techniques for reference purposes.[20] Because the aetiology of disease was considered divine, demonic, from ghosts (*qāt eṭemmi* "hand of a ghost"), witches (*kišpu* "witchcraft"), curses (*mamītu*), or anything evil (*mimma lemnu* "whatever is evil"), the *āšipu* ("exorcist") specialized in the incantations and liturgy used to appeal to deities who had the

[14] See C. J. Gadd, "Omens Expressed in Numbers," *JCS* 21 (1967), 52–63, especially p. 61. The periods in question are Saturn 59, Venus 8, Mars 15, and Jupiter 12. See J. P. Britton, "Studies in Babylonian Lunar Theory Part II. Treatments of Lunar Anomaly," *Archive for History of Exact Sciences* 63 (2009), 357–431, especially p. 349.

[15] J. P. Britton, "Remarks on Strassmaier Cambyses 400," in M. Ross (ed.), *From the Banks of the Euphrates: Studies in Honor of Alice Louise Slotsky* (Winona Lake, IN: Eisenbrauns, 2008), pp. 7–33, with bibliography.

[16] H. Hunger and D. Pingree, *Astral Sciences in Mesopotamia* (Handbuch der Orientalistik; Leiden, Boston, MA, and Cologne: Brill, 1999), pp. 26–7.

[17] Rochberg, *Babylonian Horoscopes*, pp. 7–11.

[18] See note 12.

[19] Reiner and Pingree, Babylonian Planetary Omens, Part 3, pp. 40–3, Group A, VAT 10218 lines 13–15.

[20] I. L. Finkel, "On Late Babylonian Medical Training," in A. R. George and I. L. Finkel (eds.), *Wisdom, God and Literature: Studies in Assyriology in Honor of W. G. Lambert* (Winona Lake, IN: Eisenbrauns, 2000), pp. 137–224; M. J. Geller, *Ancient Babylonian Medicine: Theory and Practice* (Oxford: Wiley-Blackwell, 2010).

power to heal and protect the patient. Incantations could be used in combination with other prophylactic and apotropaic acts such as fumigation, the topical application of salves, and the use of amulets as means to appease the divine sources of illness and pain. Incantations were collected, standardized up to a point, and had a kind of canonical force in the same way as did divinatory texts.[21]

While divinatory texts began with discrete genre boundaries (celestial, physiognomic, birth, etc.), in the last half of the first millennium, during Persian (Achaemenid), Hellenistic (Seleucid) and Parthian (Arsacid) Babylonia, interrelations are increasingly integrated, and Late Babylonian scholarly commentaries establish more direct connections between celestial and terrestrial realms. The integration of astral with terrestrial divinatory sciences seems to have been made possible by the development of astrology, that is, by the application of celestial signs for the human being (and the human body) in general, no longer focusing, as did *Enūma Anu Enlil*, only on the king. In one commented text, for example, omen series concerning human appearance, health, and births were brought into relation with celestial signs. Its opening lines:

> (The omen series) "If a Malformed fetus," (the omen series) "Symptoms," (the omen series) "Physical Characteristics." Aries, Taurus, Orion are for predicting the appearance. When they (the planets?) "reach" (the various zodiacal signs) it refers to physical characteristics. Observe the secret of heaven and earth![22]

The laconic nature of the commentary leaves open the question of exactly what the connections between birth, medical, physiognomic, and astrological phenomena were, yet a decidedly astrological, that is genethlialogical (birth astrology), turn has been taken. An important feature of later astronomical texts (Diaries and Almanacs) was to track when the planets "reached" each zodiacal sign. The colophons of Almanacs state that these texts are "measurements of the 'reachings' of the (divine) planets" (*mešhi ša kašādī ša* ^{d}*bibbī*). The arrival of a planet into a sign was presumably astrologically significant, though the statements of these "reachings" in astronomical texts do not provide an astrological meaning. These same astronomical texts, namely the Almanacs and Diaries, were very likely

[21] For example, the series *Šurpu* (see E. Reiner, *Šurpu: A Collection of Sumerian and Akkadian Incantations* (AfO Beiheft 11; Graz: Archiv der Orientforschung, 1958)); the anti-demon text *Udug-hul* (see M. Geller, *Forerunners to Udug-Hul: Sumerian Exorcistic Incantations* (Wiesbaden: F. Steiner Verlag, 1985)); the ritual series *Muššu'u* (see B. Böck, "'When You Perform the Ritual of Rubbing': On Medicine and Magic in Ancient Mesopotamia," *Journal of Near Eastern Studies* 62 (2003), 1–16, with further bibliography).

[22] See B. Böck, "'An Esoteric Babylonian Commentary' Revisited," *Journal of the American Oriental Society* 120 (2000), 615–20.

used as sources for the zodiacal positions of the planets quoted in cuneiform horoscopes.[23]

Astrology finds a connection to extispicy in the Late Babylonian Period as well, when traditionally ominous parts of the liver are associated with a god, one of the twelve months, and a heliacally rising star, thus: "the Path (of the liver) is Šamaš, Ajāru, Taurus; the gall bladder is Anu, Tašrītu, Libra," and so on.[24] Magical practice also established new connections to the zodiac, such as in a list of spells with their correlated regions in the zodiacal signs.[25] Even the term "sign" or "ominous part" (literally "flesh" UZU = *šīru*) found earlier in liver omens occurs again in a Seleucid astrological context where dodekatemoria (1/12ths) of zodiacal signs are referred to as 12 UZU.MEŠ HA.LA ša $^{\text{múl}}$LÚ.HUN.GÁ (variant, $^{\text{múl}}$LU) "the 12 signs (ominous parts) of the zodiacal sign Aries."[26]

Astronomy and the preservation by the *ummânu* of esoteric learning in magical, medical, and liturgical texts dominated the activities of the Late Babylonian literati in the last centuries of the cuneiform tradition. Testimony to the continuation of the practice of Babylonian astronomy in the first century CE comes from the Elder Pliny (23–79 CE), who claims to have seen the astronomers in Babylon in the "Temple of Jupiter-Bēl."[27] On the surface, the astronomical contents of the quantitative predictive texts of the Late Babylonian Period do not appear to relate to divinatory knowledge, yet the texts that predict such events as rains and floods, enemy attacks, and market prices, as well as the few preserved colophons on ephemerides, all reflect the fact of the identification of the astronomers as members of the classes of *ṭupšar Enūma Anu Enlil, kalû*, and *āšipu*. Furthermore, a few preserved rubrics indicate that these tablets were classified by the scribes, along with other texts of *ṭupšarrūtu*, as "secret" (*pirištu*), as in the following colophon from a text dated to the second century BCE:

> On eclipses of the moon.
>
> Tablet of Anu-bēl-šunu, lamentation priest of Anu, son of Nidintu-Anu, descendant of Sin-lēqi-unninni of Uruk. Hand of Anu-[aba-utêr, his son, scri]be of *Enūma Anu Enlil* of Uruk. Uruk, month I, year 12[1?] Antiochus [. . .]

[23] F. Rochberg, "Babylonian Horoscopy: The Texts and Their Relations," in N. M. Swerdlow (ed.), *Ancient Astronomy and Celestial Divination* (Cambridge, MA: MIT Press, 1999), pp. 39–60, and see also Rochberg, *Heavenly Writing*, pp. 145–57.

[24] SpTU 14, 159; see U. Koch-Westenholz, *Babylonian Liver Omens: The Chapters Manzāzu, Padānu, Pān tākalti of the Babylonian Extispicy Series Mainly from Assurbanipal's Library* (The Carsten Niebuhr Institute of Near Eastern Studies Publications, 25; Copenhagen: Museum Tusculanum Press, 2000), pp. 24–5.

[25] BRM 4 20, with BRM 4 19 (and parallels with STT 300 although without the zodiacal references), see M. J. Geller, *Melothesia in Babylonia: Medicine, Magic, and Astrology in the Ancient Near East* (Science, Technology, and Medicine in Ancient Cultures 2; Berlin: De Gruyter, 2014), pp. 28–57.

[26] TCL 6 14:11, 12, 13, and 20; see A. J. Sachs, "Babylonian Horoscopes," *JCS* 6 (1952), 49–75, p. 66.

[27] Pliny, *The Natual History* VI 123, VII 193.

Whoever reveres Anu and Antu [. . .]
Computational table. The wisdom of Anu-ship, exclusive knowledge of the god [. . .]
Secret knowledge of the masters. The one who knows may show (it) to an[other one who knows]. One who does not know may not [see it. It belongs to the forbidden things] of Anu, Enlil [and Ea, the great gods].[28]

CELESTIAL OBSERVATION[29]

Astronomical texts do not appear prior to the Old Babylonian period (2000–1600 BCE). No astronomical texts are known in the Sumerian language; however, some rudimentary recognition of observed astronomical phenomena is attested in the early third millennium BCE (Uruk level IV) in an Uruk cultic text concerning offerings to the goddess Inanna as the morning and evening star. A cult to the astral Inanna continued in Sumerian city-states through the third millennium. Otherwise, poetic descriptions, such as of the moon god and his many "cattle," i.e., the stars, are found in Sumerian literature,[30] and lists of star names are found in Sumerian lexical texts of the Old Babylonian period (*Ura* 5, together with geographical names[31]), roughly contemporaneous with the earliest evidence for systematic attention to the celestial bodies as signs.

One of the modern debates about the role of observation in the cuneiform tradition has to do with the origins of omen divination, namely whether the observation of co-occurences of phenomena led to the idea that one phenomenon (P) could indicate another (Q).[32] The evidence is clear that signs were studied for their appearances, regularities, and irregularities, and the patterns of their occurrence. No evidence for an observational connection of signs to portents, however, can be demonstrated. A variety of non-observational principles can

[28] O. Neugebauer, *Astronomical Cuneiform Texts*, 3 vols. (London: Lund Humphries, 1955), No. 135, reading of the date is uncertain, see p. 19 for discussion; also H. Hunger, *Babylonische und Assyrische Kolophone* (Alter Orient und Altes Testament 2; Kevelaer: Butzon & Bercker; Neukirchen-Vluyn: Neukirchener Verlag des Erziehungsvereins, 1968), No. 98.

[29] See also F. Rochberg, "Scientific Observation and Knowledge of the World in Cuneiform Culture" in E. Robson and K. Radner (eds.), *Oxford Companion to Cuneiform Culture* (Oxford: Oxford University Press), pp. 618–36.

[30] See F. Rochberg,"Sheep and Cattle, Cows and Calves: The Sumero-Akkadian Astral Gods as Livestock," in S. Melville and A. Slotsky (eds.), *Opening the Tablet Box: Near Eastern Studies in Honor of Benjamin R. Foster* (Culture and History of the Ancient Near East, 42; Leiden and Boston, MA: Brill, 2010), pp. 347–59.

[31] See http://oracc.museum.upenn.edu/dcclt/Q000042, lines 387–410.

[32] For example, P. Huber, "Dating by Lunar Eclipse Omina with Speculations on the Birth of Omen Astrology," in J. L. Berggren and B. R. Goldstein (eds.), *From Ancient Omens to Statistical Mechanics: Essays on the Exact Sciences Presented to Asger Aaboe* (Copenhagen: University Library, 1987), pp. 3–13.

be identified in the texts to explain how a sign was correlated with a portent.[33]

As seen in the examples quoted above, one such principle was analogy. The water inside the gall bladder signifies the flood. The gall bladder wrapped around the "finger" (identified with the band of tissue called the *processus caudatus*, or possibly the *processus pyramidalis*) was read as a visual analogue for the king's taking of an enemy. Not every omen exhibits such analogic reasoning from protasis to apodosis, but in many cases an association by homophony or synonymy will explain the connection between protasis and apodosis. Whatever observational dimension is to be found in the divinatory sciences is restricted to the observation of the signs themselves, not to the form of reasoning known as "after this, therefore because of this" (*post hoc ergo propter hoc*). Observation itself is an important part of why the omen texts have been classified as "scientific" in modern scholarship, and legitimately so, as the observation of physical phenomena was foundational to the development of knowledge about their intrinsic properties and behavior.

A systematic accounting of seasonal astronomical phenomena is first attested in a Middle Assyrian Period compendium entitled MUL.APIN "Plow Star."[34] Included are lists of stars; lists of periods of visibility for the five naked-eye planets; periods of invisibility for the planets Venus, Jupiter, Mars, and Saturn; intercalation schemes to reconcile the solar year with the lunar month; equinoxes and solstices; the variation in length of daylight; and omens (not identical to those of *Enūma Anu Enlil*). The interest in the planets is in the seasonal recurrence of heliacal risings and settings, phenomena that were taken to be signs, as is clear from *Enūma Anu Enlil* Tablet 63, discussed below, p. 19. Of six star-lists in MUL.APIN, two reflect an interest in the seasonal reappearance of stars or constellations. These lists give schematic dates (on the 1st, 5th, 10th, 20th, and 25th of an ideal month, that is, one of twelve thirty-day months) for the heliacal risings of thirty-five stars (constellations) and the associated intervals between their risings.[35] While not explicitly divinatory, MUL.APIN includes some omens. Its astronomical content, however, is consistent with that of *Enūma Anu Enlil*, and it devotes one of its star lists to the associations between stars and gods. MUL.APIN provides a representation of the state of astronomical knowledge of the end of the second millennium BCE.

The phenomena compiled in the Old Babylonian forerunners to *Enūma Anu Enlil*, MUL.APIN, and the Astrolabes are based presumably on

33 J.-J. Glassner, "Pour un lexique des termes et figures analogiques en usage dans la divination mésopotamienne," *Journal asiatique* 272 (1984), 15–46, and Rochberg, *In the Path of the Moon*, pp. 399–409, especially pp. 400–4.

34 See H. Hunger and D. Pingree, *MUL.APIN: An Astronomical Compendium in Cuneiform* (Archiv für Orientforschung Beiheft 24; Horn: Verlag Ferdinand Berger & Söhne, 1989).

35 Further discussion is found in Hunger and Pingree, *Astral Sciences*, pp. 57–73.

longstanding knowledge derived from naked-eye astronomy, though in no way do these represent contemporary observational texts. In celestial omens there is a sense in which the phenomena considered ominous were expected to "appear" or be "observed," as is clear from the frequent use of these verbs in the protases of celestial omens. However, the ominous astral phenomena written in the protases of omen texts cannot be taken to represent datable observations of phenomena. True observations of astral phenomena are found in two groups of texts. The first, in which the phenomena are clearly to be interpreted as omens, is a group of Neo-Assyrian scholarly texts from the reigns of Esarhaddon (680–669 BCE) and Assurbanipal (668–627 BCE), now termed "Reports."[36] The second, where the relationship to omens is less clear, is the archive compiled in the city of Babylon from the eighth century BCE (only extant from the seventh century) to the first century BCE, now termed "Diaries."[37]

Concern about visibility of ominous phenomena is revealed in the following Report:

> Twice or thrice we watched for Mars today (but) we did not see (it); it has set. Maybe the king my lord will say as follows: "is there any (ominous) sign in (the fact) that it set?" (I answer): "There is not."[38]

Or:

> Concerning Mercury, about which the king my lord wrote to me: yesterday Issar-šumu-ereš had an argument with Nabû-ahhe-eriba in the palace. Later, at night, they went and all made observations; they saw (it) and were satisfied.[39]

The Reports are clearly concerned with the meaning of the phenomena as signs, and evaluation of relevant omens are given in the Reports by means of citations from *Enūma Anu Enlil*. Within the period covered by the Reports, mostly during the first half of the seventh century, and from cities such as Assur, Babylon, Nippur, Uruk, Dilbat, Cutha, and Borsippa, the Reports were written very soon after the observations were made and transported to Nineveh, and in some cases ritual apotropaic measures were taken to avoid bad portents. The observations were made for the purpose of determining portents, but some ominous phenomena that were known to be periodic were by then already the objects of numerical schematization.[40]

[36] Hunger, *Astrological Reports.*

[37] H. Hunger and A. J. Sachs, *Astronomical Diaries and Related Texts from Babylonia*, 6 vols. (Vienna: Österreichische Akademie der Wissenschaften, 1988–2006).

[38] Hunger, *Astrological Reports*, 7: 5–rev.4.

[39] Ibid., 83:4–rev.3.

[40] See the discussion of ideal planetary schemes in D. Brown, *Mesopotamian Planetary Astrononomy-Astrology* (Cuneiform Monographs, 18; Groningen: Styx, 2000), pp. 113–22.

The duration of the moon's visibility at night is a good example, as seen in *Enūma Anu Enlil* Tablet 14 (discussed below, pp. 19–20).[41]

In the Diaries, the combination of astronomical data about the moon and the planets with political and economic events is somewhat reminiscent of *Enūma Anu Enlil*. A text will combine a nightly record of the positions of the planets with respect to a certain set of ecliptical stars (Normal Stars), eclipses, synodic appearances of the planets, and information as to the prices of barley, wool, mustard, and sesame. Additionally, the level of the Euphrates River will be reported, as well as the zodiacal signs in which each planet is found at the end of the month. Other notable events are included, such as cultic or military activities affecting the king. The following Diary reports on the routing of Darius III's army by Alexander the Great at Gaugamela in the Diary for –330:

> On the morning of the twenty-fourth of the month of Ulûlu, the king of the world [Alexander] raised his standard [lacuna]. The armies engaged each other and the king's soldiers suffered a heavy defeat. The troops abandoned their king [Darius] and headed back to their cities. They fled to the lands in the east.[42]

Finally, for the pre-Seleucid period, seventh-century tablets with observations of the synodic phenomena of the planets are sparsely attested, one from years 2 to 10 of Šamaš-šumu-ukīn and one from years 1 to 14 of Kandalanu, of first and last visibilities of Mars and of Saturn respectively.[43] Another Saturn observation text from sixth-century Uruk includes first and second stations to the first and last visibilities during the period from year 28 to 31 of Nebuchadnezzar II.[44] During this period, on the basis of the few extant planetary observation texts, the practice of citing planets with respect to a certain set of Normal Stars and the use of cubits, fingers, and degrees (UŠ) was still in the process of standardization.[45]

Typical of cuneiform scholarly texts is the long lacuna between the seventh-century copies of MUL.APIN and the next group of extant astronomical sources, from the fourth century and later. As is the case in the interim period between the abundance of Neo-Assyrian texts relating to astronomy and celestial divination and the bulk of extant texts from the

[41] The edition is in A. George and F. N. H. Al-Rawi, "Enuma Anu Enlil XIV and Other Early Astronomical Tables," *Archiv für Orientforschung*, 38–39 (1991), 52–73.

[42] Hunger and Sachs, *Astronomical Diaries and Related Texts*, vol. 1, No. –330: 14'–18', translation from http://www.livius.org/aj-al/alexander/alexander_z7.html.

[43] Hunger and Pingree, *Astral Sciences*, pp. 173–4.

[44] H. Hunger, "Saturnbeobachtungen aus der Zeit Nebukadnezars II," in Joachim Marzahn and Hans Neumann (eds.), *Assyriologica et Semitica: Festschrift für Joachim Oelsner* (Münster: Ugarit Verlag, 2000), pp. 189–92.

[45] A. Jones, "A Study of Babylonian Observations of Planets Near Normal Stars," *Archive for History of Exact Sciences* 58 (2004), 530–34. See also J. M. Steele, "Celestial Measurement in Babylonian Astronomy," *Annals of Science* 64 (2007), 293–325.

Diaries archive from Babylon (only one survives from the seventh century, one from the sixth century, and four from the fifth century) and astronomical table texts from Babylon and Uruk, the state of astronomical knowledge before the Hellenistic period must be inferred from the evidence of relatively few texts. One such is the tablet known as "Strassmaier Cambyses 400," after its original publication from 1890.[46] This tablet contains astronomical data for the year 523 BCE, or year 7 of Cambyses the Achaemenid ruler. The contents show that already from the early sixth century BCE, Babylonian astronomers had been making and no doubt recording observations of the Lunar Six[47] and of planetary synodic phenomena, and had also begun calculating the latter phenomena on the basis of goal-year periods. Confirmation of the early, i.e., seventh-century, collection of Lunar Sixes is found in tablets devoted solely to these data.[48] Interestingly, among these sources is one from Nippur, possibly for the year 618 BCE.[49] What is clear is that observation and prediction developed hand-in-hand, with prediction by calculation and measurement supplying data not accessible due to weather or other visibility conditions.

PREDICTION AND CALCULATION

Early identification of periodic phenomena such as the duration of the moon's visibility at night and the variation in the length of daylight throughout the year gave rise to quantitative methods for predicting these phenomena based on linear mathematical models. Ultimately, eclipses, equinoxes and solstices, the synodic appearances of the planets, and the date of first visibility of the moon were also predicted by means of similar, though more complex models involving relations between variables, functions to account for their variation, and excellent numerical parameters.

Astronomy as the study of the periodic behavior of astral phenomena and the development of mathematical models for their prediction seem to have arisen together with the attention to the phenomena as signs. Among the earliest examples are the "Venus Tablet of Ammiṣaduqa" (= *Enūma Anu*

[46] J. Strassmaier, *Inschriften von Cambyses* (Babylonische Texte 9; Leipzig: Eduard Pfeiffer, 1890), No. 400; Britton, "Remarks on Strassmaier Cambyses 400."

[47] The Lunar Six phenonena are: at the beginning of the month in the evening, NA is the time interval between sunset and moonset on the evening of the first lunar visibility after conjunction of sun and moon; in the middle of the month, ŠÚ is the interval from moonset to sunrise when the moon set for the last time before sunrise; in the middle of the month, *na* is the interval between sunrise and moonset when the moon set for the first time after sunrise; ME the interval between moonrise to sunset when the moon rose the last time before sunset; GE_6 the interval between sunset and moonrise when the moon rose the first time after sunset; at the end of the month, in the morning KUR is the interval between moonrise and sunrise when the moon was visible for the last time before conjunction.

[48] P. J. Huber and J. M. Steele, "Babylonian Lunar Six Tablets," *SCIAMVS* 8 (2007), 3–36.

[49] Ibid., pp. 15–16, Text B (N.2349).

Enlil Tablet 63) and the lunar visibility tablet *Enūma Anu Enlil* Tablet 14. In addition to its importance for reconstructing the series *Enūma Anu Enlil*, the Venus Tablet continues to play a role in modern astronomical chronology for the ancient Near East.[50] The text, or a part of it, was thought to be based on a source from the reign of King Ammiṣaduqa of Babylon in the mid-seventeenth century BCE, in which the appearances and disappearances of the planet would have been directly observed and recorded (the alleged text is no longer extant). Surviving exemplars, all written during the Neo-Assyrian period or later, demonstrate an awareness that 5 synodic cycles of the appearances of Venus (as evening and morning star, that is, morning rise and set and evening rise and set) occur every 8 years (that is, every 99 Babylonian months minus 4 days). The following is the first line:

> (Year 1) If on the 15th of Month XI (*Šabaṭu*) Venus disappeared in the West (Evening Last), remained invisible 3 days, and reappeared in the East on the 18th day of Month XI: Catastrophes of kings; Adad will bring rains, Ea will bring floods, one king will send greetings to another king.[51]

The omens of *Enūma Anu Enlil* Tablet 63 are constructed from a sequence of synodic phenomena of Venus over a period of 21 years (the length of the reign of Ammiṣaduqa) formulated as conditional statements "If Venus . . .," together with associated events "then . . ." In its extant form, however, the tablet does not preserve a list of Venus observations from the Old Babylonian period, but is a composite text, and includes some computed values for the phenomena and the periods of invisibility that themselves have been copied and corrupted in the manuscript transmission.

The other tablet from *Enūma Anu Enlil* that represents the use of an arithmetic model for the description of an astronomical phenomenon is Tablet 14.[52] It provides a linear zigzag scheme for the length of visibility of the moon each night for the 30 days of the two equinoctial months (when day and night are of equal length as the sun crosses the equator). The interest in duration of lunar visibility is tied to the ominous nature of the moon when visible. The lunar visibility scheme is based on an ideal year of 360 days and a ratio of longest to shortest daylight of 2:1, also a schematic consequence of the mathematical model.

The use of the sexagesimal number notation as well as agreeable sexagesimal numbers (12, 30, 360) is typical of these early astronomical texts.

[50] T. de Jong, "Astronomical Fine-Tuning of the Chronology of the Hammurabi Age," *Journal Ex Oriente Lux* 44 (2012–13), 147–67.

[51] E. Reiner and D. Pingree, *Babylonian Planetary Omens, Part 1: Enūma Anu Enlil Tablet 63: The Venus Tablet of Ammiṣaduqa* (Malibu, CA: Undena Publications, 1975), p. 29.

[52] See Hunger and Pingree, *Astral Sciences*, pp. 44–50.

Babylonian sexagesimal place value notation entered the stream of the Western scientific astronomical tradition, where it continues in the modern practice of measuring degrees of time and arc. Related to the lunar visibility scheme is another zigzag scheme for the duration of daylight throughout the year, attested in the so-called "Astrolabe" or "Three Stars Each," which assigned heliacal risings of prominent stars to the twelve thirty-day months of the ideal year, as well as the astronomical compendium MUL.APIN.[53] In the Astrolabe (and in *Enūma Anu Enlil* Tablet 14), daylight (and lunar visibility) increases and decreases by a constant between two extrema. The phenomena are thereby modeled schematically with a minimum of observational input, probably principally that, at the equinoxes, day and night are each one-half of an entire day, i.e., 3,0 UŠ (time degrees, hence 180 degrees, or 12 hours of an equinoctial day).

David Brown suggested that the ideal schemes were suited to divinatory purposes as they sufficed to indicate whether phenomena were timely or not (expressed as *ina la minâtišu* "not in accordance with its normal number," or *ina la simanišu* "not at its proper time"), and therefore interpretable as of favorable or unfavorable portent.[54] The early astronomical models attest to the interdependence of observational and schematic components, a relationship that lies at the very core of the connection between observation and theory in science generally.

The period between the Neo-Babylonian Empire and the arrival of Alexander the Great (i.e., between ca. 626 and 331 BCE) was a creative one for predictive astronomy in Babylonia. Texts from this intermediate stage can be characterized by a progressive utilization of period relations in the prediction of lunar as well as planetary phenomena. By the Seleucid period, goal-year methods were regularly in use, and the astronomical records, now termed "Goal-Year Texts," were prepared by means of compiling lunar and planetary observations from the Diaries the requisite number of years for one period before the goal year. Thus a Goal-Year Text presents the raw data for predicting the future occurrences of the dates and zodiacal signs for the synodic phenomena of Jupiter, Venus, Mercury, Saturn, and Mars (in that order), and the moon, giving data for the Lunar Sixes as well as eclipses. According to the rubric given to these texts by the scribes, they included data for "the 1st day, appearances, passings, and eclipses which have been established for the year x."[55] Goal-year predictions were then introduced into Diaries when weather prevented observation, as well as into Almanacs and Normal Star Almanacs.[56]

[53] See Hunger and Pingree, *Astral Sciences*, pp. 50–7.
[54] Brown, *Mesopotamian Planetary Astronomy-Astrology*, p. 106.
[55] A. J. Sachs, "A Classification of Babylonian Astronomical Tablets of the Seleucid Period," *Journal of Cuneiform Studies* 2 (1948), 271–90, pp. 284–5.
[56] L. Hollywood and J. M. Steele, "Acronycal Risings in Babylonian Astronomy," *Centaurus* 46 (2004), 145–62, pp. 154–5.

Predictive models, making use of mathematical schemes using either a periodic zigzag function or a periodic step function came to full maturation in Seleucid Babylonia.[57] These are tabulations of the dates, positions, and values instrumental for their calculation. Such tables depended upon the use of the fixed intercalation scheme of the nineteen-year cycle and the Seleucid Era (312 BCE), as well as the establishment, around 400 BCE, of the fixed zodiac of twelve thirty-degree signs.[58] Evidence for the astrological application of such calculated astronomical data is clear in Late Babylonian natal omens that forecast the fortunes of an individual on the basis of planetary synodic phenomena.[59] But Late Babylonian omen astrology was not limited to genethlialogical concerns, as the subjects traditionally of importance to celestial divination (enemy attacks,[60] rise and fall of the market, rains and floods, etc.) are also well attested in omens composed later than and clearly building on the content of *Enūma Anu Enlil*.[61] Specific mention of the periodicity of such terrestrial occurrences shows that the predictive science of astronomy was carried out in concert with an effort to know the periodicities of future occurrences of events of concern to human beings, such as enemy attacks, fall of the economy, ruination of the harvest, and so on:

> [. . .] Pay attention to [. . .], the appearance and last visibility of Venus and Mercury and [. . .] which on the wrong day is not . . . an omen. (If) the rain begins [. . .] from the 1st day to the 15th day and remains: its rain is not good. At the end of the year it is good . . . There will be as many storms now as there were storms in the past. You will predict a swollen flood. [. . .] waters will flow from breaches (in the dykes). (If) a planet appears and (there is) rain and flood: now (there will be) rain and flood.[62]

[57] For a full exposition of the mathematical methods of Systems A and B, see Neugebauer, *Astronomical Cuneiform Texts* and Otto Neugebauer, *A History of Ancient Mathematical Astronomy* 3 vols. (Berlin and New York: Springer Verlag, 1975). See also the chapter by Steele, in this volume.

[58] See J. P. Britton, "Studies in Babylonian Lunar Theory: Part III. The Introduction of the Uniform Zodiac," *Archive for History of Exact Sciences* 64 (2010), 1–47.

[59] TCL 6 14, edited in Sachs, "Babylonian Horoscopes," Appendix II, pp. 65–75. In this text the synodic phenomena mentioned are first appearance (IGI) and last appearance (ŠÚ), otherwise the text refers to the planets rising (È) and setting (ŠÚ).

[60] TCL 6 13 rev. ii 28 gives the rubric "periodic occurrence of the enemy attack" (*adanni tīb nakri*). For the word *adannu* "period," or "periodic cycle," see *The Assyrian Dictionary of the Oriental Institute of the University of Chicago* (Chicago, IL: Oriental Institute, 1956–2010).

[61] For a text concerned with predicting market prices, see H. Hunger, *Spätbabylonische Texte aus Uruk* (Berlin: Mann, 1976), vol. 1, No. 94. For a text that predicts enemy activities (attacks and on which regions, whether or not booty will be taken, etc.), see TCL 6 13, in F. Rochberg, "TCL 6 13: Mixed Traditions in Late Babylonian Astrology," *Zeitschrift für Assyriologie* 77 (1987), 207–28. And for the prediction of rains and floods, see TCL 6 19 and 20, in H. Hunger, "Astrologische Wettervorhersagen," *Zeitschrift für Assyriologie* 66 (1976), 234–260

[62] TCL 6 20 obv. 16'–18' and rev. 9–10, see Hunger, "Astrologische Wettervorhersagen," p. 239.

EXPLANATION

Cuneiform scholarship developed modes of explanation appropriate to the various contexts and norms of its knowledge, and this had principally to do with texts. As divination was a scribal-scholarly endeavor it stands to reason that its development would involve philological techniques consistent with other scribal-scholarly practice. Ominous signs and cuneiform signs were related in the sense that both were to be read and interpreted. Explanation, therefore, took on a number of functions depending on different text types. Variously, explanatory texts could focus on elucidation (of words by means of synonyms), exposition (of a phenomenon by means of description), or instruction (by means of the procedural steps involved in making, calculating, or performing something), all arguably subsumable under the rubric "explanation." Explanation was a vital part of the cuneiform project of knowing and interpreting the world of signs, the correspondences between things, and the meaningful relationships between words and the world. In the form of commentaries and procedures, explanation found its way into other areas of scribal knowledge apart from divination, e.g., into mathematics, medicine, and astronomy.

Numerous reports from the Neo-Assyrian scholars refer to the explanation of omens with the term *pišru*, as in,

> tonight Saturn approached the moon. Saturn is the star of the sun, (and) the relevant interpretation (*pi-še-er-šú*) is as follows: it is good for the king. The sun is the star of the king.[63]

Other explanations in divinatory reports take the form of philological commentary on the words used in the omen text, as in the following, which first quotes an omen and then explains the verb (*ṭerû* "pierce") of its protasis:

> If the moon's right horn at its appearance pierces (*tirāt*) the sky: there will be stable prices in the land; a revolt will be staged in the Westland. "Its right horn pierces the sky," as it says, means it will slip into the sky and will not be seen; DIRI – pronounced dir – is "to slip," said of a horn.[64]

The explanation in the passage is the elucidation of the verb *ṭerû* on analogical-phonological grounds with the verb *halāpu* (written with the logogram DIR) "to slip in or through." The association to *halāpu* is based on the sound of its logographic spelling, whereby /ṭer/ can be associated with /dir/. The explanation of words in this fashion, by phonological

[63] Hunger, *Astrological Reports*, 95 rev. 1–7.
[64] Ibid., 57: 5–rev. 4.

analogy and etymography, is quintessential cuneiform scribal philology, which stemmed ultimately from the translation methods of the compilers of lexical lists.[65]

Astronomical texts have different predictive and explanatory properties compared to divinatory material. Given that they are both parts of the intellectual output of scribes, however, it is not surprising that they share a common intertextual nature. Exemplary of explanatory texts on the astronomical side are collections of procedures to explain rules for computing table texts, thus establishing an intertextual reference to the tables themselves. Also common to most procedural texts, as Ossendrijver pointed out, is the second-person address, as though in or from a dialogue, perhaps the vestige of an earlier rhetorical form.[66] The second-person rhetorical form also occurs in the context of omens, such as:

> Observe his last visibility [on the 28th, variant: 29th of *Kislīmu*], and you will predict an eclipse. The day of last visibility will show you the eclipse.[67]

Using the same rhetorical device, the following procedure for a Jupiter 4-zone System A′ model gives instructions for how to calculate longitudes with a step function, using the characteristic zone (or arc subdivision) boundaries for the planet, designated by zodiacal degrees:

> From 9 Cancer until 9 Scorpius you add 30. (The amount) by which it exceeds 9 Scorpius you [multiply] by 1;7, [30].
>
> From 9 Scorpius until 2 Capricorn you add 33;[45]. (The amount) by which it exceeds 2 Capricorn you multiply by 1;[4].
>
> From 2 Capricorn until 17 Taurus you add 36. (The amount) [by which] it exceeds 17 Taurus you multiply by 0;56,15
>
> From 17 Taurus until 9 «Cancer» [you add] 33;[45]. (The amount) «by which» it exceeds 9 Cancer you multiply by 0;53, 20.[68]

In addition to the second person address, procedures referring to the subdivision of the synodic arc and describing the interval, or distance (ZI = *nishu*), a planet goes from one synodic phenomenon to the next within the total synodic cycle, called a "push," expressed in distance (*birītu*) or days (*ūmū*), simply adopt a third-person descriptive style. Thus, from the same text (Ossendrijver No. 32), the four zones of the model are clarified:

[65] See the exposition of translation methodology in C. Jay Crisostomo, "Bilingual Education and Innovations in Scholarship: The Old Babylonian Word List Izi" (PhD thesis, University of California, Berkeley, CA, 2014).

[66] M. Ossendrijver, *Babylonian Mathematical Astronomy: Procedure Texts* (Berlin: Springer Verlag, 2012), p. 15.

[67] F. Rochberg-Halton, *Aspects of Babylonian Celestial Divination: The Lunar Eclipse Tablets of Enuma Anu Enlil* (AfO Beiheft 22; Horn: Berger and Sons, 1988), p. 180.

[68] Ibid., pp. 288–9, No. 32 lines 1–4.

> From [9 Cancer] until 9 Scorpius the small one. From 9 Scorpius [until 2] Capricorn the middle one. From 2 Capricorn until 1[7] Taurus the large one. From 17 Taurus until 9 Cancer the middle one.[69]

A late seventh century Goal-Year procedure text (BM 45728), perhaps the earliest such explanatory astronomical text known, gives sidereal periods for planetary synodic phenomena and corrections for the dates.[70] The tablet enumerates the periods for the appearances (IGI.DU$_8$.A = *tāmartu*) of the moon, Venus, Mercury, Mars, Saturn, and Sirius in a procedural manner. A selected passage reads:

> [Appearance of] Venus. 8 years [you go back] behind you . . . 4 days you subtract. You observe (it).
>
> [Appearance of] Mercury. Your 6 years you go back behind you . . . to it you add . . . 10 you add to the (date of) appearance. You observe (it).
>
> [Appearance] of Mars 47 years you go back [behind you]; 12 days in addition [. . .] 10 you add to the (date of) appearance and you observe (it).
>
> [App]earance of Saturn, 59 years you go back [behi]nd you. To the day (lit. "day by day"), (it = Saturn) appears (again).[71]

In the segment of the text quoted, the period for observing Venus is given as 8 years – 4 days, Mercury 6 years + 10 days, Mars 47 years + 12 days, and Saturn 59 years "to the day," where the addition or subtraction of days is a feature of similar texts that correct for dates of observations, and the periods for Venus, Mercury, and Mars are those known in Goal-Year Texts.[72]

Later astronomical procedure texts explain the computational methods of the ephemeris tables directly. The verbal idiom of late procedures is partly consistent with that of the divinatory tradition, for example in the phrase used to tell someone to take something into consideration, literally, "hold x in your hand" (*ina qātika tukâl*), or perhaps "bear in mind." The contexts in which the phrase is employed can refer to a wind (or direction), times, positions, or a goal-year. The following shows its usage in *Enūma Anu Enlil* Tablet 20: "You observe his (the god's) eclipse and bear in mind the north

[69] Ibid., No. 32 lines 5–6.

[70] J. P. Britton, "Treatments of Annual Phenomena in Cuneiform Sources," in J. M. Steele and A. Imhausen (eds.), *Under One Sky: Astronomy and Mathematics in the Ancient Near East* (Münster: Ugarit Verlag, 2002), pp. 21–78, Appendix A, pp. 59–61.

[71] BM 45728: 5–14. Cf. translation in Ossendrijver, *Babylonian Mathematical Astronomy*, p. 23.

[72] J. M. Steele, "Goal-Year Periods and their Use in Predicting Planetary Phenomena," in G. J. Selz with the collaboration of K. Wagensonner (eds.), *The Empirical Dimension of Ancient Near Eastern Studies/Die empirische Dimension altorientalischer Forschung* (Vienna and Berlin: LIT Verlag, 2011), pp. 101–10, p. 105.

(IM.SI.SÁ ina ŠU-ka tu-kal)."[73] In the late astronomical procedures, times and positions are to be "held in one's hand (or hands)," as in Ossendrijver's Procedure Text No. 46 (= ACT 812) rev. ii 1–2: "The day when Venus appears in the west (EF) or sets in the east (ML): you hold the times and positions for the *igigubbû*-coefficients [. . .] in your hands."[74]

Another feature shared with the divinatory corpus was the collection of explanatory texts into series. Some explanatory texts, such as the "commentary" (*mukallimtu*) or "explanatory word list" (*ṣâtu*), constituted multi-tablet series of their own, and a few astronomical procedure texts appear to similarly belong to series.[75] A set of procedures referring to Saturn, giving a number of procedures explaining Systems A, B, and B" for Saturn, has a colophon with the catchline for the next tablet in the series, with an incipit "the displacement (or progress in longitude) of Mars."[76] Similar to the interpretative material prepared and collected for intuitive predictive texts (omens), the mathematical tables too had an interpretative corpus prepared and collected as procedures.

An exploration of the context of explanation must recognize the importance of schemes (both divinatory and numerical), models, and analogies (both divinatory and numerical). As divination and astronomy both had predictive aims, explanation was embedded in the predictive undertaking. Each branch of cuneiform predictive knowledge was tied to programs of observation and a tradition for recording and dating those observations. From our point of view, observation, prediction, and explanation belong within the purview of science. Although the methods of investigation into the world of perception and experience, and ideas of what was usual, unusual, regular, irregular, normative, and anomalous were determined by the particular phenomena – mostly ominous phenomena – that interested the Assyrian and Babylonian scholars over time, in terms of the empirical, predictive, and explanatory dimensions of cuneiform knowledge, particularly in its persistent attempts to grasp an order of things and to resolve what is anomalous into a system, it is hardly possible not to see the features of its kin in the later history of science.

MODERN NOMENCLATURE OF BABYLONIAN ASTRONOMICAL/ASTROLOGICAL TEXTS

Since Neugebauer's pioneering work first to publish the cuneiform ephemerides and procedure texts,[77] and then to incorporate this material into

[73] Rochberg-Halton, *Aspects*, p. 241.
[74] Ossendrijver, *Babylonian Mathematical Astronomy*, p. 329, No. 46 rev ii 1–2.
[75] Ibid., p. 12.
[76] Ibid., p. 310, No. 42 rev. 7.
[77] Neugebauer, *Astronomical Cuneiform Texts*.

a treatment of ancient mathematical astronomy alongside Egyptian, pre-Ptolemaic. and Ptolemaic Greek astronomy,[78] cuneiform astronomical texts have attained a pride of place in the history of ancient science for their quantitative and predictive nature, their use of period relations for lunar and planetary synodic phenomena, their use of number-theoretic functions for construction of predictive models for the calculation of these periodic lunar and planetary synodic phenomena, and the direct influence they had on Greek astronomy.

An effective classification scheme for the Babylonian astronomical texts was introduced in 1948 by Abraham Sachs in what is by now a classic paper.[79] There, Sachs divided the corpus of Seleucid astronomical tablets into tabular and non-tabular texts. The "Astronomical Tables" included planetary and lunar ephemerides, which were recognized as intimately related to procedural texts.[80] Sachs' study and classification of the astronomical texts began with the problem of distinguishing between observations and predictions, and he stated that "a somewhat startling result is that all classes of Seleucid astronomical texts contain at least some predictions."[81] It is in the methods of prediction that the texts are to be differentiated, in the tabular texts by the use of linear functions based on period relations and in the non-tabular texts by a simpler use of appropriate (synodic or sidereal) periods (goal-year periods) for the astronomical body in question.

In 1980 Asger Aaboe revisited the question of the relation of observation and prediction.[82] His goal was to explain how the excellent mathematical schemes of the Seleucid Babylonian ephemerides could have been derived from what he called the "crudity" of observational data in the texts he termed "non-mathematical astronomical texts," i.e., in the very same set of sources classified as "non-tabular" by Sachs, principally, the Diaries. The "non-mathematical astronomical texts" were to be differentiated from the texts in Neugebauer's ACT volume on the basis that they were largely observational, whence the distinction of mathematical versus non-mathematical astronomical texts. Aaboe's distinction rested on the difference between the singular and theoretical nature of the ephemerides' mathematical schemes on the one hand and the (non-mathematical) largely observational data of the Diaries on the other. His question was: how might the Diaries have served as a possible source for the development of the mathematical schemes of the ephemerides?

Sachs' terminology, however, referred to format not content, and emphasized the fact that both tabular and non-tabular texts contain predicted data. Thus it is not the case that only the ACT texts are quantitative and

[78] Neugebauer, *A History of Ancient Mathematical Astronomy*.
[79] Sachs, "A Classification of Babylonian Astronomical Tablets."
[80] Ibid., p. 273.
[81] Ibid., p. 271.
[82] Asger Aaboe, "Observation and Theory in Babylonian Astronomy," *Centaurus* 24 (1980), 14–35.

predictive. Their distinguishing feature is the highly developed use of functions (zigzag and step) in the form of difference sequences to generate positions in the zodiac or dates in the year (using months and a mathematical unit 1/30 of a month, termed "tithi" after its Sanskrit coinage, and fractions of tithis).[83] The non-tabular astronomical texts (i.e., Aaboe's non-mathematical astronomical texts, also referred to by Neugebauer as GADEx texts[84]), though comprised largely of observations, also contain predicted data derived from the use of numerical parameters, based in turn on a knowledge of planetary and lunar periods.

A distinction between mathematical and non-mathematical for cuneiform astronomical texts may well be useful in that the numbers in the tabular texts (ephemerides) reveal the quantitative methods (zigzag and step functions), predictive goals (dates and positions of synodic phenomena of the moon and planets), and theoretical quantities (mean synodic arc or mean synodic time) or constructs (the mean sun) associated with and exclusive to mathematical (or tabular) astronomical texts. That distinction, however, should not obscure the broader relationships among astronomical, astrological, as well as medical, magical, and divinatory texts. Intertextual connections, both direct (references in Diaries to ominous events) and indirect (calculations for planetary positions given in horoscopes), point to a more integrated culture of knowledge.

Observational and predictive, as classificatory descriptors, therefore apply to two closely related approaches to astral knowledge that not only complemented one another from the beginning of the history of Babylonian astronomy but had some relation to divination as well. Early Babylonian astronomy aimed to schematize and model phenomena consistent with ominous signs, such as the appearances and disappearances of Venus formulated as omens in *Enūma Anu Enlil* 63. Later, in Goal-Year Texts, Diaries, and Almanacs, the phenomena that were predicted (to supplement those observed) were the synodic appearances of the planets, their passages by Normal Stars, Lunar Threes, Lunar Sixes, and eclipses. In the ephemerides, synodic phenomena of the moon and planets were calculated by means of mathematical models based on period relations and the use of zigzag or step functions. It is important to note that these phenomena do not appear as such in omen texts.

Terms associated otherwise exclusively with astronomical texts, such as "latitude" (NIM "high" and SIG "low") and the Lunar Three (length of the previous month, expressed as either 30 or 1, mid-month *na*, i.e., the *na*

[83] Of course, foreshadowing of these methods can be seen earlier, though at a much cruder level, for example in Enūma Anu Enlil 14 (see note 61) and the Astrolabe Daylight scheme, both of which employ zigzag functions. A later extension of this method can also be seen in a group of *ziqpu* star texts that deal with the problem of zodiacal rising times; see Rochberg, *In the Path of the Moon*, pp. 271–302.

[84] See Neugebauer, *A History of Ancient Mathematical Astronomy*, p. 351.

opposite the sun, and KUR, the time between the last visible moonrise before conjunction and sunrise) do, however, appear in Late Babylonian astrological contexts, e.g., horoscopes. In the case of latitude, the horoscopes refer only to lunar latitude,[85] but a late Uruk omen text clearly deals with the relevance of planetary latitude for the forecasting of market prices, as follows:

> If Jupiter is faint, or attains minimum latitude, or disappears, and Mars is bright or attains maximum latitude, or Mars and Jupiter are in conjunction: Business will greatly decrease and the people will experience severe famine.[86]

What emerges is that the earliest selective attention to phenomena, especially but not exclusively of the heavens, comprised a number of separate but interdependent strands of knowledge and practice – divination, astronomy, and astrology – each of which took complex forms and utilized a variety of methods. Each played a role in the scribes' engagement with the world over the millennia of their tradition. That engagement centered on the search for and understanding of order and anomaly within phenomena, both in order to observe and interpret phenomena as signs and to predict those that were periodic.

[85] Rochberg, *Babylonian Horoscopes*, pp. 42–3.
[86] Hunger, *Spätbabylonische Texte aus Uruk*, vol. 1, Text 94: 12–14.

2

BABYLONIAN MEDICINE AS A DISCIPLINE

Markham J. Geller[1]

Herodotus famously said that he found among the wisest of Babylonian customs the following:

> They have no doctors, but bring their invalids out into the street, where anyone who comes along offers the sufferer advice on his complaint, either from personal experience or observations of a similar complaint in others. Anyone will stop by the sick man's side and suggest remedies which he has himself proved successful in whatever the trouble may be, or which he has known to succeed with other people. Nobody is allowed to pass a sick person in silence but everyone must ask him what is the matter.[2]

This passage is important since it more-or-less defined Babylonian medicine among medical historians for the next two millennia.[3] At first one tends to discount this statement as whimsical and transparently false when compared with the very large quantity and quality of Akkadian medical texts from Babylon and elsewhere in Mesopotamia, attesting to professional medical practices of all kinds. However, since Herodotus (485–425 BCE) lived a generation before Hippocrates (460–370 BCE), he would know little of "scientific" Greek medicine, but obviously there must have been something in this account of unusual interest, which was either very different or very similar to healing practices in his contemporary Greek world. One striking

[1] Work on this article was carried out under the auspices of the ERC Advanced Grant Project BabMed, and an earlier version was presented at the Rencontre Assyriologique Internationale 61 in Bern, 25th June 2015. Comments on this paper were generously offered by Philip van der Eijk and the ERC Project BabMed team colleagues (Cale Johnson, Ulrike Steinert, and Strahil Panayotov), and the approach taken here was strongly influenced by the work of BabMed.

[2] Herodotus I 197, translated by A. de Sélincourt (Harmondsworth: Penguin, 1965), 94.

[3] For the first Assyriologist to assess the accuracy of this passage, see G. Contenau, *La médecine en Assyrie et en Babylonie* (Paris: Maloine, 1938), in which he concludes that Herodotus refers to healing in rural contexts, far from centers of medical learning, and that Herodotus was badly informed.

feature of this Herodotus passage is that the patient is carried into the "street" for diagnosis[4], and more troubling is that those suggesting remedies are laymen, not the *asû*, a professional Babylonian physician, and it is far from clear why one could not withhold an opinion about a patient's health. So how much does Herodotus actually reflect state-of-the-art Babylonian medicine?

To a certain extent, Herodotus' comment coincides with documentary evidence from Babylonia, in which the evidence for the *asû*-physician almost completely dries up, judging from colophons of tablets and administrative records,[5] but this information may be misleading, since the physician as a private entrepreneur (as in Greece) would not normally have appeared in either palace or temple records,[6] nor would doctors have earned sufficient sums of money for their transactions to have been recorded. There is little doubt, however, about the downgrading of the medical profession in late periods in comparison with the heady days of eighteenth-century BCE Babylon, in which the *asû*-physician was the only medical practitioner featured in Hammurabi's famous Law Code.[7] Hammurabi's Code even specified how much the *asû*-physician could charge for his services, depending upon the status and relative wealth of the client. In fact, the Code is the best surviving evidence for the fact that the *asû*-physician indulged in surgery, although this in no way excludes other types of therapy, since the Code only intended to highlight obvious high-profile cases of malpractice against the *asû*;[8] other kinds of treatments were not relevant to the legal questions being regulated.

4 In fact, both Babylonian sources and the Hippocratic oath assume that the doctor visits the ill patient in his home, in order to make a prognosis or diagnosis. The incipit of the Babylonian Diagnostic Handbook stipulates, "When the KA.PIRIG-exorcist goes to the patient at home," while the Hippocratic Oath stipulates that the doctor, whenever he visits the patient at home, must do him no harm.

5 See P. Clancier, "Teaching and Learning Medicine and Exorcism at Uruk During the Hellenistic Period," in Alain Bernard and Christine Proust (eds.), *Scientific Sources and Teaching Contexts Throughout History: Problems and Perspectives* (Boston Studies in the Philosophy and History of Science, 301; Berlin: Springer, 2014), pp. 41–66, p. 6, stating that "the *asû* had disappeared by the end of the first millennium at least from the written documentation." Nevertheless, it is unlikely that the profession had disappeared entirely, since the Aramaic cognate term '*sy*' remained the normal term for "healer, physician," and late documentary evidence can be misleading, because the *asû* no longer had any official status. By the time Herodotus comments on Babylonian medicine many new changes had occurred with professional relationships. The traditional divining by examining entrails was virtually replaced by the new science of astrology, which also gave birth to new disciplines of astral magic and medicine (see E. Reiner, *Astral Magic* (Philadelphia, PA: American Philosophical Society, 1995)). Babylonian temples were administered by Persian *magoi*, whose prestige probably served as a model for the *mašmaššu*-exorcist.

6 Royal correspondence from palace records makes frequent reference to both physicians and exorcists among professions serving the royal court, but such data is not representative of society as a whole, and in any case, after Persian rule such documentation ceased to exist.

7 See Contenau, *La médecine*, pp. 30–5, and P. Attinger, "La médecine mésopotamienne," *Le Journal des Médecines Cunéiformes* 11–12 (2008), 1–96, 3, 50–1.

8 Stefan Maul translates *asû* as *Wundarzt* (S. Maul, *Die Wahrsagekunst im Alten Orient* (Munich: C. H. Beck, 2013), p. 278). See also I. L. Finkel, *The Ark before Noah: Decoding the Story of the Flood* (London: Hodder and Stoughton, 2014), p. 45.

Nevertheless, the *asû* was by no means the only professional healer in Hammurabi's Babylon, since rich collections of incantations and healing rituals attest to the activities of an *āšipu*, a priest who functioned as an exorcist.[9] The further back one goes into records of medical and magical texts from Mesopotamia, the distinction between medicine and magic appears to be more clearly delineated; medical texts do not appear to employ incantations, and exorcistic texts are clearly non-medical compositions in format and content. By the end of the second millennium BCE, however, the roles of professional healers had altered substantially, and it appears that the exorcist, now usually known as a *mašmaššu*,[10] had become well-versed in medical literature, and the previously clear boundary between the two professions was becoming blurred.

This causes numerous difficulties for modern studies of Babylonian medicine. According to one widely accepted reconstruction, the exorcist (*mašmaššu* or *āšipu*) was interested in the *reasons* for illness, usually attributed to the patient's immoral behavior, divine anger, or witchcraft, which unleashed demons or natural causes, and his treatments were essentially incantations and rituals. The *asû*-physician, on the other hand, was a pragmatic healer who was committed to treating symptoms (with drugs, etc.) rather than in determining what motivated the illness on a higher plane.[11] The weakness of this approach is the assumption that the exorcist exclusively operated within a theoretical framework, based on a magic- and divinity-orientated explanation for causes of illness and therapy, while the physician's remedies essentially had no theoretical framework beyond trial-and-error and observation.[12] A counterargument proposed by

[9] See now D. Schwemer, "The Ancient Near East," in D. J. Collins (ed.), *The Cambridge History of Magic and Witchcraft in the West from Antiquity to the Present* (Cambridge: Cambridge University Press, 2015), pp. 17–51.

[10] Although modern scholarship prefers the term *āšipu* for "exorcist," in ancient records the term *mašmaššu* for this profession predominates (e.g. in colophons), most often with the Sumerian logogram MAŠ.MAŠ, which the dictionaries assume to be rendered as *āšipu*, although the term *mašmaššu* is well attested. In Assur, for instance, the title *mašmaššu* for exorcist was standard (see S. Maul, "Tontafelbibliothek aus dem Haus des Beschwörungspriesters," in S. Maul and N. Heeßel (eds.), *Assur-Forschungen* (Wiesbaden: Harrassowitz, 2010), pp. 189–228). The difference between these two terms eludes us, but it is possible that nuances reflect the professional role of this healer, both as a priest responsible for many different kinds of temple and purification rituals, as well as hemerologies, and at the same time serving as a specialist healer working with incantations and healing rituals. For this reason, we will employ the term *mašmaššu* for the profession and *āšipūtu* for the practice of incantatory-ritual healing. Nevertheless, the term *mašmaššūtu* was also employed, and this may refer as well to the more general duties within the priesthood, as well as the professional healing qualifications.

[11] See N. Heeßel, *Babylonisch-assyrische Diagnostik* (Münster: Ugarit Verlag, 2000), p. 95, and for the latest attempt to explain the complex relationship between *āšipūtu* and *asûtu*, see now Schwemer, "The Ancient Near East," p. 27.

[12] Edith Ritter proposes two postulates to distinguish between the methods of *āšipūtu*, which can be summarized as follows: *āšipūtu* assumes that disease results from indiscernible "malignant influences" or transgression, which the exorcist can diagnose and make a prognosis based upon duration, crises, remission, and recurrence of symptoms. The *asû*-physician meanwhile views disease as complex symptoms but without assigning supernatural causes, often avoiding diagnosing or even

J. Scurlock assumes that because the majority of known medical tablets known were owned by exorcists,[13] this allows us to alter the basic picture known from numerous other sources and consider the *mašmaššu*-exorcist to be the actual physician, while the *asû* simply functioned as an apothecary.[14] The potential for confusion is predictable. A. Leo Oppenheim preferred to harmonize any contradictions by assuming that all healing in Mesopotamia was a complementary mixture of both magic and medicine, and hence modern scholarship should simply accept this fact and assume that magic and medicine are not distinguishable entities within Mesopotamian healing practices:

> One further characteristic of our texts may be pointed out here: the occasional admixture of magical elements, such as short conjurations, the preference for magic numbers, the request for the special timing of certain tasks, their performance by special agents, etc. Such references must be judged in relation to others which clearly indicate that the Mesopotamians were fully aware of the effectiveness of both of the two media, medicine and magic, separately. Under certain circumstances, they did not mind using both together, and we should neither attempt to explain away nor overstress such practices.[15]

The weakness of all these approaches is that they fail to grasp that there are two competing disciplines within Babylonian medicine, namely *āšipūtu* versus *asûtu*, i.e. incantation-ritual versus recipe-based healing, each with its own distinctive system of theory and practice.[16] Furthermore, the fact the *mašmaššu*-exorcist owned, studied, or copied medical recipes does not change this basic paradigm. J. Scurlock sees the practice of medicine as a 'team effort' between the *asû* and *mašmaššu*,[17] and she posits a collaborative and collegial cooperation between two healing professions in which

labeling the condition (E. Ritter, "Magical Expert (= Āšipu) and Physician (= Asû): Notes on Two Complementary Professions in Babylonian Medicine," in H. G. Güterbock and T. Jacobsen (eds.), *Studies in Honor of Benno Landsberger on his Seventy-fifth Birthday, April 21, 1963* (Chicago, IL: University of Chicago Press, 1965), pp. 299–321, 301–2, although these ideas probably came from Benno Landsberger).

13 See Maul, "Tontafelbibliothek," pp. 212–14 for Assur and Clancier, "Teaching and Learning," pp. 8–10 for Uruk examples of medical tablets owned by exorcists.

14 See J. Scurlock, "Physician, Exorcist, Conjurer, Magician: A Tale of Two Healing Professions," in T. Abusch and K. van der Toorn (eds.), *Mesopotamian Magic* (Groningen: Styx, 1999), p. 77. In fact, in J. Scurlock, *Sourcebook for Ancient Mesopotamian Medicine* (Atlanta, GA: SBL Press, 2014), the author consistently employs the misleading translation "physician" for *mašmašsu*.

15 A. L. Oppenheim, "Mesopotamian Medicine," *Bulletin of the History of Medicine* 36 (1962), 97–108, 103.

16 See G. Lloyd, *Disciplines in the Making* (Oxford: Oxford University Press, 2009), p. 3, where he defines disciplines as "more or less systematic inquiries in different societies and in different periods"; Lloyd views ancient "medicine" as a discipline, but without considering the sister disciplines of magic and divination as relevant to healing arts (ibid., pp. 76–92).

17 Scurlock, *Sourcebook*, p. 3.

diagnosis (*āšipūtu*) was seamlessly combined with treatment (*asûtu*). In fact, these two professions were likely to have been worlds apart and may have had little to do with each other, apart from both professions being represented at the royal court. The *mašmaššu* was a priest who enjoyed all of the benefits of a steady temple income and an elevated social status associated with the priesthood, while in later periods the *asû* was a private entrepreneur, subject to the whims of the free market and competition.[18]

It is clear from all previous studies that trying to get to the bottom of Babylonian medical practice by the "who does what to whom?" route will not get us very far; our awareness of how and under what conditions incantations and prescriptions were disseminated and received is extremely limited, since the bulk of our medical and magical compendia represent library or archival copies of reference tablets, probably for scholastic use, rather than actual applications used in everyday situations. For this reason, we have chosen to examine the *science* – rather than the *sociology* – of medical and magical healing within these frameworks, since the genres of *āšipūtu* (the art of healing through incantations/rituals) and *asûtu* (the art of healing through recipes) have very distinctive characteristics, independent of who is employing these genres and how they are used.[19]

What is required is a new approach to the relationship between these two professional disciplines, which involves the idea of *incursion* between genres. The basis for this approach is that each discipline maintains its integrity on a theoretical assumption that originally incantations, rituals, and even prognoses formally belong to *āšipūtu*, while medical recipes and prescriptions formally belong to *asûtu*. However, in both cases, over time, new needs

[18] The same could be said for the Greek *iatros*:

> The Hippocratic physician is a craftsman. As a craftsman, he practices either as a resident or as an itinerant; he may also settle for a while in some town, leave again, work in another town, or wander all over the country. When he is in a town, he works in his shop or in his patient's home. The shop is a place in which today one person, tomorrow another, plies his trade – not a hospital, not a consulting room in the physician's house. Sick people come to the shop for examination and treatment, or the physician goes to his patient's home. (L. Edelstein, *Ancient Medicine* (Baltimore, MD and London: Johns Hopkins University Press, 1987), p. 87)

Edelstein goes on to explain that the physician is like other craftsman and "holds a low position in society"; meetings between physician and patient were hardly private consultations, since he would be accompanied by his own students and an audience would be present, and the physician would be competing with other physicians for clientele (ibid., pp. 88–9). All of these statements would apply equally well to the Babylonian *asû*.

[19] The fact that so many more genres belonged to the exorcist's training is not surprising, since his own activities far surpassed those of the physician, who was only responsible for composing and maybe administering therapeutic recipes. The *mašmaššu*/*āšipu*, by contrast, had numerous jobs within the temple for various purification rituals and a large repertoire of incantations for removing the ill effects of omens (namburbi-rituals; see S. Maul, *Zukunftsbewältung, eine Untersuchung altorientalischen Denkens anhand der babylonisch-assyrische Löserrituale (Namburbi)*, (Mainz: von Zabern, 1994)), or witchcraft in all its many forms (D. Schwemer, *Abwehrzauber und Behexung* (Wiesbaden: Harrassowitz, 2007)), or the results of a violated oath or taboo (see K. van der Toorn, *Sin and Sanction in Ancient Israel and Mesopotamia* (Assen: van Gorcum, 1985), pp. 50–5).

arose for both of these disciplines resulting in important changes, perhaps introduced at the scholastic level of the academy, but reflecting actual practices. The evolved perception for *asûtu* was that technical prescriptions were not sufficient to convince the patient that the illness could be treated successfully; *asûtu* needed to "borrow" some techniques (such as incantations) from *āšipūtu* to improve the reception of medical preparations.[20] The discipline of *āšipūtu*, for its part, needed to extend its repertoire of techniques to handle certain kinds of ailments (e.g. epilepsy, strokes, and seizures) which were usually defined through magic as being directly caused by angry gods or demons, etc.[21] So while mutual needs allowed for physicians to use incantations and for exorcists to use medical-like recipes, the genres themselves remained unchanged: magic remained magic and medicine remained medicine.[22]

In fact, the problem of defining Babylonian medicine within the confusing array of incantations, rituals, and recipes available to practitioners is not a new problem but was already confronted in antiquity by a famous Babylonian scholar, Esagil-kīn-apli, who was attributed with making order out of the chaos of medical and magical texts of various kinds; in fact, it may have been this scholar (or his school) who was responsible for the first canonical recensions of numerous magical and medical texts.[23] Like the later Hippocrates, Esagil-kīn-apli was the scion of a famous family of healers, with his direct forebear being credited with having been *mašmaššu* to Hammurabi himself. Also like Hippocrates, Esagil-kīn-apli altered the

[20] This is a trend which is already obvious in Egyptian medicine from the mid-second millennium BCE, such as in Papyrus Ebers, which includes incantations amongst the medical recipes. But Ebers is still considered to be a medical and not a magical papyrus (see W. Westendorf, *Handbuch der altägyptischen Medizin*, 2 vols. (Leiden, Boston, MA, and Cologne: Brill, 1999), vol. 1, 92)

[21] There is rare documentary evidence dating to the eleventh century BCE (Middle Assyrian period) from a *mašmaššu*-exorcist acknowledging the receipt of *materia medica* for a salve, indicating that medical recipes were available to a non-physician; see W. Farber and H. Freidank, "Zwei medinische Texte aus Assur," *Altorientalische Forschungen* 5 (1977), 255–8 (ref. courtesy S. Panayotov).

[22] Jean Bottéro preferred to see the state of the healing arts as "infiltrations et contamination" of magic into rational medicine, resulting in what he terms "la médicine exorcistique" which eventually drove out "la médicine empirique et rationnelle" (J. Bottéro, "La médecine de l'ancienne Mésopotamie," *Bulletin du Centre d'Etude d'Histoire de la Médecine* 16 (1996), 4–17, 14–15). The problem with this approach is that it fails to recognize the integrity of *asûtu* and *āšipūtu* as separate discrete disciplines and the fact that both borrowed features from the other. The fact that medical recipes used incantations did not reflect a 'contamination' but rather a broadening of the techniques employed by *asûtu*, just as the anti-witchcraft corpus employed medical recipes as well as incantations and rituals. There are also numerous examples of amulets being employed within medical recipes, see W. Farber, "ina KUŠ.DÙ.DÙ(.BI) = ina maški tašappi," *Zeitschrift für Assyriologie* 63 (1973), 59–68.

[23] The redacting of texts continued long after Esagil-kīn-apli's lifetime. One unusual colophon of a tablet in Assurbanipal's Library (seventh century BCE) has the king boasting (on behalf of his scribes) that he created a new edition of the canonical plant lists by putting previous recensions into a proper sequence and removing redundancies; see B. Böck, "Sourcing, Organizing, and Administering Medicinal Ingredients," in K. Radner and E. Robson (eds.), *The Oxford Handbook of Cuneiform Culture* (Oxford: Oxford University Press, 2011), pp. 690–705, 692–3. This colophon dates roughly from the same period as catalog tablets ascribed to Esagil-kīn-apli's school, showing that there was a serious interest in this period in producing new editions of works relevant to medicine.

direction of medicine in his own day by establishing a corpus of medical and magical texts which would eventually become a standard curriculum for healing practices, based on a theoretical foundation that differentiated between medical prescriptions, incantations and rituals, and prognosis/diagnosis. Like Hippocrates, the name of Esagil-kīn-apli became associated with a corpus of texts, although we have no idea to what extent he either composed or edited the texts in this corpus; the list of works comprising this Esagil-kīn-apli Corpus is only known to us from catalogs compiled several centuries after he flourished (yet another similarity with the Hippocratic Corpus).[24]

The evidence for these three distinctive genres or curricula from the Esagil-kīn-apli-school is concrete and identifiable, since we are lucky to have these three surviving catalog tablets listing the incipits or titles of all relevant texts within each genre. Each tablet lists numerous works, which probably constituted the training of specialists in *asûtu* and *āšipūtu*, with the latter discipline including both "exorcistic" and "mantic" texts relevant to healing arts.[25] The discipline of *asûtu*, on the other hand, concentrated on medical prescriptions, but as we will soon see, the borders between these disciplines were not hermetically sealed, and overlapping grey areas of shared interests and techniques were common features of all three genres of texts.[26] There are three standard models of texts relevant to Babylonian medicine:

[24] See J. Jouanna, *Hippocrates* (Baltimore, MD: Johns Hopkins University Press, 1999), pp. 63–5, crediting the first century CE scholar Erotian with the earliest list of titles of works in the Hippocratic Corpus, to which other works were added; Erotian also produced a glossary based on Hippocratic works. Galen had his own list of Hippocratic works, upon which he based his own commentaries (ibid., pp. 355–6), although the earliest commentaries on the Corpus were already being produced in Alexandria in the third century BCE (ibid., p. 348). By this period, the lexical studies of Bacchius on the Hippocratic Corpus had already produced an early list of eighteen works ascribed to Hippocrates; see H. von Staden, "Lexicography in the Third Century BC: Bacchius of Tanagra, Erotian, and Hippocrates," in J. J. López Férez (ed.), *Tratados Hipocráticos* (Madrid: Universidad Nacional de Educación a Distancia (España), 1992), pp. 549–69, 563. These extensive lists of Hippocratic works and commentaries on these same works reflect the same kind of scholastic activity taking place in Babylonian academies and on the same genre of texts.

[25] Paul Unschuld has argued for a distinction between "Heilkunde" and "Medizin" within Chinese medicine, with the distinction being that the former term refers to the attribution of illness to gods and demons, which eventually gave way to a more technical secular "medicine" based on natural causes. See P. Unschuld, *Was ist Medizin? Westlich und östlich Wege der Heilkunst* (Munich: C. H. Beck, 2003), pp. 82–3, 92–3. This is a distinction which unfortunately cannot be applied to Babylonian medicine, which maintained parallel traditions of magic (diseases attributed to gods and demons) and a more secular tradition of medical recipes.

[26] Within Greek medicine, listings of the Hippocratic Corpus served as useful guides to distinguishing this group of texts from rival authors (such as Diocles or Herophilus), whose works were not included within the Corpus and only survive in fragments; for these two authors, see P. van der Eijk, *Diocles of Carystus, a Collection of the Fragments with Translation and Commentary* (Leiden, Boston, MA, and Cologne: Brill, 2000) and H. von Staden, *Herophilus, the Art of Medicine in Early Alexandria* (Cambridge: Cambridge University Press, 1998), as well as A. Krug, *Heilkunst und Heilkult* (Munich: C. H. Beck, 1985), p. 59–63. In fact, to a certain extent, the Hippocratic Corpus protected the works assigned to it, since those within the Corpus fared better in being preserved intact. According to Schironi, the works collected into a canon in Alexandria were those that were "suggested" or even "compulsory" reading in the curriculum, and although some works outside a canon were preserved and others within a canon were lost, the general picture is that canonized

	medical models[27]	exorcistic models[28]	mantic models[29]
Genre	*asûtu*	*āšipūtu*	*āšipūtu*
Profession	*asû*	*mašmaššu*	LÚ.KA.PIRIG

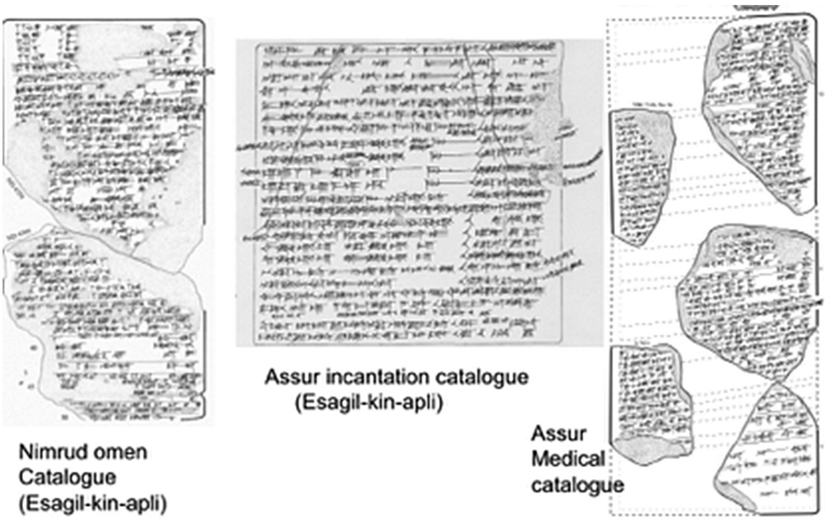

Figure 2.1. Catalog tablets from the Esagil-kīn-apli school.

works were usually those commented upon as "classic" texts (see F. Schironi, "Greek Commentaries," *Dead Sea Discoveries* 19 (2012), 399–441, 439–40). The Esagil-kīn-apli Corpus operated in a similar fashion, especially with incantations, since canonical incantations formed into a standard series had a much better survival rate in the first millennium BCE than did non-canonical incantations.

[27] The Assur Medical Catalogue is now published in a full edition in U. Steinert (ed.), *Assyrian and Babylonian Text Catalogues: Magic, Medicine, and Divination* (Boston, MA: de Gruyter, 2018), pp. 203–91. See also E. Frahm, *Babylonian and Assyrian Text Commentaries: Origins of Interpretation* (Münster: Ugarit Verlag, 2011), pp. 328–9, for a discussion of Esagil-kīn-apli's role as responsible for the Assur medical catalog.

[28] This catalog of incantations (known as KAR 44) lists some texts which pertain to general priestly functions and rituals, but the overwhelming majority of entries relate to incantations and healing arts. This text is re-edited within the BabMed Project, see the sources cited in the previous footnote.

[29] Cf. D. Wiseman and J. Black, *Literary Texts from the Temple of Nabû, Cuneiform Texts from Nimrud* 4/I. L. Finkel, "Adad-apla-iddina, Esagil-kin-apli, and the Series SA.GIG," in Leichty et al. (eds.), *A Scientific Humanist*, pp. 143–59, as well as J. V. Kinnier Wilson, "The Nimrud Catalogue of Medical and Physiognomical Omina," *Iraq* 24 (1962), 52–62. These are catalogs of diagnostic omens (the Diagnostic Handbook; see R. Labat, *Traité akkadien de diagnostics et pronostics médicaux* (Leiden: Brill, 1951), Heeßel, *Babylonisch-assyrische Diagnostik*, and Scurlock, *Sourcebook*, pp. 13–271) and physiognomic omens (see B. Böck, *Die babylonisch-assyrische Morphoskopie* (Vienna: Berger and Söhne, 2000)).

At first glance, this tripartite scheme appears to offer three clearly demarcated areas of practical applications of Babylonian *Heilkunde*, namely recipes, incantations, and prognostic/diagnostic omens, often organized in head-to-foot arrangements. These listings are referred to as "new editions" (lit. "new weavings") of the texts listed, which could either imply new editions of individual compositions or a new arrangement of texts into these three specific genres.[30] All the catalogs have certain features in common, but can briefly be summarized: in all three cases, the listings are organized into a two-part structure, with the second parts containing more advanced or more theoretical kinds of texts, or at least more surprises, such as magical texts within the medical catalog, or medical texts within the incantation catalog; these usually represent diseases which fall into grey areas between magic and medicine,[31] such as seizure, stroke, and epilepsy (although not in the modern clinical usage of these terms).[32] In the same vein, the catalog of "mantic" incipits consists of diagnostic and prognostic omens (i.e. symptoms, listed in the so-called Diagnostic Handbook), while the second part of this catalog consists of physiognomic omens, with a more complex logical structure.[33] Furthermore, although "mantic" works clearly represented a sub-branch of *āšipūtu*, the diagnostic omens were assigned in

30 Cf. M. Stol, in M. Roth, W. Farber, M. Stolper, and P. von Bechtolsheim (eds.), *Studies Presented to Robert D. Biggs* (Chicago, IL: Oriental Institute, 2007), pp. 233–42, 241–2, which makes the sensible comparison with the word "text" as derived from Latin *textus*, "woven, textile." The unusual Assurbanipal Library colophon mentioned above employs the same terminology as the catalogs referring to a new edition of the canonical plant list.

31 A recipe dealing exclusively with "stroke" ends with an interesting clause: "in the first month, second month, third month you must [keep performing] *āšipūtu* [and he will get better]" (F. Köcher, *Babylonisch-assyrische Medizin in Texten und Untersuchungen* (1–6) (henceforth "BAM"), 523 iii 8', the text has not been edited). The text mentions employing *āšipūtu* in dealing with this condition, probably because no drugs were available which could influence the symptoms of stroke, which was essentially hopeless, medically speaking.

32 A good example of this phenomenon is the exorcistic catalog (KAR 44; see C. Jean, *La Magie Néo-Assyrienne en Contexte* (Helsinki, Helsinki University Press), pp. 62–82), which lists medical themes in its second part under the headings of *bulṭu*, "prescription," which would appear to be extracurricular for an exorcist. However, these particular *bulṭu*-prescriptions deal with seizure-like illnesses comparable to "epilepsy," being deemed as directly caused by a demon, which is precisely what the Hippocratic treatise *On the Sacred Disease* recognizes and attempts to reject. Another unexplored area of comparison between Babylonian and Greek medicine is the idea in this same Hippocratic treatise that epilepsy is essentially a disease of the brain, which may also be reflected in a unique Persian-period tablet from Uruk which lists stroke and epilepsy as diseases associated with the "mind"; cf. M. J. Geller, *Melothesia in Babylonia* (Berlin/New York: de Gruyter, 2014), pp. 2, 24.

33 The diagnostic omens were edited into the "Diagnostic Handbook," while the physiognomic omens were never edited into a single work. In fact, there was a "non-canonical" Diagnostic Handbook which dealt mostly with diagnostic omens related to witchcraft and epilepsy; see T. Abusch and D. Schwemer, *Corpus of Anti-Witchcraft Rituals* (Leiden and Boston, MA: Brill, 2011), pp. 434–43, and M. Stol, *Epilepsy in Babylonia* (Groningen: Styx, 1993), pp. 91–8. Another probably non-canonical exemplar of physiognomic omens is known from Assur, for which the rubric reads, "older physiognomic *alamdimmû*-omens which Esagil-kīn-apli did not unravel"; see N. Heeßel, "Neues von Esagil-kīn-apli, die ältere Version der physiognomischen Omenserie *alamdimmû*," in Maul and Heeßel (eds.), *Assur-Forschungen*, pp. 139–87, 145, 150.

colophons to a specific category of exorcist with the title of LÚ.KA.PIRIG.[34] Since this rather exotic title for the diagnostic-prognostic specialist is virtually unknown in other contexts, we can assume that this nomenclature in colophons is a way of designating the arts of diagnosis/prognosis to be a sub-speciality of *āšipūtu*. It appears, therefore, that Esagil-kīn-apli's reconstruction of medical literature had three different configurations associated with three different professional titles (*asû, mašmaššu*, LÚ.KA.PIRIG), and two different general disciplines (*asûtu, āšipūtu*),[35] the latter resuming both exorcistic and mantic texts relevant to healing. Each of these disciplines and sub-disciplines had its own unique theoretical basis, with mutual influences developing over time.[36] Let us consider all three categories in turn, with accompanying theories.

MEDICAL MODELS: RECIPES AND THEIR THEORETICAL UNDERPINNINGS

By the first millennium BCE, medical recipes all have a very standard format, which is easily recognizable and easily distinguishable from magical texts or divination, despite certain similarities to these other genres.[37] Medical recipes nearly always begin with the semiotic markers DIŠ NA (in Sumerian logograms) corresponding to Akkadian *šumma amīlu* "if a patient [lit. man]," often (but not always) followed by the term GIG (Akkadian *maruṣ*), "suffers from," completed by some medical condition

[34] A designation which is never adequately explained beyond the fact that he is a type of exorcist (see C. B. F. Walker and M. Dick, *The Induction of the Cult Image in Ancient Mesopotamia* (Helsinki: Helsinki University Press, 2001), p. 15, citing a bilingual incantation in which LÚ.KA.PIRIG is translated by "*āšipu*"). The title KA.PIRIG is likely to be a phonetic revival of an archaic priestly title abrig/*abriqqu*, which was no longer current in first millennium temples. The latter term appears in "mouth-opening" rituals for divine statues, with other archaic titles, such as the išib (*išippu*), guda (*pāšišu*), and abgal (*apkallu*), none of which were current in later periods. See Walker and Dick, *The Induction of the Cult Image*, p. 161. Why this title KA.PIRIG < abrig was revived specifically for diagnostic omens remains a mystery.

[35] It is obvious from the Exorcism Catalogue (KAR 44; Jean, *La Magie Néo-Assyrienne*, pp. 62–82) that the first entries concern texts which deal exclusively with the *mašmaššu*'s role as a priest rather than exorcist, highlighting the sacerdotal basis of his profession. Nevertheless, this is not like so-called "temple medicine" in Greece, which primarily involved the cult of Asclepius and incubation rituals to affect cures (V. Nutton, *Ancient Medicine* (London: Routledge, 2004), pp. 103–14; Krug, *Heilkunst und Heilkult*, pp. 120–87), so far unattested in Babylonia. However, one medical prescription against "ear-disease" contains several instructions for the patient to seek out the "sanctuary" of a god, where "he will see something favorable" (sig5 igi-*mar*; see BAM 503 = Scurlock, *Sourcebook*, pp. 376 ff.). This latter phrase occurs in dream omen texts and we may indeed have a reference to incubation within "temple medicine" in a Mesopotamian context. See also Finkel, *The Ark before Noah*, p. 112.

[36] The contradictions and conflicting views of health and therapy between these disciplines and sub-disciplines are consistent with what we know from other ancient societies; as G. Lloyd suggests, "in no case was there a single rigid orthodoxy in medical theory and practice even within the learned elite" (Lloyd, *Disciplines in the Making*, p. 82).

[37] See now Scurlock, *Sourcebook* and B. Janowski and D. Schwemer (eds.), *Texte zur Heilkunde* (Gütersloh: Gütersloher Verlagshaus, 2010).

(disease, syndrome, or symptom). Further symptoms may follow in a more detailed description of the problem. The next step is to list the *materia medica*, which can be introduced by a phrase such as, "in order to cure it" or "in order to remove it (the disease)," but neither phrase is obligatory.[38] Prescriptions usually ended on a positive note, that the patient will "improve" (*inêš*) or "get better" (*iballuṭ*).

Since many of the recipe ingredients were ordinary plants which could also be used as condiments or spices, the way these plants were prepared and administered as drugs was an essential part of the theory and practice of medicine. The *materia medica* are not simply listed, but in good recipe fashion the preparation of the ingredients is also described, such as in what form the substances are to be applied – to be swallowed, drunk, injected as a suppository, or applied externally as a salve or ointment, and for this the ingredients normally have to be treated in a variety of different ways. Individual amounts of ingredients are usually unreliable or unrealistic (often with all ingredients having the same dry or liquid measures), and dosages are usually not given, except for occasional mention of the number of times per day to be applied or ingested, and we are sometimes told that the recipe has been "tested." Finally, despite the reassurances that the patient should improve, recipe collections always have additional recipes for the same conditions, with the symptoms or disease simply introduced by "ditto," followed by a new listing of *material medica* and procedures for application.

One of the distinctive conventions of recipes of all kinds (not just medical ones) is the use of second person singular verbal forms as instructions for dealing with prescription drugs (e.g. you crush, you pound, you stir, you mix, etc.). Unlike others,[39] we are not particularly concerned with to whom this "you" refers, since the theoretical basis for recipes has nothing to do with its user but rather with its application, and in fact medical recipes could have been used by exorcists without impinging on the discipline itself.

Another characteristic aspect of *asûtu* is its inherent secularity, since although it invokes the healing goddess Gula or employs incantations, the theoretical mechanism for the use of recipes is based upon the perceived properties of active drug ingredients, which are independent of the influence of divinities or demons. The best example of this can be seen within the pharmacy employed in *asûtu*, which will be discussed below.[40]

[38] See P. Herrero, *La Thérapeutique mésopotamienne* (Paris: Editions Recherche sur les civilisations, 1984), pp. 39–40 for other similar terms.

[39] See Scurlock, "Physician, Exorcist, Conjurer, Magician."

[40] For an opposing view, see B. Böck, *The Healing Goddess Gula: Towards an Understanding of Ancient Babylonian Medicine* (Leiden and Boston, MA: Brill, 2014), pp. 178–9, who argues for magic and religion as the primary components of Babylonian medicine, with great significance given to appeals to the healing goddess Gula in medical texts. Though Greek medicine did not exclude reference to divinities, its secular character is unchallenged. Secularity is not atheistic but describes a system of healing which relies primarily upon what is understood as natural properties (*pace* ibid.,

DIAGNOSIS THEORY WITHIN *ASÛTU* (I.E. OUTSIDE THE DIAGNOSTIC HANDBOOK)

Once the idea has been proposed that Babylonian medicine spans the two separate disciplines of *asûtu* and *āšipūtu*, one result that we might expect from such a complex system[41] turns out to be verifiable: that there are actually two different systems of diagnosis operating side-by-side; one system of diagnosis operates within the context of medical recipes of *asûtu*, and the second within the Diagnostic Handbook of *āšipūtu*. There is a fundamental difference between these two approaches. Medical prescriptions used diagnostics to identify the specific conditions or symptoms which could be treated through drugs, and *asûtu* hardly ever indulged in *prognosis*, which was a primary function of the Diagnostic Handbook.[42]

Most descriptions of symptoms in *asûtu* recipes are quite simple and extremely laconic. The phrase, "if a man suffers from X-disease," served as a shortcut to listing descriptions of symptoms, since it was understood that the disease name was a shorthand notation for all relevant symptoms. The cases in which medical recipes contain elaborate or detailed descriptions of symptoms, with or without a diagnosis, represent unusual or non-textbook cases, always of more interest to physicians. This distinction is not to be found in the Diagnostic Handbook, which displays its own system of listing and classifying disease symptoms, from head to foot, which is *not* the case with symptoms listed in therapeutic prescriptions. In fact, the focus of the Diagnostic Handbook was on diseases and their underlying causes, while that of medical recipes was on symptoms and how to treat them. Even descriptions of the same symptoms, such as the color of discharged urine, are described differently within the Diagnostic Handbook and corresponding recipes relating to kidney ailments.[43] What this means is that *asûtu* does not

p. 194). For example, even the use of amulet stones and other ingredients (e.g. human bone) hung from a patient's neck in Babylonian medical prescriptions were not perceived as magical *per se* but as a means to harness the inherent properties of stones and minerals, since the uses of leather bag-amulets are predominantly found in medical recipes rather than in magical rituals; see W. Farber, "*ina* KUŠ.DÙ.DÙ(.BI) = *ina maški tašappi*," *Zeitschrift für Assyriologie* 63 (1973), 59–68.

41 The complexities of *asûtu* and *āšipūtu* can already be seen in the Assur Medical Catalogue see Steinert (ed.), *Assyrian and Babylonian Text Catalogues,* pp. 203–91.

42 This division is not unique to Babylonia but also existed within Greek medicine. As Ludwig Edelstein pointed out,

> The physician's art and prognostic acumen do not necessarily go together; on the contrary, profound medical learning may even be the basis for rejecting prognosis. It is possible to be a physician without practicing prognosis. (Edelstein, *Ancient Medicine*, p. 75)

43 BAM 7 No. 4 and No. 49 = M. J. Geller, *Renal and Rectal Disease Texts* (Die Babylonisch-assyriche Medizin und Untersuchungen 7; Berlin and New York: de Gruyter, 2005), pp. 65 and 251.

"borrow" or rely upon the diagnostics elaborated within its sister discipline, *āšipūtu*, but rather developed its own independent system of diagnosis and recording of symptoms.

Babylonian diseases often reflect anatomically-based nomenclature, and we can classify diseases according to their anatomical features, e.g. ailments associated with the head (including forehead and temples), eyes, nose, ear, tooth, etc., which often simply reflect these anatomical parts ("eye-disease," "nose-disease," "ear-disease," etc.). The anatomical diseases progress further down the body in descending order (*de capite ad calcem*) to include inter alia "lung disease,"[44] "cough," "abdominal-disease,"[45] "hip (or groin)-disease," "kidney-disease," "rectal-disease," "foot-disease," and various skin lesions, all of which are fairly clearly understood as referring to a pathology associated with a particular organ. Nevertheless, there are complications. First of all, not all anatomical organs are identified with certainty, such as the *takaltu*, which may (or may not be) the stomach.[46] Furthermore, there are no descriptions of tumors or skeletal problems (hardly surprising without the use of microscopes or x-rays), but only obvious abnormalities such as bleeding, pain, or inflammation, which can be detected through external examination without the aid of instruments.[47] There are other disease categories, however, which one could classify as non-anatomical, since symptoms are not associated with any particular organ.[48] Here we could include all gynecological problems (including pregnancy and childbirth), seizures (including epilepsy), stroke and palsy, any kind of paralysis or joint disease, fevers, mental illnesses, and many other diseases mentioned by name, which we cannot identify with certainty. Moreover, it is often difficult to distinguish between a disease and its symptoms, since dropsy (oedema), fevers, and inflammation often have the status of a malady; the same can be said for "constriction" (*hiniqtu*), although diarrhea and vomiting are usually limited to symptoms.

[44] It remains to be proven whether Babylonians considered the lungs to be responsible for respiration.

[45] One word for abdominal organs (including the stomach) is *libbu*, which can refer to the "heart," but as an organ of cognition; Babylonians (like many Greek medical writers and even Aristotle) thought of the heart as the "mind," since the physiological functions of the heart were not understood.

[46] Although the dictionaries give the meaning of this word as stomach (perhaps derived from *akālu*, "to eat", or *kullu*, "to hold"), it essentially means a bag or sack.

[47] Cf. Scurlock, *Sourcebook*, pp. 576, 580, "[if a woman] is sick with a crab" (*allutu*), which Scurlock associates with parasites (J. Scurlock and B. Andersen, *Diagnoses in Assyrian and Babylonian Medicine: Ancient Sources, Translations, and Modern Medical Analyses* (Urbana, IL: University of Illinois Press, 2005), p. 20), but the gynecological text in which this symptom appears deals with much more serious medical problems, such as uncontrolled menstrual bleeding. Galen referred to a tumor (usually of the breast) as a "crab" (cancer) because of its crab-like appearance (K.-H. Leven (ed.), *Antike Medizin, ein Lexikon* (Munich: C. H. Beck, 2005), p. 538).

[48] There is also no direct evidence in Babylonian medicine for a system of humors (blood, black bile, yellow bile, phlegm).

Beyond these basic terms, it is often difficult to translate disease names in a satisfactory way. One example is *bu'šānu*, a disease of the nose and mouth which might resemble diphtheria in some respects, and the term might be derived from the word *bu'šu* "stench." The term *amurriqānu* "a yellow-like ailment" has respectable Semitic cognates for "jaundice,"[49] but this does not explain another name for this ailment, *ahhāzu*.[50] Even well-known diseases have obscure names, such as *bennu* for epilepsy (a general term for all kinds of seizures), or alternatively *bēl ūri* "Lord-of-the-roof-disease" (also a term for epilepsy).[51] Another term, *miqtu*, is a calque on Sumerian an.ta.šub.ba, lit. "fallen from heaven," and as such could mean "falling sickness," a common term for epilepsy in Hippocratic medicine.[52] However, it is often difficult to distinguish this disease from another known as *mišittu*, lit. "stroke" (< *mašādu*, "to strike"),[53] distinguishable from the near synonym *mihṣu*, "stroke," which never refers to the medical condition of aphasia or paralysis. The disease of leprosy is known as *saharšubbû*, a Sumerian loanword (lit. "thrown dust") which is hardly self-explanatory, but synonyms for this term are *epqu* "scales" (and *epqēnu*, a similar term) and *garābu* "scab," indicating a skin complaint; in no case do these terms reflect the pathology of Hansen's disease, the modern classification for leprosy.[54] The disease-labels reveal little about the disease itself, and we are almost entirely dependent upon descriptions of the symptoms for the characteristics of the diseases, on the basis of which we can try to match up the disease-terminology with recognizable modern equivalents, to a limited extent. Infectious diseases come under the general label *li'bu*-disease, derived from *la'abu*, "to infect," but this disease name occurs more often in non-medical contexts (e.g. incantations or omens) and appears not to be a *terminus technicus*.[55]

The point is that nothing among these terms resembles modern disease taxonomy, but many conform to the basic concepts of what constitutes

[49] See Jewish Babylonian Aramaic *yrqn'*, M. Sokoloff, *Dictionary of Jewish Babylonian Aramaic* (Ramat Gan: Bar Ilan University Press, 2002), p. 543b.

[50] The disease *ahhāzu* has symptoms pointing to "jaundice" but attributed to a demon, and the disease was considered fatal and usually not treatable. M. Stol has discovered two different varieties of jaundice described in a Hippocratic treatise, the second of which includes a black tongue, and one of the distinguishing features of *ahhāzu*-disease is indeed a black tongue (M. Stol, "An Assyriologist Reads Hippocrates," in H. Horstmanshoff and M. Stol (eds.), *Magic and Rationality in Ancient Near Eastern and Graeco-Roman Medicine* (Leiden: Brill, 2004), pp. 63–78, 77).

[51] Based on Sum. *lugal-ùr-ra*, which is equally obscure, but cf. T. Kwasman, "The Demon of the Roof," in I. L. Finkel and M. J. Geller (eds.), *Disease in Babylonia* (Leiden and Boston, MA: Brill, 2007), pp. 160–86 for Aramaic parallels.

[52] O. Temkin, *The Fallen Sickness: A History of Epilepsy from the Greeks to the Beginnings of Modern Neurology* (Baltimore, MD: Johns Hopkins University Press, 1994), pp. 85–6. The disease *miqtu* often affects a small child, which is more likely to refer to epilepsy than to a stroke.

[53] See J. V. Kinnier Wilson and E. H. Reynolds, "On Stroke and Facial Palsy in Babylonian Texts," in Finkel and Geller (eds.), *Disease in Babylonia*, pp. 67–99.

[54] *Pace* Scurlock and Andersen, *Diagnoses in Assyrian and Babylonian Medicine*, pp. 70–3.

[55] See M. Stol, "Fevers in Babylonia," in Finkel and Geller (eds.), *Disease in Babylonia*, pp. 1–39, 11–15.

a disease in the Hippocratic Corpus.[56] Greek physicians employed similar descriptions of diseases they knew, mostly taken from external signs and symptoms, much in the same way that their Babylonian colleagues worked. In the course of time, however, Babylonian and Greek medicine parted company: one tactic among some Greek physicians was to base diagnosis on the theory of four bodily humors,[57] and Greeks developed the methodology of the case history,[58] which examined all of the symptoms of a single patient, mentioned by name, rather than relying upon a generic list of diseases and symptoms, as was done in Babylonian medicine.

Finally, there is the question of the incursion of magic into medicine, through the use of incantations. This manifests itself in two ways. First, there are treatments for *māmītu*-disease in medical texts, but the etiology of this disease is really magical: illness caused by a curse resulting from the violation of an oath, as described in standard magical compositions.[59] While we cannot be certain as to who exactly was responsible for recognizing and dealing with health issues resulting from a broken oath, there are two routes for treating such problems. One was to employ the prominent genre of nam. érim.búr.ru.da magical incantations, "to absolve the (bad effects of a) *māmītu*-oath," while another way was to use medical recipes. In the latter case, however, the medical condition *māmītu* is a *disease* rather than a magical event, as can be seen from the following clause dealing with the lungs: "if a man is afflicted with *ṣētu*-fever (and) it turns into 'oath-disease' (*māmītu*)."[60] A disease can only mutate into another disease, not into magic.[61]

[56] See disease terminology in W. Golder, *Hippokrates und das Corpus Hippocraticum* (Würzburg: Königshausen und Neumann, 2007), pp. 143–55 and E. Craik, *The "Hippocratic" Corpus: Content and Context* (London and New York: Routeledge, 2015), pp. 292–4.

[57] Although the theory of four humors did not become standard doctrine until Galen, Diocles already accepted it in the fourth century BCE (see Nutton, *Ancient Medicine*, p. 121). See also Craik, *The "Hippocratic" Corpus*, pp. 207–8, for Hippocratic treatises which discuss the four bodily "fluids" without specifically referring to them as "humors." Although Craik comments that "the importance of the humors has been exaggerated in Hippocratic interpretation" (ibid., p. 208), nevertheless the basic assumptions upon which the theory was based were Hippocratic. See also M. J. Geller, "Phlegm and Breadth – Babylonian Contributions to Hippocratic Medicine," in Finkel and Geller (eds.), *Disease in Babylonian*, pp. 173–85, for some Babylonian parallels.

[58] Golder, *Hippokrates und das Corpus Hippocraticum*, pp. 36–9; Nutton, *Ancient Medicine*, p. 89.

[59] One thinks here of Šurpu, a magical composition which aims at breaking the effects of a - violated *māmītu* or oath (see E. Reiner, *Šurpu, A Collection of Sumerian and Akkadian Incantations* (Archiv für Orientforschung, Beiheft 11; Graz: Biblio-Verlag, 1958); see also van der Toorn, *Sin and Sanction*, pp. 138–9). See also BAM 49:32', "if a man's 'epigastrium' (lit. 'mouth of his stomach') is like an oath (^{d}nam.érim) has gripped him . . ." On the other hand, a notation added to a medical recipe for "stricture of the bladder" (*hiniqtu ellibuhhi*) prescribes a cocktail of juices mixed with grains and injected into the patient's anus to relieve constipation, while at the same time his hemorrhoids are treated to facilitate defecation. However, the recipe appends a note to say that "this is a lotion for 'absolving an oath' (nam.ri.búr.da) and for any illness (gig dù.a.bi)," also showing the influence of magical thinking in a standard medical recipe; cf. Geller, *Renal and Rectal Disease Texts*, No. 34, pp. 52–3.

[60] DIŠ NA U_4.DA TAB-*su-ma ana* NAM.ERIM$_2$ GUR-*šú* BAM 558 rev. 25, courtesy J. Cale Johnson.

[61] The point is that the usual paradigm may not always apply, namely that the exorcist is concerned with aspects of the patient's behavior, which may have brought about illness, while the physician

There is one other qualification to the use of *āšipūtu* within medical recipes (*asûtu*): the incantations which accompany medical recipes often have a very different character than incantations which appear within formal magical compilations such as *Utukkū Lemnūtu*.[62] Medical incantations are most often in vernacular Akkadian rather than Sumerian-Akkadian bilinguals, or mumbo-jumbo, or represent rather jejune illustrations of the nature of the illness (e.g. toothache characterized as a worm, or eye disease caused by a mote).

EXORCISTIC MODELS: INCANTATIONS AND RITUALS AND THEIR THEORETICAL UNDERPINNINGS

Mesopotamia has provided an extremely rich corpus of incantations of various types, aimed at all kinds of demonic activity and associated misfortunes,[63] but not all incantations are specifically aimed at healing. One of the primary incantation compositions which specifically renders the victim as "patient" is *Utukkū Lemnūtu*, in which the exorcist claims to enter the patient's home to treat his condition, probably from head to foot:[64]

When I approach the *patient*,
when I entered the *patient*'s house,
my hand was present at his head
and I studied the sinews of his limbs.

The exorcist's actions are in response to the actions of the demons, who "studied (the patient's) entire anatomy in order to lodge themselves in a diseased place."[65] Just as medical treatises and prescriptions had distinctive genre markers, incantations and rituals were usually accompanied by traditional labels, such as Sumerian ÉN and KA.INIM.MA for "incantation/spell," and a concluding label TU_6.ÉN with the same meaning; magical rituals had a designation DÙ.DÙ.BI or KÌD.KÌD.BI, meaning "its ritual." These were graphic reminders that the texts found between the markers

was strictly concerned with alleviating symptoms. The term *māmītu*, even as a designation of "oath-disease," may still reflect the etiology of the disease, even if this may not necessarily be relevant to the specific drug therapy.

62 See M. J. Geller, *Healing Magic and Evil Demons, Canonical Udug-hul Incantations* (Die Babylonisch-assyrische Mediuzin in Texten und Untersuchungen 7; Boston, MA and Berlin: de Gruyter, 2016), p. 30. For examples of incantations that accompany medical recipes, see T. J. Collins, "Natural Illness in Babylonian Medical Incantations" (PhD thesis, University of Chicago, 1999).

63 See Schwemer, "The Ancient Near East."

64 Geller, *Healing Magic and Evil Demons*, p. 197.

65 Ibid.

technically belonged to *āšipūtu*, and the fact that both *asûtu* and *āšipūtu* could be marshaled to treat illnesses encouraged a good deal of interdisciplinarity, even to the extent of harnessing competing methods.

One should note a formal similarity between magical and medical texts, that the recipe format of 2.p.s instructions is also employed for magical rituals. This is primarily true, however, only in Akkadian magical texts, since formal Sumerian or bilingual incantations generally use an imperative verbal form for ritual instructions.[66] The recipe format in magical texts is therefore likely to be a later development within Mesopotamian magic.

The theoretical underpinnings of exorcism are so wellknown and frequently discussed that there is no need to review the evidence here, except to say that it is essentially religious rather than secular, dominated by appeals to gods against demons.[67] Witchcraft, however, is exceptional to a certain extent, since it can be assigned to human agency as a kind of *technē* and hence secular, which may be why recipes are commonly employed in this domain as well as incantations and rituals. Already in the second millennium BCE recipes (resembling *asûtu*) were being employed against witchcraft:

> If a man (DIŠ NA) is bewitched: You parch (*turrar*) "apricot-turnip," kupad-salt, the kidney of a lamb that has not yet eaten grass, (and) *ernīnu*-(plants). He eats it and will recover.[68]

This type of text represents an incursion of *asûtu* into *āšipūtu*, or the borrowing of medical techniques in the fight against illnesses which are defined by magic, not medicine.

[66] An example typical of such instructions is found in *Utukkū Lemnūtu* incantations (ibid., p. 250): "Bind the limbs of that man and cast the spell of Eridu. Sprinkle that man with water, and pass the censer and torch over him." All of the verbal forms are imperatives (ibid., p. 176). However, the ritual tablet of Šurpu, designed to break the bad effects of a violated oath or taboo, uses the 2.p.s. verbal instructions of recipes, "when you perform the (Šurpu) ritual, you set up a brazier," etc. (see Reiner, *Šurpu*, p. 11).

[67] An example of how demons affect one's health is a dominant theme in incantations, such as the following extract from *Utukkū Lemnūtu* VII 34–40 (Geller, *Healing Magic and Evil Demons*, pp. 137 and 221):

> (The demon) approached that man and touched his hand,
> and chased after him, went to his house,
> and made him neglect his body (lit. limbs).
> His eyes are open, but he sees no one,
> his ears are open, but he hears no one.
> That man is miserably depressed by the hand of Fate (Namtar)
> the "sacrilege"-(*asakku*)-disease has overwhelmed him gravely (lit. "bitterly").

There are no judgments within this text regarding the patient's moral state and whether he deserves to be ill or not, but the dominant theme concerns illness as an external attack on the patient's body, personified as demons.

[68] Abusch and Schwemer, *Corpus of Anti-Witchcraft Rituals*, p. 66.

MANTIC MODELS: THE DIAGNOSTIC HANDBOOK AND ITS THEORETICAL UNDERPINNINGS

There are two genres of texts listing physical human symptoms, often from head to foot, as a way of forecasting future health- and life-events. The first is the Diagnostic Handbook consisting of forty tablets or "chapters," and the second are physiognomic omens, which were never collected into a single series. Both of these genres of diagnostic and physiognomic omens have similar standard casuistic formats, which supply the basic theoretical logic for noting symptoms as omens. We will deal first with symptoms as omens (the Diagnostic Handbook) and come later to physiognomic omens as a sub-category of this same system.

The opening two "chapters" (tablets) of the Diagnostic Handbook differ significantly from all other parts of this work and probably represented later accretions to the text.[69]

> If he (the exorcist) sees a black pig: that patient may die; (alternatively) he may suffer acutely but then recover. If he (the exorcist) sees a white pig: that patient may get better; (alternatively) stress (*dannatu*) may take hold of him. If he (the exorcist) sees a red pig: that patient may die within three months, (alternatively) within three days. If he (the exorcist) sees a spotted pig: that patient may die from dropsy; it is worrying, one should not approach him.[70]

The unusual character of such omens within a work intended as "medical" (in some form) is striking, since the text resembles terrestrial omens much more than prognostic ones; although referring to the patient, the signs enumerated in these two tablets bear no relation to the patient's physical condition or state of health, but to external objects which have portentous significance.[71] A good example of the logic is the exorcist's encounter with a pregnant woman en route to the patient, a bad portent which Irving Finkel explains as follows: one is entering (an impending birth) and another is departing (the patient's demise), as reflecting

[69] The first two tablets of the Diagnostic Handbook were not edited in Scurlock, *Sourcebook*. The entire character of the first two tablets of the Diagnostic Handbook, resembling *bārûtu*, "divination," also impinges upon these healing genres and derives omens from animal entrails or flights of birds, smoke, oil on water, and other methods of forecasting the future; *bārûtu* could also be employed in health contexts, perhaps as a way of double-checking the results of diagnostic or prognostic omens proposed by the *mašmaššu* (see Maul, *Die Wahrsagekunst*, pp. 77, 116 f., 174–5, 178).

[70] A. R. George, "Babylonian Texts from the Folios of Sidney Smith. Part Two: Prognostic and Diagnostic Omens, Tablet I," *Revue d'Assyriologie et d'Archéologie Orientale* 85 (1991): 137–67, 142f.

[71] In fact, in later periods the actual meaning of the first two tablets of the Diagnostic Handbook puzzled ancient scholars to the extent that no less than three extensive commentaries survive explaining how these omens were intended to reflect the patient's health. For instance, commentaries on this passage insist that the word for "pig" is actually a code word for pandemic fever. See George, "Babylonian Texts from the Folios of Sidney Smith. Part Two" for a thorough treatment of these commentaries.

some sort of balance of human fates.[72] The point is that while adhering to the basic "if P then Q" formulation of all genres of Mesopotamian divination,[73] the essential worldview which originally governed the Diagnostic Handbook was that divinities were in control, and that symptoms reflected supernatural decisions about the fate of patients and the course of disease. This is one of the significant distinctions between *asûtu* and *āšipūtu*, in that disease diagnosis within recipes is secular while diagnosis and prognosis within the Diagnostic Handbook originally derives from Mesopotamian religion. Over time, however, this situation changes, and secularity was eventually adopted within the Handbook as well, as we will soon see.

The critical characteristic of a "mantic" medical text is that it deals with *acute* diseases; the Diagnostic Handbook recorded diseases in a critical or terminal stage,[74] which explains why the prognoses within the Diagnostic Handbook were usually pessimistic.[75] The opening line of the Handbook explains that the specialist (LÚ.KA.PIRIG) is visiting the patient at home, which is likely to be because the patient is on his deathbed and not ambulatory. This practice generally conforms with the Hippocratic treatise *On Prognosis of Acute Diseases*.[76] One of the obvious problems of dealing

[72] Finkel, *The Ark before Noah*, pp. 322–3.

[73] All systems of prediction in Mesopotamia employ a similar casuistic form, namely if P then Q, probably originally borrowed from legal compendia (law codes), but then spreading to divination and omen collections, in which an event is associated with another event in a rather loose logical manner (often described as *post hoc ergo propter hoc*); see F. Rochberg, *In the Path of the Moon: Babylonian Celestial Divination and Its Legacy* (Leiden and Boston, MA: Brill, 2010), pp. 378 f. Naturally, there is something very different about prognostic omens compared with celestial or terrestrial omens, which associate totally unrelated phenomena, such as the birth of an ill-formed fetus with the fate of the king. From a modern perspective, diagnostic omens (symptoms) could be perceived as the products of many generations of observations, with Karl Popper-like verifiable results. After years of studying fevers, it would not be difficult to devise a list of signs of symptoms closely associated with a specific type of fever, and to conclude whether the patient would be likely to either survive or succumb to this fever, depending upon circumstances. This was not, however, how the Diagnostic Handbook was understood by most ancient scholars, who assumed omens to be messages from gods, although there are indications that a more secular approach took hold in late periods.

[74] Ulrike Steinert reminds me that one tablet of the Diagnostic Handbook dealing with pregnancy ("chapter" 36) does not deal with acute diseases but whether the foetus is likely to be male or female, and the prediction that the "foetus will be rich" resembles apodoses of physiognomic omens. Nevertheless, the following tablet in the Diagnostic Handbook predicts whether the pregnant woman will live or die, returning to the dominant theme of acute disease. See M. Stol, *Birth in Babylonia and the Bible* (Groningen: Styx, 2000), pp. 194–203 and Scurlock, *Sourcebook*, pp. 145–257.

[75] Medical prescriptions had their own method of describing disease, usually with a quite different vocabulary and purpose. Despite a few parallels between this work and therapeutic recipes (pointed out by Scurlock, *Sourcebook*, p. 11), the majority of diagnostic comments found in medical recipes are not duplicated in the Diagnostic Handbook.

[76] Ludwig Edelstein's succinct description of prognosis in the Hippocratic corpus bears uncanny resemblance to the Babylonian Diagnostic Handbook, although Edelstein himself did not see the connection. First, Hippocratic prognoses were limited to acute cases, "in certain diseases which are called acute because the crisis is reached rapidly and because death usually ensues very quickly" (Edelstein, *Ancient Medicine*, p. 67). Moreover, Edelstein points out that the Hippocratic corpus defines acute diseases as "pleurisy and pneumonia and phrenitis and *kausos* and the others of this sort where the fever generally is continuous, and that individual diseases with common

with acute diseases is the difficulty posed by adequate treatment, and all systems of ancient medicine warned practitioners against offering prognoses (or treatment) in terminal cases.[77]

Finally, this text is not about patients but about symptoms, and it attempts to be a comprehensive description of externally noted symptoms of specific diseases, gathered from the collective experience of observing numerous patients; since there are no case histories of individual patients, it is virtually impossible to establish a retrospective diagnosis corresponding to modern diseases or disease taxonomies.[78]

The theoretical nature of the Diagnostic Handbook can be seen from alternative propositions put forward to provide a systematic range of symptoms. For instance, one passage reads, "If his (a patient's) forehead contains heat, he should recover, if his forehead contains sweat, he should recover," followed by similar symptoms, "if his forehead is cold," "if his forehead is white (and) his tongue is white," "if his forehead and face are white," "if his forehead is yellow," "if his forehead is red," "if his forehead is black."[79] The scheme of characteristics is hot or cold, moist (dry is missing in this passage but appears elsewhere), white, yellow, red, or black, similar to the Hippocratic elementary qualities of hot, cold, dry, or wet.[80] Although similar sequences of characteristics appear throughout the Diagnostic Handbook, one cannot conclude that Babylonians had invented a precursor to a theory of four humors, in which these same colors represented phlegm, blood, and yellow and black bile, since these fluids (phlegm, bile, and blood) are rather randomly distributed within the Diagnostic Handbook and do not themselves form the basis for a comprehensive system of diagnosis. In fact, the Babylonian concept

characteristics or symptoms could be linked under these general categories of acute diseases." Also, prognoses were based upon the patient's behavior rather than on any "idea of a disease *per se* considered apart from the organ which is affected by it" (ibid., p. 68). The Hippocratic method of prognosis is clearly reminiscent of Babylonian practice. Edelstein gives a pithy description of Hippocratic prognosis as "made either semeiotically, by taking the way the organs are functioning and the whole constitution of the body as signs, or mantically, by divination, without paying attention to the patient's physical condition" (ibid., p. 70).

All three of these descriptions would apply equally well to the Babylonian Diagnostic Handbook. See also P. van der Eijk in chapter 16 in the present volume, pointing to a checklist of symptoms within the Hippocratic Corpus, which certainly characterizes the Diagnostic Handbook.

[77] See Edelstein, *Ancient Medicine*, p. 97: "the physician is forbidden to give treatment when the disease is in a critical stage because, according to theory, intervention at that time is injurious." There is similar Babylonian evidence for this practice, such as, "the *asû*-physician should not lift a hand to this particular illness; that patient will die and not [get better]" (Scurlock, *Sourcebook*, pp. 518, 527). The phrase that a physician "should not lift a hand" is a direct quote from hemerologies which prohibit treatment on unfavorable days of the month, and since hemerologies belonged to the exorcist's area of expertise, the context is clear. The exorcist is the one who gives prognoses in cases of acute disease, and it is he who advises the physician to withhold treatment.

[78] Galen criticized Methodists for using general rather than specific terminology for diseases, and in fact Galen's criticisms of Methodist practice could equally apply to Babylonian compilers of the Diagnostic Handbook; see Geller, *Melothesia*, p. 21.

[79] Scurlock, *Sourcebook*, pp. 33, 40

[80] See P. van der Eijk in this volume (chapter 16).

of disease throughout the Diagnostic Handbook remained one in which disease was mostly the result of external invasion from demons or natural forces (e.g. wind), rather than any perceived imbalance or superfluity of bodily fluids or humors, although this Babylonian concept of external attack was identified in Hippocratic terms as antiquated notions of old-fashioned (pre-Hippocratic) medicine.[81]

The Diagnostic Handbook provides useful and detailed information regarding a theoretical view of Babylonian anatomy, which resembled Hippocratic medicine in that it was not based upon autopsy or dissection; both were equally vague and ill-informed regarding internal human anatomy.[82] Apart from the observed symptoms, parts of the body are represented by overlapping terms for stomach (heart), belly, abdomen, veins, nerves, sinews, muscles, tissues, joints, loins (hips), liver, kidneys and bladder, spleen, lungs, brain, etc. Major organs, such as the pancreas, were unknown (also to Hippocratic physicians), and functions of these organs (even the heart and probably lungs as well) remained unclear. In fact, the internal organs of sacrificial sheep were much better known to diviners than were human organs known to physicians or exorcists. There is virtually no late evidence for Babylonian dissection expressly for the study of anatomy, as happened in Alexandria and elsewhere.

CAUSALITY IN THE DIAGNOSTIC HANDBOOK

In contrast to *asûtu*-recipes, the Diagnostic Handbook is primarily concerned with causality, and the patient's condition and symptoms were frequently attributed to the "hand of a god" or demon or witch.[83]

[81] See now M. Asper, "Medical Acculturation? Early Greek Texts and the Question of Near Eastern Influence," in B. Holmes and K. -D. Fischer (eds.), *The Frontiers of Ancient Science, Essays in Honor of Heinrich von Staden* (Berlin: de Gruyter, 2015), pp. 19–45, 26. It would be useful, for the purposes of the present argument, to assume that the Hippocratic corpus, including such seminal works as *On the Sacred Disease*, represented a dramatic change in direction for Greek medicine (although not to be confused with *le miracle grec*). The designation of Cnidian medicine for pre-Hippocratic medical practices has now been abandoned by Greek specialists (so ibid., pp. 21–2), regrettably in some respects, since this was a convenient term to designate earlier forms of Greek medicine which resembled Babylonian practice but were later abandoned by Hippocrates and his followers. V. Nutton appears to have grasped the point of a major change occurring in Greek medicine by the mid-fourth century BCE, since he argues that while "chants, charms, and so-called sympathetic or white magic" continued to be used within medicine, "practitioners who relied primarily on such procedures for their cures were now marginalised, or at least excluded from the new idea of medicine" (Nutton, *Ancient Medicine*, p. 113).

[82] Babylonian scholarship (probably within the discipline of *āšipūtu*) created one composition (Ugumu) listing all known parts of the human body from head to foot, concluding with entries identifying more general features, such as the body's shadow, general stature, and skeleton (N. Veldhuis, *History of the Cuneiform Lexical Tradition* (Münster: Ugarit Verlag, 2013), pp. 158–9; this academic study of anatomy in Ugumu has remarkably little overlap with the medical corpus).

[83] Some "hand of X" entries can be rather baroque, such as "hand-of-the-shrine" (ŠU ZAG.GAR.RA); see Scurlock, *Sourcebook*, pp. 77 and 81, translated as "hand of the tithe"). However, the subject of causation in the Diagnostic Handbook deserves a much more nuanced treatment, since although the "diagnoses" are expressed in simplistic terms as the "hand" of an agent, the entire process of

The point is that the earliest known exemplars of the Diagnostic Handbook incorporated the expression "hand of a god" (mentioned by name) or "hand of a ghost" as a diagnostic entry, and this became standard practice for all first-millennium recensions of the text. It seems plausible that the expression "hand of a god/ghost" originally implied the literal intervention of divine agency, or even human agency (i.e. witchcraft), in the patient's fate: a god, goddess, demon, ghost,[84] or malevolent human had actively influenced the patient's health through supernatural or magical means. The concept that the art of diagnosis and prognosis technically belongs to exorcism (*āšipūtu*) rather than "medicine" *per se* is counterintuitive to the modern mind, but A. Leo Oppenheim grasped this idea many years ago when describing the Diagnostic Handbook:

> whenever a treatment is prescribed – and that is only rarely the case – it is not medical but exclusively magical. Even the names of the disease mentioned are not medical but point as a rule to the deity or demon that has caused them.[85]

However, by the first-millennium BCE recensions of the Diagnostic Handbook, causality began to take on a somewhat different shape. The new causality has a more graphic character, e.g. if a man has pain in his head, (entire) body, top of his nose, and his lips and suffers spasms, "someone (var. a ghost) among his relatives who died of thirst has seized him."[86] Other kinds of causality based upon a patient's own (usually immoral) behavior is considered as a factor in his illness, as can be seen in the same passage which describes a patient's body as being covered with sores, "being with a woman in bed has caught up with (him)"; it is the "Hand of the Moon god";[87] a similar statement attributes an erratic pulse to the "the Hand of the Sun-god because of a man's wife" (Scurlock 2014: 39).[88] The "hand of a god" is no longer sufficiently causal and requires additional information to explain the disease, i.e. illicit sexual relations. An even more extraordinary cause attributes the patient's condition to the fact that he has

arriving at a diagnosis (in later periods) was probably the result of more sophisticated criteria, based upon exhaustive observation of the external human body and symptoms, and then inferring causes. The process is reminiscent of Galen's comment that "the diagnosis of bodies is prior, on the basis of signs, of course, and after which [comes] the discovery of their causes" (von Staden, *Herophilus*, p. 104). Since there is no philosophical tradition in Babylonian schools, the process of determining diagnoses was never put to writing but was no doubt a topic for discussion among ancient scholars.

84 Ghosts are occasionally described in the Diagnostic Handbook in graphic terms, such as a ghost of one burned to death (Scurlock, *Sourcebook*, p. 204), or a ghost struck down in water (ibid., p. 209).

85 A. L. Oppenheim, *Ancient Mesopotamia, Portrait of a Dead Civilization* (Chicago, IL: University of Chicago Press, 1964), p. 224.

86 See ibid., pp. 16, 22; the text implies that the descendant had neglected to make the proper funerary offerings to his forebears.

87 Scurlock, *Sourcebook*, pp. 17 and 23.

88 Ibid., p. 39.

robbed something from a boat and the "god of the quay" has seized him.[89] In another entry, "if a patient gnaws at his arms, he had someone strangled in a murder and the double of a dead man has seized him,"[90] also providing a moral judgment. Other causes are more prosaic, such as a patient simply being fed dirt.[91]

All of these additional "causes" for diseases show an erosion from the original concept of personal and arbitrary divine intervention in favor of causes to be found within a patient's own behavior or moral failings. Although in modern terms such "causes" would be considered as "religious" in origin, in the ancient world they imply a degree of secularity, in the sense that disease was not simply the whimsical decision of a god deciding one's fate, but it was the concomitant result of human activity, with the implicit message that avoiding such behavior would also prevent the disease; one's fate was in one's own hands.[92]

There is further evidence for changes in medical thinking in the first millennium. The twenty-eighth tablet of the Diagnostic Handbook features prognoses based upon diseases which have altered their characteristics and morphed into other diseases, such as "seizure" (*miqtu*) changing into Hand-of-the-Ghost-disease;[93] the Hand-of-the-ghost here can only designate a disease and not the direct action of a ghost against a patient. Further evidence is supplied by a double listing in the Diagnostic Handbook Tablet 33 in tabular form,[94] in which the "hand of a god" in one column matches up with a technical disease name in the second column (in standard commentary layout). In effect, in some circles at least, the attributions to the "hand" of a god or ghost or demon is no longer to be taken literally but is seen as a diagnosis, which makes us reflect again on the Hippocratic treatise, *On the Sacred Disease*. There is no doubt that certain diseases (e.g. epilepsy) would never lose their sacred character, at least in public perception, but as diagnosis and therapy progressed over time, a more technical notion of disease took hold in mid-first-millennium BCE Mesopotamia, which would bring medicine more in line with its Greek counterpart, even within the discipline of *āšipūtu*.

[89] Ibid., pp. 17 and 23.
[90] Ibid., pp. 76 and 80.
[91] Ibid., pp. 84 and 88.
[92] See P. van der Eijk, *Medicine, Philosophy, Classical Antiquity: Doctors and Philosophers on Nature, Soul, Health and Disease* (Cambridge: Cambridge University Press, 2005), pp. 48–52, contrasting the divine character of the "sacred disease" with the human factors which are also causes. The analogies are not perfect, since in Greek terms the "divinity" of a disease consists of natural factors such as the effects of climate, while human factors include the patient's age, physiology, and for how long the patient has been affected by the disease. Nevertheless, these "factors" represent an attempt by the Hippocratic author of the treatise to counter the previous claims of magical and religious causes, such as divine punishment bringing on disease, and it is these very attitudes from pre-Hippocratic medicine – although discredited by the Greek author – which resemble the Babylonian Diagnostic Handbook.
[93] See Heeßel, *Babylonisch-assyrische Diagnostik*, pp. 307–317, and Scurlock, *Sourcebook*, pp. 211–13.
[94] Heeßel, *Babylonisch-assyrische Diagnostik*, pp. 358, 363; see Scurlock, *Sourcebook*, pp. 235, 240.

MANTIC MODELS: PHYSIOGNOMIC OMENS AND THEIR THEORETICAL UNDERPINNINGS

One of the most remarkable colophons ever found on cuneiform tablets appears on the catalog of omens cited under "mantic models" above.[95] The colophon attributes a new edition of these texts to Esagil-kīn-apli, and then characterizes the diagnostic omens (of the Diagnostic Handbook) as referring to both (physical) illness and "depression," while physiognomic omens are taken "from a person's form and external appearance, which would determine his fate." Neither of these genres is to be "recited" or "invoked" aloud, presumably meaning that they are not to be disseminated abroad to laymen who would not be able to interpret them correctly, but the practitioner is advised to know both diagnostic and physiognomic omens thoroughly in order to be able to form his diagnosis.[96]

What this colophon tells us is that both diagnostic and physiognomic omens constituted the art of diagnosis and prognosis and should be taken together as aspects of the same process, with the full integration of physiognomy into the practice of healing arts. For instance, many of the physiognomic omens are based on body marks or lesions, such as the *umṣatu* or *pindû* "moles," which are also the subject of medical recipes intended to remove them. Apart from physical characteristics, one entire text of physiognomic omens is derived from the patient's own words, which may reflect things said under duress or crisis, perhaps during a breakdown or acute illness ("I want to die," versus "I want to live!," or "I want to be rich!" versus "I have nothing").[97] Babylonian physiognomic omens are relatively easy to decipher but difficult to understand, such as the propositions, "if (a man's) right eyebrow is thick," "if a man's nose is long," or "if a man's tongue shines," having associated predictions of success, longevity, or premature death.[98] In fact, the same kind of logic can be found in the physiognomic

[95] See Finkel, "Adad-apla-iddina, Esagil-kin-apli, and the Series SA.GIG."

[96] Ibid., pp. 148–50.

[97] See Böck, *Die babylonisch-assyrische Morphoskopie*, p. 131. The same pattern can be observed in the Diagnostic Handbook, in which many of the symptom-omens report the patient's own words, such as "my stomach, my stomach!" as an indication of severe pain or discomfort. Chapter 36 of the Diagnostic Handbook notes that certain symptoms will predict that the foetus will be "rich" (see Stol, *Birth in Babylonia and the Bible*, pp. 194–203, and Scurlock, *Sourcebook*, pp. 145–255), providing a clear link between diagnostic and prognostic omens.

[98] Böck, *Die babylonisch-assyrische Morphoskopie*, pp. 38–40) makes an attempt to explain the logic of these omens as examples of analogy (e.g. between human and corresponding animal characteristics) and association of ideas, such as long and short referring both to body parts and the length of one's life. See also F. Rochberg, *The Heavenly Writing* (Cambridge: Cambridge University Press, 2004), p. 50: "Even within the physiognomic omen series, when an untoward event was forecasted for the individual – such as 'If the hair on his head is sparse: His days (= life) will be short; he will be critically ill' – such forecasts do not appear to have been viewed as amenable to magical manipulation or appeal to the gods." Rochberg is correct that no rituals were employed to counter the effects of physiognomic omens (although such rituals exist with other types of omens), but the reason is that physiognomic omens were intended to provide information about a person's qualities,

work of Pseudo-Aristotle, which records analogous examples of human physiognomy being interpreted through similarities to animals and then to the general characteristics of those same animals.[99] However, the actual language and expressions of the Babylonian physiognomic omens reflect the standard apodoses of Babylonian omen literature, rather than remarks about the patient's character or personality.[100] In other words, physiognomic omens become useful as a way of determining whether the patient is in general good health or not, since a favorable prediction (e.g. that he or she will have length of days or wealth) implies that the patient's condition is essentially healthy, despite an encounter with acute disease. The insistence by the Esagil-kīn-apli-school that diagnostic and physiognomic omens be

although employing the standard language of omen texts; they were not meant to be interpreted literally.

99 Cf. S. Vogt, *Aristoteles Physiognomonica* (Berlin: Akademie Verlag, 1999) on physiognomic omens in Aristotle's *Physiognomonica* and *Historia animalium*, based upon characteristics known from animals which are employed analogously to identify personal qualities (especially human psychology) associated with similar physiognomic features; there are many similarities here with Babylonian physiognomic omens which remain to be studied. Vogt (ibid., p. 117) mentions Babylonian parallels but also draws attention to a certain Zopyros, a contemporary of Socrates who hailed from the East and was known for his expertise in this discipline. The comparisons between Greek and Akkadian physiognomy are rather striking, although in order to make such comparisons one has to take account of the different literary formats of each genre. The typical casuistic style of Babylonian *Listenwissenschaften* gives each omen separately, with its accompanying interpretive forecast or interpretation. For example, the omens derived from a woman's breasts read as follows:

> If a woman's breasts are abnormally large, she is a "god's wife."
> If a woman's breasts are small, she will be wanting.
> If a woman's breasts are drooping, she will be fruitful.

The apodoses here are both descriptive and predictive. Let us contrast these with physiognomic omens in which the (male) subject has animal features:

> If a man has the head of a bird, he will have success.
> If a man has the head of a raven, his days will be long.
> If a man has the head of a falcon, he will die in the bloom of life . . ..
> If a man has the head of an ass, he will be poor.

There is an unexplained logic behind these clauses, probably linked up with animal features which partially explain the apodoses, as is the case with Aristotle. His own remarks (*Hist. anim.* I.8) are as follows: "Menschen mit großer Stirn sind schwerfällig, die mit kleiner beweglich; bei denen sie flach ist, die sind leicht erregbar, bei denen sie ründlich ist, nachgiebig (ibid., p. 134). As for animal parallels with human physiognomy, Aristotle makes the following general observation (*Hist. anim.* IX.44): "Über die Charaktere der Tiere sind . . . am meisten die Unterschiede hinsichtlich Mut und Feigheit, danach die hinsichtlich Sanftmut und Wildheit auch bei den wilden Tieren zu berücksichtigen" (ibid., p. 143). The objectives are similar, namely to determine the characteristics or qualities of a person with certain features, but allowances must be made for differences in processes. Pseudo-Aristotle is attempting to explain how these omens work, while Babylonian scribes are confined by the rather fixed language and formats of omens, which by their nature are predictive, not descriptive.

100 See T. Barton, *Power and Knowledge: Astrology, Physiognomics, and Medicine under the Roman Empire* (Ann Arbor, MI: University of Michigan Press, 2005), p. 100, noting that Babylonian physiognomic omens are designed to predict the future rather than reveal a person's character, although she qualifies this by noting that omens may also foretell how a person will behave in future (p. 207). Oppenheim also saw Babylonian physiognomic omens as intended to note a person's qualities or "nature" (Oppenheim, *Ancient Mesopotamia*, p. 223).

used in tandem could be seen as a mechanism to evaluate a patient's general condition and outlook, from all possible points of view, in order to assess their state of health.[101]

PHARMACOLOGY AND ITS THEORETICAL UNDERPINNINGS

The study of plants and minerals was not invented by the Greeks. Already in eighth- to seventh-century Mesopotamia we have two texts, known as *Šammu Šikinšu*, "a plant – its characteristic," and *Abnu Šikinšu*, "a stone – its characteristic," which provide much useful information regarding the physical appearance and medical uses of medicinal plants and minerals.[102] The term *šiknu* in both instances corresponds to the use of Greek *dunameis* "powers, qualities" in relation to drugs,[103] although the information provided is not sufficient to establish a Dioscorides-like systematic pharmacology, nor can we show that Babylonian drugs (plants and stones) could have been used as opposites, as described by Galen ("wet" versus "dry," "cold" versus "hot," etc.). Nevertheless, the categories of "wet," "dry," "hot," and "cold," and the four colors associated with humors in Greek medicine (red, black, white, yellow) appear in standard descriptions of symptoms in the Diagnostic Handbook, and it is theoretically likely that Babylonian drugs were understood as being hot or cold or wet or dry, but we lack any explanatory treatise to provide us with relevant information. The Babylonians did not excel at expounding theory.[104]

Nevertheless, we can get some important clues from the rather laconically written drug commentary, *Šammu Šikinšu*, regarding commonly used medicinal plants, such as the following (ŠŠ I 19): "the plant is called 'fox-vine' (*karān šēlebi*), it is good for coldness (*kuṣṣi*) (caused by) the *lilû*-demon (i.e. an illness)."[105] The plant itself is often used for abdominal disorders or scorpion stings or applied against a skin lesion, or even used in urinary-tract ailments,[106] but it is unclear whether any of these ailments could be classified

[101] This is not so very far from the Greek idea of *hygeia*, which the *iatros* was also advised to take into account; see P. van der Eijk in this volume (chapter 16).

[102] These are not the only texts with explanatory information about drugs, but they are the most informative. See also the preliminary edition of the so-called Vademecum, an explanatory plant-list (see Scurlock, *Sourcebook*, pp. 273–80).

[103] See the remarks of P. van der Eijk in this volume (chapter 16).

[104] The Babylonian Talmud, however, does give a general rule for applying drugs, within the context of a lengthy passage providing medical recipes for diseases of the ear. The rule of thumb is, "moist" (drugs) for "dry" (ailments) and "dry" (drugs) for "moist" (ailments) (b. Ab. Zar. 28b, *rṭyb' lybšṭ' wybšt' ltyb'*). A bit later on the same page, the Talmud recommends "hot (drugs) for a scorpion (sting) and cold (drugs) for a wasp (sting)" (*ḥmymy lʿqrb' wqryry lzybwr'*).

[105] H. Stadhouders, "The Pharmacopoeial Handbook Šammu šikinšu. A Translation," *Le Journal des Médecines Cunéiformes* 19 (2012), 1–21, 2, and H. Stadhouders, "The Pharmacopoeial Handbook *Šammu šikinšu* – An Edition," *Le Journal des Médecines Cunéiformes* 18 (2011), 3–51, 7.

[106] For convenience, see Chicago Assyrian Dictionary (henceforth "CAD"), K 202.

as "cold." The name of the plant following this particular entry is lost, but it is listed as being good for "heat" (*ana ummi*). A second group of plants described by *Šammu Šikinšu* (ŠŠ I 79'–86') all grow "on the surface of the water" (*ina pān mê*) or more specifically on "the reclaimed (dirt) of the sea where there are neither marsh-reed nor plants."[107] This group of water-based plants is mentioned as being effective against "shivers" and "sun-fever," perhaps because all of these plants are presumably "wet," and the usage instructions call for the plants to be "dried out." One could presume here that "sun-fever" (*ṣētu*) is both hot and dry and needs to be treated by plants which are cold and wet. Another plant, *kamkadu*, is useful against lesions "which give off sweat";[108] this plant would presumably have drying properties, and it is also synonymous with "field-clod-plant," which might well be a description of a drying substance.[109] This entire area requires much further investigation.[110]

Furthermore, there was a rich tradition of *Dreckapotheke* in Babylonian drug lore, most of which can be shown to be *Deckname* or secret names for ordinary plants or minerals, probably to discourage non-specialists from dabbling in medical recipes.[111] It is hardly coincidental that many of the drugs used in Greek recipes against epilepsy were described by O. Temkin as being "magical," including such exotic items as hare's rennet, weasel stomach and blood, ass's liver and hoof, lichen of horses, he-goat liver, seal's rennet, land turtle-blood, and stork dung, counterparts for which existed in Babylonian *Dreckapotheke*.[112] The point is often made that Babylonian scholars neglected to write theoretical treatises to explain their thinking,

[107] Stadhouders, "A Translation," p. 12 and Stadhouders, "An Edition," p. 5; this same plant is described in a commentary as a "panacea" (M. J. Geller, *Ancient Babylonian Medicine in Theory and Practice* (Chichester: Wiley-Blackwell, 2010), p. 169, 172, l. 18).

[108] Stadhouders, "A Translation," p. 17 and Stadhouders, "An Edition," p. 8.

[109] See CAD K 123 and A. Attia and G. Buisson, "BAM 1 et consorts en transcription," *Le Journal des Médecines Cunéiformes* 19 (2012), 22–50, 28 (= BAM 1 ii 53–5), in which *kamkadu* features in association with two other drugs to "dry out the surface of a lesion."

[110] Unfortunately, the only study of Mesopotamian plants is from the British scholar R. Campbell Thompson, whose *Dictionary of Assyrian Botany* is regularly cited by non-specialists, but the methodological basis for this work and the resulting botanical identifications are almost entirely unreliable and groundless. See R. C. Thompson, *A Dictionary of Assyrian Botany* (London: British Academy, 1949). Nevertheless, data from Babylonia reflect O. Temkin's comment on Greek pharmacology:

> The Greek physicians also tried to establish a scientific pharmacology and to explain the actions of the various drugs by their "qualities." Galen, above all others, laid down rules how these qualities could be recognized. Once the quality of a disease was known, it would be treated with drugs of the opposite qualities. Epilepsy, being "cold and moist," would need heating and drying substances, and since it was caused by a thick phlegm, the physician would prescribe phlegmagogues and "sharp" drugs able to "cut" the viscous humour (Temkin, The Fallen Sickness, p. 78).

[111] See Geller, *Ancient Babylonian Medicine*, pp. 155–7.

[112] See Temkin, *The Fallen Sickness*, pp. 78–9. Temkin's list also included items which could be associated with astral magic or medicine, such as amulet-stones found in the stomach of swallows at the waxing moon, or selenite stone found by night at the waxing moon, and although these particular items are not found in Babylonia, there is a rich tradition of such amulet-stones which could be compared; see A. Schuster-Brandis, *Steine als Schutz-und Heilmittel. Untersuchung zu ihrer*

but whatever logic was employed in this kind of explanatory text may reflect the forerunners to later Greek pharmacology.

CONCLUSION

The survey of Babylonian medicine, with its complex relationships between different healing disciplines involving recipes, incantations and rituals, and diagnosis/prognosis, makes us wonder again about Herodotus' brief comments on Babylonian healing arts. How could Herodotus get things so wrong? Why did Herodotus describe Babylonian diagnosis and therapy in such comic and seemingly inaccurate terms?

In fact, Herodotus' description of the Babylonian healer as a layman is not entirely groundless. The distinction between a physician and layman in the Babylonian world is rather blurred, in the absence of anything resembling a diploma,[113] while a Greek physician's reputation was partly based upon his skills as an orator, to convince patients of his superior medical abilities and craft.[114] Herodotus' observation that everyone passing by offered medical advice is clearly an exaggeration, but it could reflect a scenario in which physicians gathered in the public square looking for clients would recommend their favorite remedies. The fact that the Babylonian patient was treated in the street (or in a public place) probably reflects an acute state of health, a last resort at seeking relief from disease. In any case, the scenario described by Herodotus is not so very far from what Ludwig Edelstein has reconstructed for the practice of the Greek physician, who operated from a shop in the street and competed with other physicians and healers for passing trade; the same working conditions probably applied to the *asû* in contemporary Babylonia. Furthermore, Herodotus' statement that no one was allowed to pass by the patient without offering advice probably reflects a commonly repeated medical adage in all systems of ancient medicine, privately advising healers to avoid making prognoses in terminal or hopeless

Verwendung in der Beschwörungskunst Mesopotamiens im 1. Jt. v. Chr. (Münster: Ugarit Verlag, 2008).

113 There is no clear distinction in Babylonian between "professionals" and "laymen," as there is in Greek (*technitai* and *idiôtai*), but Babylonian academies distinguished between "those who knew" (*mūdû*) and those "who did not know" (*lā mūdû*), and many texts were expressly prohibited in colophons from being shown to the latter. Nevertheless, the absence of a diploma or its equivalent was common to both societies; for the training of Greek physicians, see Krug, *Heilkunst und Heilkult*, pp. 190–3, and possibly for Babylonian medical training, see I. L. Finkel, "On Late Babylonian Medical Training," in A. R. George and I. L. Finkel (eds.), *Wisdom, Gods and Literature, Studies in Assyriology in Honour of W. G. Lambert* (Winona Lake, IN: Eisenbrauns, 2000), pp. 137–223.

114 See Edelstein, *Ancient Medicine*, pp. 89–90 and 99–100, based upon the Hippocratic treatise *On Diseases I*, on the art of medical oratory.

cases; such reticence was not likely to have been acceptable to patients.[115] After surveying all options, it seems possible that Herodotus was describing a somewhat caricatured view of medicine of his own day, and that pre-Hippocratic and Babylonian medicine were perceived to have had much more in common than modern historians realize.

[115] See the comments of van der Eijk in this volume (chapter 16).

3

MESOPOTAMIAN MATHEMATICS

Jens Høyrup

The term "Mesopotamian mathematics" refers to the mathematical knowledge and the mathematically based practices of the cuneiform tradition from the mid-fourth millennium BCE until its disappearance around the beginning of the Common Era.[1] All dates in the following should thus be understood to be BCE (and according to the "middle chronology") when CE is not indicated explicitly.

The reference to the writing system is not peripheral. Throughout its history, the development and orientation of Mesopotamian mathematics was intimately bound up with written administration and the scribal craft, and all documentation we possess derives from documents written on clay tablets (here, as mostly, the mathematical regularities of buildings and other artifacts give little conclusive information about the kind of mathematical *knowledge* which was involved in their production).

Attentive reading of the written sources reveals, however, that the written tradition must have received important inspiration from traditions carried by non-scribal (and, at least until the advent of Aramaic alphabetic literacy in the first millennium, scarcely literate) specialists: surveyors, master-builders, traders, and/or similar groups. In all likelihood, these "lay" practitioners also borrowed from the literate tradition, but this is more difficult to document.

[1] The changing approach to the field and the increasing awareness that Mesopotamian mathematics has *a history* is described in Jens Høyrup, "Changing Trends in the Historiography of Mesopotamian Mathematics: An Insider's View," *History of Science* 34 (1996), 1–32. An exhaustive annotated bibliography until 1982 is Jöran Friberg, "A Survey of Publications on Sumero-Akkadian Mathematics, Metrology and Related Matters (1854–1982)," *Department of Mathematics, Chalmers University of Technology and the University of Göteborg* 17 (1982). A very detailed account of mathematical knowledge and techniques is Jöran Friberg, "Mathematik," in *Reallexikon der Assyriologie und Vorderasiatischen Archäologie*, vol. 7 (Berlin and New York: de Gruyter, 1990), pp. 531–85. Eleanor Robson has published *Mathematics in Ancient Iraq: A Social History* (Princeton, NJ and Oxford: Princeton University Press, 2008).

LONG-TERM DEVELOPMENTS

So-called protoliterate writing was created in Southern Mesopotamia (the later "Sumerian" area) after the mid-fourth millennium, in connection with the earliest formation of a bureaucratic state (understood as a social system characterized by an at least three-tiered system of control and by extensive specialization of social roles) headed by a temple institution. The root of the invention was an accounting system based on clay tokens (probably standing for various measures of grain, heads of livestock, etc.) that had been used in the Near East since the eighth millennium, and various transformations and extensions of this system introduced in response to the needs created by increasing social complexity.[2]

In the protoliterate period, metrological notations were created that depicted the traditional tokens. A notation for an "almost abstract" number may have been created by adaptation of the system of grain or hollow measures to existing spoken numbers, with basic signs for 1, 10, 60, and 3,600, and composite signs for 60·10 and 3,600·10. Sub-unit extensions of all metrologies, an administrative calendar, and a combined metrology for length and area measurement replacing older "natural" (ploughing or irrigation) measures may also be new creations.[3]

Mathematics was fully integrated with its bureaucratic applications – school texts are "model documents," distinguishable from real administrative documents only by lacking the name of a responsible official and by the prominence of nice numbers. But the integration was mutual: bureaucratic procedures, centered on accounting, were mathematically planned, for instance around the new area metrology and the calendar. The cognitive integration corresponds to social integration – the literate and numerate class seems to coincide with the stratum of temple managers.

The third millennium continued the mutual fecundation of administrative procedures and the development of mathematics (in a process whose details we are unable to follow). The reach of accounting systems increased gradually, and metrologies were modified intentionally so as to facilitate managerial planning and accounting. At the same time, there is a trend toward "sexagesimalization," expanding use of the metrological step factor 60 – see below, "Numbers, Number Systems, Tables, and their Computational Use."

[2] On the token system and its development, see for instance Denise Schmandt-Besserat, *Before Writing.* I. *From Counting to Cuneiform* (Austin, TX: University of Texas Press, 1992).

[3] A broad summary of fourth- and third-millennium mathematical techniques (including the details of metrologies) is Hans J. Nissen, Peter Damerow, and Robert Englund, *Archaic Bookkeeping: Writing and Techniques of Economic Administration in the Ancient Near East* (Chicago, IL: Chicago University Press, 1993). The interplay between state formation and the shaping of mathematical techniques and thought is analyzed in Jens Høyrup, *In Measure, Number, and Weight. Studies in Mathematics and Culture* (New York: State University of New York Press, 1994), pp. 52–7, 68–74. Pages 45–87 of the same volume may serve as a general reference (with extensive bibliography) for the links between statal bureaucracy, scribal craft and culture, and the transformations of mathematics until the mid-second millennium.

Around 2600, however, when a distinct scribal profession emerged, numeracy and literacy outgrew the full cognitive subservience to accounting and management. For the first time, writing served to record literary texts (proverbs, hymns, and epics); and we find the first instances of "pure" or (better) "supra-utilitarian" mathematics – mathematics starting from applicable mathematics but going beyond its usual limits. It seems as if the new class of professional intellectuals set out to test the potentialities of the professional tools – the absolute favorite problem was the division of very large round numbers by divisors that were more difficult than those handled in normal practice.[4]

The language of the protoliterate texts is unidentified, whereas the language of the third-millennium southern city states was certainly Sumerian. Toward 2300, however, an Akkadian-speaking dynasty conquered the whole Sumerian region, and soon, for a while, the entire Syro-Iraqian area (Akkadian is a Semitic language, later split into the Babylonian and Assyrian dialects; names show it to have been present in the area at least since 2600). Sumerian remained the administrative language (and hence the language of scribal education), but new problem types suggest inspiration from a lay, possibly non-Sumerian surveyors' tradition – area computations that are very tedious unless one knows that $\square(R-r) = \square(R) + \square(r) - 2\sqsubset\!\sqsupset(R, r)$ ($\sqsubset\!\sqsupset$ and $\square$ stand for rectangle and square, respectively), and the bisection of a trapezium by means of a parallel transversal.

The twenty-first century is of particular importance. After a breakdown of the "Old Akkadian" empire, a new territorial state ("neo-Sumerian" or "Third Dynasty of Ur") established itself in 2112. A military reform under king Šulgi in 2074 was followed immediately by an administrative reform, in which scribal overseers were made accountable for the outcome of every $^{1}/_{60}$ of a working day of the labor force allotted to them according to fixed norms; at least in the South, the majority of the working population was apparently subjected to this regime, probably the most meticulous large-scale bureaucracy that ever existed.

Several mathematical tools were apparently developed in connection with the implementation of the reform (all evidence is indirect): a new book-keeping system – not double-entry book-keeping, but provided with similar built-in controls; a place-value system with base 60 used in intermediate calculations; and the various mathematical and technical tables needed in order to make the place-value system useful (described below, p. 64).

[4] Two specimens with divisor 7 are analyzed in Jens Høyrup, "Investigations of an Early Sumerian Division Problem, c. 2500 BC," *Historia Mathematica* 9 (1982), 19–36. A similar problem with divisor 33 from Ebla in Syria (whose mathematics was borrowed from Sumer) is analyzed in Jöran Friberg, "The Early Roots of Babylonian Mathematics. III: Three Remarkable Texts from Ancient Ebla," *Vicino Oriente* 6 (1986), 3–25.

No space seems to have been left to autonomous interest in mathematics; once again, the only mathematical school texts we know are "model documents."

For several reasons (among which were probably the exorbitant costs of the administration) even the Ur-III state collapsed around 2000. A number of smaller states arose in the beginning of the succeeding "Old Babylonian" period (2000 to 1600), all to be conquered by Hammurapi around 1760. Without being a genuine market economy, the new social system left much space to individualism, on the socio-economic as well as the ideological level. In the domain of scribal culture, this individualism expressed itself in the ideal of "humanism" (*sic* – n a m-l ú-u l ù, Sumerian for "being human"): scribal *virtuosity* beyond what was needed in practice. This involved the ability to read and speak Sumerian, now a dead language known only by scribes, as well as supra-utilitarian mathematical competence.

The vast majority of Mesopotamian mathematical texts come from the Old Babylonian school (teachers' texts or copies from these, except for the training of simple calculation and copies of tables not student production as all third-millennium specimens). They are in Akkadian (notwithstanding sometimes heavy use of Sumerian word signs), another indication that the whole genre of "humanist" mathematics had no Ur-III antecedents. Its central discipline was a geometrically based second-degree "algebra," probably inspired from a collection of geometrical riddles circulating among lay, Akkadian-speaking surveyors (to find the side of a square from [the sum of] "the side and the area" or from "all four sides and the area," etc.), but transformed into a genuine mathematical discipline and a general analytical technique.

A first classification divides the text corpus into table texts and problem texts. The second category can be subdivided in different ways: into (1) theme texts, whose problems have a common theme; (2) anthology texts which have no common theme; and (3) single-problem texts. Alternatively it can be subdivided into (I) procedure texts that teach how to obtain a solution; and (II) catalog texts listing mere problem statements (most catalogs are theme texts). It is noteworthy that anthology texts, even if mixing different kinds of mathematics, do not mix mathematics with other topics (not even sacred numerology); Old Babylonian mathematics was clearly a cognitively autonomous field.

Some of the texts come from excavations, but most from illegal diggings.[5] For these, provenience and dating must be derived from paleography, orthography, and characteristic differences in terminology. In a region

[5] The basic text editions are O. Neugebauer, *Mathematische Keilschrift-Texte*, 3 vols. (Berlin: Julius Springer, 1935–7) = MKT; O. Neugebauer and A. Sachs, *Mathematical Cuneiform Texts* (New Haven, CT: American Oriental Society, 1945) = MCT; and E. M. Bruins and M. Rutten, *Textes mathématiques de Suse* (Paris: Paul Geuthner, 1961) = TMS. MKT and MCT are very careful editions, TMS alas not. Only TMS contains archaeologically excavated texts. Single texts with

encompassing the former Sumerian South, the Center (Babylon and surroundings) the Center-to-North-East (Ešnunna), and even the eastern periphery (Iranian Susa), the global character of Old Babylonian mathematics is largely the same (from the Assyrian North, never dominated by Ur III or Babylonia, no mathematical texts but only accounts are known). Close attention to language and procedures reveals, however: that the adoption of lay material has taken place simultaneously in Ešnunna and in the South; that pre- and post-Šulgi-reform Sumerian mathematics coexisted in eighteenth-century Ešnunna without being fully merged; that a number of schools tried to develop a strict terminological canon but did not agree in their choices; and that all texts that try to explain procedures abstractly and not only through paradigmatic numerical examples are close to the lay oral tradition – the school seems to have given up abstract formulation as pedagogically inefficient.[6]

Inner weakening followed by a Hittite raid put an end to the Old Babylonian state in 1600. A warrior tribe (the Kassites) subdued the Babylonian area, for the first time rejecting that managerial-functional legitimization of the state which, irrespective of suppressive realities, had survived since the protoliterate phase and made mathematical-administrative activity an important ingredient of scribal professional pride. The school institution disappeared, and scribes were trained henceforth as apprectices. Together, these events had the effect that mathematics disappears almost completely from the archaeological horizon for a millennium or more (one Kasssite problem text and one table text have been found; the problem text offering a sham solution to a very difficult problem seems to derive from the style of the Old Babylonian northern periphery); metrologies were modified in a way that would fit practical computation in a mathematically less sophisticated environment (e.g. making use of normalized seed measures in area mensuration; though no longer an object of pride, mathematical administration did not disappear).

Around the "Neo-Babylonian" mid-first millennium, mathematical texts turn up again, for instance concerned with area mensuration, the conversion between various seed measures, and some supra-utilitarian problems of the kind that had once inspired Old Babylonian "algebra." This and other features may reflect renewed interaction between the scribal and the lay traditions, which so far cannot be traced more precisely.

known provenience have been published by Taha Baqir, Jöran Friberg, and others, many in the journal *Sumer*.

6 The analysis is presented in Jens Høyrup, "The Finer Structure of the Old Babylonian Mathematical Corpus. Elements of Classification, with some Results," in Joachim Marzahn and Hans Neumann (eds.), *Assyriologica et Semitica. Festschrift für Joachim Oelsner anläßlich seines 65. Geburtstages am 18. Februar 1997* (Münster: Ugarit Verlag, 2000), pp. 117–77. The idea that traces of pre-Šulgi mathematics might be present in texts from the northern periphery was first proposed by Eleanor Robson in her dissertation from 1995, now published as *Mesopotamian Mathematics 2100–1600* BC. *Technical Constants in Bureaucracy and Education* (Oxford: Clarendon Press, 1999). The precise chronology for the adoption of various kinds of lay material, as far as it can be known, is investigated in Jens Høyrup, "A Hypothetical History of Old Babylonian Mathematics: Places, Passages, Stages, Development," *Gaṇita Bhārati* 34 (2012), 1–23.

One Neo-Babylonian text combines the sacred numbers of the gods with a metrological table. This breakdown of cognitive autonomy corresponds to what the texts tell us about their owners and producers (such information is absent from the Old Babylonian tablets); they identify themselves as "exorcists" or "omen priests" (another reason to believe that their practical geometry was borrowed from lay surveyors).

A final development took place in the Seleucid era (311 onwards). Even this phase is only documented by a few texts: some multi-place tables of reciprocals possibly connected to astronomical computation, though of no direct technical relevance; one theme text; an anthology text focusing on practical geometry; and an unfocused anthology text. The unfocused anthology text shows some continuity with the Old Babylonian tradition (including its second-degree "algebra") but also fresh developments (e.g., formulas for $\sum 2^n$ and $\sum n^2$). The theme text contains "algebraic" problems about rectangles and their diagonals, of which only one type is known from the earlier record, but where even this is solved in a different way. It seems to be a list of *new* problem types or procedures, either borrowed from elsewhere or freshly invented. The Seleucid texts make heavy use of Sumerian word signs, but in a way that sometimes directly contradicts earlier uses. To some extent at least they represent a new translation into Sumerian of a tradition that must have been transmitted outside an erudite scribal environment.

In connection with the creation of a planetary astronomy based on arithmetical schemes, the Neo-Babylonian period (in particular the Seleucid phase) developed a set of highly sophisticated numerical techniques; these are dealt with in chapter 4 by John Steele, in this volume.

NUMBERS, NUMBER SYSTEMS, TABLES, AND THEIR COMPUTATIONAL USE

The original number system was based on specific signs for 1, 10, 60, 600, 3,600, and 36,000, multiples of which were produced by repetition in fixed patterns (· , ·· , ··· , :: , ·.·.· , etc.). In the third millennium, adjunction of the sign g a l, "great," allowed upwards extension of the system by a factor 60, whereas the sign gín, borrowed from weight metrology, was used in the sense of $^1/_{60}$ (the same tricks were used to expand the reach of metrological sequences). From the Old Akkadian epoch onward, calculators can be seen to have experimented with the system, thus approaching the place-value principle – but all the relevant texts commit errors, thus showing that no place-value *system* was yet available.[7]

[7] See Marvin A. Powell, "The Antecedents of Old Babylonian Place Notation and the Early History of Babylonian Mathematics," *Historia Mathematica* 3 (1976), 417–39.

The *system* seems to have been created in the wake of the Ur-III administrative reform. It employed the traditional sign for 1 for any integer power 60^n, and the sign for 10 for any $10 \cdot 60^n$ – still with repetitions in fixed patterns to express 2, 3, . . .,9, and 20, 30, . . ., 50. It was a floating-point system, with no indication of absolute order of magnitude, as the slide rule engineers would use until some decades ago. Nor were "intermediate zeroes" indicated. For both reasons, the notation could only be used for rough work – final results had to be inserted in the documents in the traditional, unambiguous notation.

The place-value notation did not facilitate additions and subtractions – these were performed on a calculating board;[8] the reason it was introduced was the importance of multiplications in Ur-III planning and accounting. If, e.g., the labor needed to produce a wall of given dimensions of bricks of a given type was to be found, one ("metrological") table would translate a thickness measured in cubits and fingers into the standard length unit (a "rod" ≈ 6m), after which the volume of the wall could be found in standard units. A "technical" table of "constant factors" would tell the number of bricks of the type in question per unit volume, another the number of bricks produced by a worker per day, a third the number carried a given distance per man-day, etc. The total consumption of labor could then be found by means of multiplications and divisions.[9]

Beyond metrological conversion tables and tables of technical constants, the system depended on the availability of multiplication tables and of tables of reciprocals (to be learned by heart in the scribe school) – the latter because division by n was performed as a multiplication by $^1/_n$. The important step in the invention of the place-value system was thus not the inception of the idea, which had been in the air for centuries; it will have been the government decision to have it spread in teaching and to produce (mass-produce!) the tables needed for its implementation.

Once introduced, the place-value system could survive in less bureaucratic settings. It became the standard system of Old Babylonian mathematical texts (only occasionally will the units of "real" life turn up in statements or final results) and of late Babylonian mathematical astronomy. It is quite uncertain whether the Indian decimal place-value system for integers depends on it, but it is undisputed that it was taken over in the minute-second fractions of Greek and later astronomy, whence it inspired the introduction of decimal fractions.

It may seem a drawback of the Babylonian division method that it only works for "regular" divisors of the form $2^p \cdot 3^q \cdot 5^r$ (p, q, and r positive, negative,

8 See Jens Høyrup, "A Note on Old Babylonian Computational Techniques," *Historia Mathematica* 29 (2002), 193–98, and Christine Proust, "La multiplication babylonienne: la part non écrite du calcul," *Revue d'Histoire des Mathématiques* 6 (2000), 293–303.

9 The most thorough treatment to date of the technical factors and their use as reflected in mathematical texts is Robson, *Mesopotamian Mathematics 2100–1600 BC*.

or 0). In practice this was no trouble, firstly because all metrological step factors were regular, and secondly because the margin on technical factors was always large enough to allow representation by a simple regular number. Factors that might turn up as divisors were always chosen thus.

Beyond the tables already mentioned, other arithmetical tables occur: n^2 (with inversions as $\sqrt{N}$, where N itself is square), and the inversions of n^3 and $n^2 \cdot (n+1)$. Tables of squares (*viz* square areas expressed in metrological units) go back to before the mid-third millennium and thus antedate the place-value system by ca. 500 years.

GEOMETRY

Very few third-millennium texts reveal the actual mathematical knowledge and procedures that went into their results. The post-Old-Babylonian texts at our disposal are also too few to suggest any global picture. For these reasons, this and the following two sections deal primarily with the mathematics and the mathematical thought of the Old Babylonian period.

In geometry, no concept of the *quantifiable* angle existed. In order to find the area of a rectangle the Babylonians would multiply the length with the width – as mentioned, the area metrology had been adapted to this already in the fourth millennium. When dealing with near-rectangular quadrangles they would choose as length and width the legs of an approximately right angle (as opposed, we may say, to a "wrong" angle). If opposite sides were slightly different, average length would be multiplied by average width (the "surveyors' formula"; also since the fourth millennium). This would always yield too large results, but with one known exception it was only used in practical mensuration when the error was negligible (when used as a mere pretext for supra-utilitarian problems in the Old Babylonian school, the formula might be employed in cases where it is blatantly absurd).

The area of approximately right triangles was found as the product of the bisected width with "the length" – as opposed to "the long length," i.e., the hypotenuse. More complex shapes would be split up into quasi-rectangles and quasi-right triangles (this is seen in Ur-III field plans). A Seleucid text computes the height of an equilateral trapezium; a text from Old Babylonian Susa suggests that the same *could* be done in earlier times when regular polygons were investigated.

The absence of the notion of the quantifiable angle did not prevent the understanding of similarity relations. It was also routinely used that the areas of similar figures are to each other as the squares on the linear dimensions.

It was known that the square on the diagonal of a rectangle augmented by the doubled area equals the square on the sum of the sides, whereas the squared diagonal minus the doubled area equals the square on their difference – and, probably as a sequel, that the squared diagonal itself equals the sum of the squared sides. The latter, of course, is what we know as the "Pythagorean theorem."

The fundamental circle parameter was the perimeter p – the area was found as $^1/_{12}p^2$, and the diameter as $^1/_3p$. In one text group from the northern region (in general close to the lay tradition, where both the separate treatment of the semicircle and the very same formula turn up in later ages) the area of the semicircle is found as $^1/_4$ of the product of diameter and arc.

In volume metrology, the area units were thought of as provided with a "standard thickness" of 1 cubit. In order to determine a prismatic or cylindrical volume, the calculator would first find the base (with this implicit thickness) and then "raise it to," i.e, multiply it with, the height. This operation was so important that "raising" became the standard term for any multiplication which was based on similar considerations of proportionality (only concrete repetition and the laying-out of rectangular areas employ other terms); all multiplications with factors taken from metrological tables or tables of technical constants were thus "raisings."

The volume of a truncated cone was calculated as the height times the mid-cross-section, that is, as that of a cylinder with the average diameter. In one case, the volume of a truncated pyramid is determined as the average base raised to the height – in another, the correct value is found, whether from a correct formula or not is unclear (a correct formula *can* be derived from relatively simple intuitive arguments).

The simple area and volume formulas are without doubt based on such intuitive insights. The restricted use of the "surveyors' formula" indicates that it was understood to be only an approximation, but nothing suggests any precise idea as to the importance of the error; probably the Babylonians would see no difference between this kind of approximation and the treatment of an inevitably uneven terrain as if it were a perfect plane.

Formal demonstration seems to be absent from Babylonian geometry. There was a certain interest in striking geometrical configurations – e.g., systems of concentric squares. Reflections on a concentric two-square system and on the appurtenant "average square" may have led to the discovery of how to bisect a trapezium by a parallel transversal. Apart from this, the only important kind of supra-utilitarian geometry was the area technique which has become known as "Babylonian algebra."

"ALGEBRA" AND OTHER "PURE" PURSUITS

When it was discovered in the late 1920s that the sequence of numbers in certain texts corresponds to the solution of second-degree equations, much of the technical terminology was still uninterpreted. It was assumed – and generally accepted for 60 years – that the underlying conceptualizations were arithmetical; that the operations involved were therefore numerical

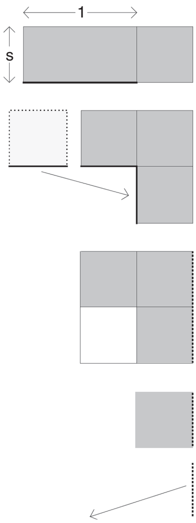

Figure 3.1. The procedure of BM 13901 nº 1.

additions, subtractions, multiplications, and extractions of roots; and that the persistent references to lengths, widths, and areas were nothing but metaphors for numerical unknowns and their products.[10]

Close attention to the vocabulary and the organization of the texts demonstrates, however, that two presumed additions are kept strictly apart; that there are two different subtractive operations; that two different "halves" are distinguished; and that "multiplications" are four in number. All of this concerns *concepts* – several of the concepts are covered by two or more synonymous *terms*. This makes no sense in the arithmetical interpretation, but everything becomes obvious if we take the words of the texts (lengths, widths, squares, areas) seriously. Babylonian "algebra" turns out to be a cut-and-paste technique which manipulates measurable line segments and areas in analytic processes which, in their numerical steps, correspond to the procedures of our equation algebra.[11] As an example we may look at the simplest of all mixed second-degree problems: the sum of a square area and the side is 45′ (i.e., $45 \cdot {}^1/_{60} = {}^3/_4$). The sequence of numerical steps in the solution (with added indications of absolute magnitude, ′ indicating minutes or sixtieths, ′′ seconds) is as follows: 45′ — 1 — 1 — 30′ (= $^1/_2$) — 30′ — 15′ (= $^1/_4$) — 45′ — 1 — 1 — 30′ — 1 — 30′.[12] What goes on can be followed in Figure 3.1. At first the side is represented by

[10] The discovery and the development of interpretations is analyzed in Høyrup, "Changing Trends," pp. 1–10.

[11] See Jens Høyrup, *Lengths, Widths, Surfaces: A Portrait of Old Babylonian Algebra and Its Kin* (Studies and Sources in the History of Mathematics and Physical Sciences; New York: Springer, 2002).

[12] BM 13901 nº 1, in MKT III, 1.

a rectangular area ⊏⊐(1,*s*), which is glued to the square □(*s*). Its length 1 is bisected, and the outer $^1/_2$ is moved so as to span with the $^1/_2$ that remains in place a quadratic complement with the area $^1/_4$. This is joined to the gnomonic area $^3/_4$ consisting of the square and the bisected rectangle. The resulting square has the area 1, and thus also the side 1. The $^1/_2$ that was moved is detached from this 1, and $^1/_2$ remains as the side of the original square. The method, as we see, is analytical in the same sense as equation algebra: the unknown side *s* is treated *as if* it were known, and the complex relation subjected to manipulations until *s* appears in isolation.

This is one of those original surveyors' riddles that were apparently borrowed by the early Old Babylonian scribe school. In the scribe school it was only one of many problems dealing with areas and segments. In non-normalized cases, a proportional scaling of figures along one dimension was used along with the cut-and-paste procedures.

Line segments and areas constituted the basis of the technique. They could then be used to *represent* entities of other kinds: numbers from the table of reciprocals, prices – or segments might represent areas or volumes. The technique thus served as a general tool for finding unknown entities involved in complex relationships. Even in this sense, it was similar to equation algebra; its "basic representation" was not numerical, it is true, but the segments and areas of this representation were as *functionally abstract* as the numbers of equation algebra. The closest kin of Babylonian "algebra," however, is pre-Viète algebra: it was and remained a *technique*, and was never associated with any algebraic theory about solvability conditions or the classification of problems (classification was based on geometrical object and not on algebraic type, as revealed by the organization of theme texts). Nor was it used to solve "real-life" problems – no single practical problem presenting itself to Babylonian calculators was of the second degree. The only "practical" purpose of treating second-degree problems in school was as a pretext for training calculation with sexagesimal numbers (much as second-degree equation algebra has served in the schools of recent centuries to train the manipulation of algebraic letter symbols).

Practical first-degree problems were solved without recourse to algebraic techniques, often by variants of the "single false position" (also used in homogeneous problems of the second and third degree). However, second-degree "algebraic" systems might include a genuine first-degree equation, of the type "the sum of the length and the width, from which $^1/_4$ of the width is detached, is 45"). A couple of texts discuss such equations and their transformation, identifying most pedagogically the coefficients of length and width and the contribution of each to the sum.

Some higher-degree problems (of biquadratic and similar types) were solved by means of the algebraic technique. Mixed third-degree problems also turn up – e.g., to find the side of a cubic excavation if the sum of the volume and the base is known. Here the algebraic technique would forsake.

Instead the calculator resorted to a combination of false-position considerations and factorization or (in the case just mentioned) the table of $n^2 \cdot (n+1)$. The trick is elegant but *only works because a simple solution is known in advance to exist* (*all* school problems were constructed backwards from known solutions).

The cubic problems are found in theme texts together with other "excavation" problems of the first or the second degree, solved on their part with algebraic methods. As regards their method, however, they are rather linked with another kind of supra-utilitarian mathematics: investigations of the properties of the regular numbers of the sexagesimal place value system. In simple cases, it involved factorizations, continued multiplication products of simple factors, etc. The high point is a tabulation, not directly of Pythagorean triplets a–b–c but of ???–z^2–b–c, where ??? stands for one or more missing columns, and $z=\frac{c}{a}$. All sets $(\overline{b},\overline{c}) = (\frac{t'-t}{2}, \frac{t'+t}{2})$ are listed for which $\sqrt{2}-1<t<5/9$, t being the quotient between two regular integer numbers no greater that 125, $t' = \frac{1}{t}$.[13]

The headings of the *b*- and *c*-columns speak about width and diagonal, and it is thus certain that a geometric rectangle and its diagonal are involved. Apart from that, the purpose of the text is obscure. As shown by Friberg, it is not the result of a pure number-theoretical investigation. Instead, he proposes, it might serve as a tool for finding an array of data that would permit some mathematical problem (e.g. concerning right triangles) to be solvable. Unfortunately, Old Babylonian problems always have very simple solutions, and consecutive problems often stick to the same solution; available evidence therefore speaks against this proposal, but no more convincing alternative is at hand. The text adds an important shade to our knowledge about *what* the Babylonians could do but so far nothing to our understanding of *why* they would do it.

"MATHEMATICS" OR "COMPUTATION"? A GLOBAL CHARACTERIZATION

In the Old Babylonian period, mathematics was a cognitively autonomous subject, and it may therefore be considered legitimate to speak of it precisely as *mathematics*, as done until now. In contrast, the term "mathematician" appears nowhere. All we know about the authors of the mathematical texts with some certainty (namely from the format of the texts) is that they will have been teachers of future scribes (even though the sophisticated matters were

[13] The tablet has been much discussed in the literature. Analysis and summary of earlier work is found in Jöran Friberg, "Methods and Traditions of Babylonian Mathematics. Plimpton 322, Pythagorean Triples, and the Babylonian Triangle Parameter Equations," *Historia Mathematica* 8 (1981), 277–318. A new profound analysis is Eleanor Robson, "Neither Sherlock Holmes nor Babylon: A Reassessment of Plimpton 322," *Historia Mathematica* 28 (2001), 167–206.

hardly tought to more than a minority of these). Much of what we find in the texts is supra-utilitarian – but its ultimate legitimacy always rests on its link to scribal activity. The scribe, however, when using mathematics, would always be interested in *finding a number*, not, e.g., in geometrical regularities; artists might have this interest, but with the exception of the abovementioned concentric squares nothing permits us to link patterns with mathematically interesting symmetries to the mathematical texts.

Strictly speaking, Old Babylonian (and, in general, Mesopotamian) mathematics might therefore better be characterized as *computation*; instead of "mathematicians" we should speak of "calculators" and "teachers of calculation"; supra-utilitarian activities represent "pure calculation" rather than "pure mathematics." The ultimate interest in finding a number is of course also a characteristic of most present-day applications of mathematics; but it remains a feature which distinguishes both the Mesopotamian and the contemporary calculating orientation from the *investigation of the properties of mathematical objects* which (since the Greeks) constitutes our ideal type of mathematics proper.

The italicized passage contains a veiled reference to another difference between our ideal type and the Mesopotamian type: "investigation." In principle, theoretical mathematics has *the problem* as its core, and then sets out to construct methods and a conceptual apparatus that permit its solution. The same characteristic holds for *applications* of mathematics (Mesopotamian as well as contemporary), with the difference that the defining problem is no mathematical problem. The core of Mesopotamian supra-utilitarian mathematics, on the contrary, was always *the method*. When the mid-third-millennium calculators were testing the potentialities of the professional tools, these tools were the starting point, and the aim was to find out how far they would reach. Similarly, the "scribal humanism" of the Old Babylonian period, aiming at handling with *virtuosity* the tools and techniques of the scribe, would be centered on these.

This does not preclude the practical existence of mathematical research, in the form of a search for problems that could be treated by available techniques and tricks. The difference between the surveyors' riddles and the algebraic discipline created in the school is indeed the outcome of this kind of search. Nor did it preclude the invention of new techniques of scarce practical utility; such inventions might be needed if new problem types were to be transformed so as to be solvable – the "quadratic completion" used to solve mixed quadratic problems is an example, already conceived in the lay surveyors' environment and then adopted into the early Old Babylonian scribe school (which knew it as "the Akkadian [method]"). Once devised, such techniques would themselves become part of the stock of professional tools, and serve in the search for problem types that might now be solved – as the quadratic completion became the basis for the whole fabulous development of second-degree algebra in the school.

REVERBERATIONS

After the discovery of the Babylonian second-degree algebraic in the late 1920s, Neugebauer proposed that the geometry of *Elements* II (characterized by Zeuthen as "geometric algebra" already in the 1880s) should be understood as a geometrical translation of the supposedly numerical algebra of the Babylonians, prompted by the discovery of irrationality and the ensuing "foundation crisis" of Greek mathematics.

The foundation crisis turned out to be a projection of the 1920s on Greek antiquity, and even the translation theory proved problematic as it was formulated. *Elements* II solve no problems, at most they can be said to prove algebraic identities of a kind that Babylonian algebra seemed to be based on. Worse was the disappearance of the main stock of Babylonian algebra more than a millennium before the creation of Greek geometry and the failing evidence that any Greek mathematician knew about Babylonian mathematics.

The geometric reinterpretation of the Babylonian technique transforms the question: Euclid's diagrams coincide with those of which the Babylonians had made use (II.6 thus with the procedure shown above), and his proofs may be said to provide a "critique" of the Babylonian procedures – verification of their legitimacy and investigation of the conditions under which they are valid. But it does not invalidate the second objection to Neugebauer's thesis.

Comparative analysis of the Babylonian material and a number of later sources – mostly treatises on practical geometry containing supra-utilitarian material, many from the Islamic Middle Ages but others belonging to classical antiquity or to the stock of Italian borrowings from lost Arabic sources – now allows us to delineate a new scenario.[14]

The original stock of quasi-algebraic surveyors' riddles can be said with fair certainty to have encompassed at least the following problems on a single square (area A, side s, "all four sides" ${}_4s$; Greek letters indicate given numbers):

$$A \pm s = \alpha \ // \ A + {}_4s = \beta \ // \ A = {}_4s.$$

On rectangles (length l, width w, all sides ${}_4s$, diagonal d) the following can be identified:

[14] The details of the scenario and fairly full arguments from the sources are found in Jens Høyrup, "On a Collection of Geometrical Riddles and Their Role in the Shaping of Four to Six 'Algebras'," *Science in Context* 14 (2001), 85–131. Supplementary material is in "Hero, Ps.-Hero, and Near Eastern Practical Geometry. An Investigation of *Metrica*, *Geometrica*, and other Treatises," in Klaus Döring, Bernhard Herzhoff, and Georg Wöhrle (eds.), *Antike Naturwissenschaft und ihre Rezeption, Band 7* (Trier: Wissenschaftlicher Verlag Trier, 1997), pp. 67–93. The changing and incoherent uses of the notion of "geometric algebra" since Zeuthen by friends and foes are analyzed in Jens Høyrup, "What is 'Geometric Algebra', and What Has It Been in Historiography?" *AIMS Mathematics* 2 (2017), 128–60.

$$A = \alpha, l \pm w = \beta \;//\; A + (l \mp w) = \alpha, l \pm w = \beta \;//\; A = \alpha, d = \beta,$$

and seemingly also

$$A = l + w // A =_4 s.$$

On two squares, finally,

$$A_1 + A_2 = \alpha, s_1 \pm s_2 = \beta \;//\; A_1 – A_2 = \alpha, s_1 \pm s_2 = \beta.$$

The lay tradition – whose geographical extension may have outranged Mesopotamia – survived the collapse of the Old Babylonian scribe school, and conserved its stock of riddles. It may have borrowed from the scribe school, but only marginally, and never anything "algebraic" of a more advanced character than the original riddles. Some of its characteristic riddles turn up in Diophantus' *Arithmetica* I, some are found in pseudo-Heronian or agrimensor treatises, and some are referred to in the *Theologoumena arithmeticae* – enough, indeed, to demonstrate that Greek theoretical geometry would have had no difficulty in running into the tradition (whether during contacts with Syro-Phoenician practitioners or in Egypt, to where it may have been brought by military surveyors or tax collectors in the wake of the Assyrian or the Persian conquests). It seems that some geometers did so before Theodoros' time (thus probably in the fifth century) and submitted the old procedures to a "critique" whose results turn up in *Elements* II, propositions 1–10; all of these, indeed, are related to the basic riddles or to the formulas $\square(R \pm r) = \square(R) + \square(r) \pm 2 ⊏⊐(R, r)$, apparently known already in the Old Akkadian school. In contrast, nothing in Euclid relates to the particular creations of the Old Babylonian scribe school: ample and intricate use of coefficients; the treatment of biquadratics and other higher-order problems; and the scaling of non-normalized problems (*Elements* VI.28–9 is likely to represent an independent though similar generalization).

In the Islamic world, the tradition and at least some of the riddles are still encountered around 1200 CE. In the ninth century CE, the cut-and-paste procedures were borrowed by al-Khwārizmī for his demonstrations of the algorithms of *al-jabr*, and thus were also adopted as a core constituent of Latin *algebra*.

References to the old tradition are also found in Mahāvīrā's ninth-century *Gaṇita-Sāra-Sangraha*. Since they do not correspond to what occurs in Islamic sources, Mahāvīrā is likely to have drawn on the Jaina tradition. He is thus a witness of a possible link between the Near Eastern tradition and Indian medieval algebra – a link which is invisible in the numerical algebra of Āryabhata and Brahmagupta. If not directly, then at least through this lay tradition, the Babylonian algebra discovered by Neugebauer and his collaborators thus had even wider repercussions than he ever dared imagine in print.

4

BABYLONIAN AND ASSYRIAN ASTRAL SCIENCE

John M. Steele

Several thousand cuneiform tablets attest to the practice of astronomy and the related astral sciences of celestial divination, astrology, and calendrics in Assyria and Babylonia over the final two millennia BCE. Modern study of these tablets began in the last quarter of the nineteenth century and has continued up to the present. As a result of these studies, we now have a good understanding of the internal content of the main groups of cuneiform astronomical and related texts, of the mathematical structure of the computational astronomy of the last four centuries BCE, and of the traditional form of celestial divination as it was practiced in the Neo-Assyrian period (eighth and seventh centuries BCE).[1] Research over the past fifteen years or so has seen significant advances in our understanding of the development of predictive and computational astronomy from the seventh to the fourth century BCE, late forms of astrology, and observational practice and record keeping during the Late Babylonian period (ca. 750 BCE–75 CE). Recent research has also begun to address questions concerning the social history of astronomy and the relationship between the astral sciences and other branches of scholarship in Assyria and Babylonia.[2] Nevertheless, there are still many unstudied tablets that are known to deal with the astral sciences, and considerable areas within Mesopotamian astronomy that remain only imperfectly understood. Thus, the history of Mesopotamian astronomy is liable to considerable changes over the next few decades as further tablets are studied.

[1] H. Hunger and D. Pingree, *Astral Sciences in Mesopotamia* (Leiden: Brill, 1999).

[2] F. Rochberg, "Scribes and Scholars: The *ṭupšar Enūma Anu Enlil*," in J. Marzahn and H. Neumann (eds.), *Assyriologica et Semitica: Festschrift für Joachim Oelsner* (Münster: Ugarit-Verlag, 2000), pp. 359–76; E. Robson, *Mathematics in Ancient Iraq: A Social History* (Princeton, NJ: Princeton University Press, 2008), pp. 214–62; M. Ossendrijver, "Exzellente Netwerke: Die Astronomen von Uruk," in G. J. Selz and K. Wagensonner (eds.), *The Empirical Dimension of Ancient Near Eastern Studies – Die empirische Dimension altorientalischer Forschungen* (Vienna: LIT, 2011), pp. 631–44.

In what follows I attempt to present a diachronic history of astronomy and the astral sciences in Assyria and Babylonia. Before beginning this history, however, it should be noted that there are several problems with this approach. The most serious problem concerns the preserved source material. For various reasons (some of which I will discuss in the next section)[3] almost all preserved cuneiform tablets dealing with the astral sciences date to only two time periods: eighth and seventh century BCE Assyria; and Babylonia between the fourth century BCE and the first century CE. A comparatively small number of astronomical tablets from Babylonia are known from the eighth to the fifth century BCE, and only a handful of tablets from before the eighth century BCE are known from either Assyria or Babylonia. Our knowledge of the astral sciences in the earlier period relies for the most part on late copies of works that it is believed were composed in the second or early first millennium BCE. It is quite possible – indeed likely – that what is preserved in late copies is only a small part of the astronomical activity, which took place at that earlier time. Furthermore, the material preserved in late copies is by definition material that was considered worth copying by later scribes, and as a result may provide a biased view of early astronomy. A second issue that needs to be borne in mind about a diachronic history of Mesopotamian astronomy is that there is considerable temporal overlap between different astronomical traditions within Mesopotamia. For example, the astronomy of the Neo-Assyrian period overlapped with the beginning of Late Babylonian astronomy by a century or so, but there are clear differences in the practice of astronomy in these two cultures, notwithstanding that Babylonia was part of the Assyrian empire at the time.

Before turning to this diachronic history of Babylonian and Assyrian astral science it is necessary to briefly address some background issues: the scribal and archival context of cuneiform astral science; the mathematical techniques and conventions found in Babylonian and Assyrian astronomy; and the calendar.

SCRIBAL AND ARCHIVAL CONTEXT

The majority of cuneiform tablets known to deal with the astral sciences were recovered during the last quarter of the nineteenth and the first years of the twentieth centuries. These tablets were acquired by European and, to a lesser extent, American museums either from unscientific archaeological excavations undertaken on behalf of those museums or, more often, through purchases made by the museums from antiquities dealers. As a consequence, little is

[3] See also D. Brown, "What Shaped Our Corpuses of Astral and Mathematical Cuneiform Texts," in F. Bretelle-Establet (ed.), *Looking at it from Asia: The Processes that Shaped the Sources of History of Science* (New York: Springer, 2010), pp. 277–303.

known about the archaeological context of these tablets; in some cases not even the city where a tablet was found is known. There are two major exceptions to this general situation: tablets recovered from Assyrian cities dating to the eighth and seventh centuries BCE, and tablets excavated from the site of Uruk in southern Babylonia dating from the fifth to the first century BCE.

The Assyrian tablets containing texts dealing with the astral sciences were found at the sites of Nineveh, Kalḫu, and Assur in the Assyrian heartland and from Ḫuzirina in Anatolia. A large proportion of the Assyrian tablets were recovered from the Neo-Assyrian capital Nineveh, mainly from the palace of Assurbanipal. These tablets include the correspondence sent by scholarly advisors to the kings Esarhaddon and Assurbanipal, copies of works such as *Enūma Anu Enlil* and MUL.APIN (for these texts, see below) from the so-called "library of Assurbanipal,"[4] as well as many tablets dealing with individual astronomical topics. Smaller numbers of tablets including copies of tablets of *Enūma Anu Enlil* and MUL.APIN were discovered at the site of the Nabu temple in Kalḫu,[5] and from a library in the house of a family of *šangû*-priests in Ḫuzirina.[6] A significant number of astronomical tablets, mainly copies of *Enūma Anu Enlil*, have also been recovered from sites in Assur. It seems, therefore, that we have evidence for two institutional contexts, within which astronomy was practiced during the Neo-Assyrian period: the palace and the temples. The preserved texts indicate that there was a considerably broader range of astronomical and astrological activity in the context of the palace than in the temples.

From Uruk we have astronomical and astrological tablets for which no archaeological context is known, tablets with limited provenience information, and tablets which were found in situ for which detailed findspot information is available.[7] The latter tablets were found in two contexts: private houses and the *Bīt Reš* temple. Detailed archaeological information concerning a very small number of astronomical tablets from Babylon and Sippar that were excavated scientifically provide further evidence that astronomical activity during the Achaemenid, Seleucid, and Parthian periods was located both in temples and in private residences.

Information concerning the scribes who wrote astronomical and astrological tablets in Babylonia – for simplicity I will call these individuals "astronomical scribes" – can be gathered from two sources: colophons written on astronomical and astrological tablets, and administrative texts

[4] O. Pedersén, *Archives and Libraries in the Ancient Near East 1500–300* BC (Bethesda, MA: CDL Press, 1998), p. 163.

[5] D. J. Wiseman and J. A. Black, *Literary Texts from the Temple of Nabû*, Cuneiform Texts from Nimrud IV (London: British School of Archaeology in Iraq, 1996).

[6] O. R. Gurney and P. Hulin, *The Sultantepe Tablets II* (London: British Institute of Archaeology at Ankara, 1964).

[7] P. Clancier, *Les bibliothèques en Babylonie dans la deuxième moitié du 1er millenaire av. J.-C.* (Münster: Ugarit-Verlag, 2009); Robson, *Mathematics in Ancient Iraq*, pp. 227–60.

from the temples in which the astronomical scribes were employed.[8] From the former source it has been possible to establish that most astronomical scribes identified themselves as either a *kalû* "lamentation priest" or a *mašmaššu* "incantation priest." Sometimes the scholars supplement this title with a second title: *ṭupšar Enūma Anu Enlil* "scribe of the (celestial omen series) *Enūma Anu Enlil*." Most astronomical scribes were members of scribal families where their fathers, brothers, and uncles were often also scribes. The scribes had broad scholarly interests, often owning or writing/copying tablets containing omens, literary, ritual, and medical texts, as well as astronomical texts. The astronomical scribes sometimes appear as either a witness in a business text or as the scribe of one of these texts, suggesting that they held a fairly highly respected position in society.

MATHEMATICAL BACKGROUND

The sexagesimal place-value number system is used extensively in cuneiform astronomical texts from all periods. Throughout this chapter I transcribe sexagesimal numbers by separating each sexagesimal "digit" using a comma. Where the absolute value of a number can be inferred from the context I use a semicolon to separate integers and fractions. Two sets of units of measurement are used extensively in the astronomical texts.[9] UŠ and *bēru* (1 *bēru* = 30 UŠ) are used in the measurement of time and for the measurement of certain distances in the heavens. In both cases, 360 UŠ corresponds to a complete rotation and for this reason UŠ are conveniently translated as "degrees." For the measurement of time, there are 360 UŠ in one day. In celestial measurement, UŠ are mainly used in two contexts: for the measurement of position in the zodiac (i.e. celestial longitude and, occasionally, latitude), and for intervals between the culmination of stars (i.e. right ascension). The measurement of position of a heavenly body relative to a star is usually stated in KÙŠ "cubits" and SI "fingers"; in the late period there were 24 SI in a KÙŠ and in principle 1 KÙŠ = 2 UŠ,[10] although in practice when measured the KÙŠ seems to have been slightly larger.[11]

[8] O. Neugebauer, *Astronomical Cuneiform Texts* (London: Lund Humphries, 1955) (henceforth, "Neugebauer, *ACT*"), pp. 13–25; H. Hunger, *Babylonische und assyrische Kolophone* (Neukirchen-Vluyn: Verlad Butzon & Bercker Kevelaer, 1968); Rochberg, "Scribes and Scholars."

[9] See further D. Brown, "The Cuneiform Conception of Celestial Space and Time," *Cambridge Archaeological Journal* 10 (2000), 103–22 and J. M. Steele, "Celestial Measurement in Babylonian Astronomy," *Annals of Science* 64 (2007), 293–325.

[10] J. M. Steele, "Planetary Latitudes in Babylonian Mathematical Astronomy," *Journal for the History of Astronomy* 34 (2003), 269–89.

[11] A. Jones, "A Study of Babylonian Observations of Planets Near Normal Stars," *Archive for History of Exact Sciences* 58 (2004), 475–536.

Three mathematical tools were commonly used in Babylonian astronomy: period relations; zigzag functions; and step functions.[12] A period relation is an equivalence of two numbers of phenomena: most commonly, the period relation equates a number of occurrences of an astronomical event and a number of months or years. In general terms, Z events of one kind are assumed to correspond exactly to Π events of another kind, and the period P can be defined as Π/Z.

Zigzag functions were used to represent variable functions where the variability of one phenomenon could be expressed as a simple function of another. The phenomenon represented by a zigzag function increases and decreases linearly from between a minimum (m) and a maximum (M) by a fixed amount (d) for each unit of the other function. In its simplest form, the zigzag function will repeat after one cycle, reaching both the minimum and the maximum in a whole number of steps. In other words, the difference $\Delta = M - m$ between maximum and minimum is a whole-number multiple of d in these simple zigzag functions. The most common examples of this simple form of a zigzag function relate a phenomenon (for example, the length of daylight) to either the month of the year or the day of a month. More complex zigzag functions are found in the mathematical astronomy of the Late Babylonian period. In these cases, the difference $M - m$ is not equal to a whole number multiple of d, and the values of the zigzag function do not repeat after one cycle. In both simple and complex zigzag functions the period P can be expressed as

$$P = \frac{2(M - m)}{d} = \frac{2\Delta}{d} = \frac{\Pi}{Z}.$$

The period P is an integer in the case of simple zigzag functions but a fraction for complex zigzag functions.

Step functions provided a second way of representing the variation in a phenomenon. Step functions are so-named because the variation in a phenomenon is represented by two or more zones of length $\alpha_1, \alpha_2, \ldots$ of constant amplitude $w_1, w_2, \ldots$ which vary discontinuously at the boundaries between the zones. In most cases, step functions are used to represent phenomena which are functionally dependent upon celestial longitude (λ), most commonly the synodic arc ($\Delta\lambda$) or the synodic time (Δt). Where the addition of w to a starting value would take the resulting value past a zone boundary, the resulting distance past the zone boundary must be modified by the factor w_{i+1} / w_i. The period of an n-zone step function can be expressed as

$$P = \frac{\Pi}{Z} = \sum_{1}^{n} \frac{\alpha_i}{w_i}.$$

[12] O. Neugebauer, *A History of Ancient Mathematical Astronomy* (New York: Springer, 1975) (henceforth, "Neugebauer, *HAMA*"), pp. 373–9.

THE CALENDAR

Two types of calendar were used in Mesopotamia during the second and first millennia BCE: a civil or cultic calendar and a schematic calendar.[13] The civil calendar operated as a luni-solar calendar: on the evening of the thirtieth day of the month (the day began at sunset) a watch was kept for the new moon crescent. If the moon was seen, the thirtieth day was renamed as the first day of the new month; if the moon was not seen then the month began on the next evening. A month could therefore contain either 29 or 30 days.[14]

Twelve lunar months contain on average about 355 days, a little more than ten days short of the length of the solar year. Thus, in order to keep the calendar in line with the seasons, an intercalary (DIRI) month was added roughly every three years.[15] The decision over whether or not to intercalate in a given year was at most periods officially made by the king, in part because of its obvious political importance (for example, in determining the date of the collection of taxes and tributes).[16] Several early sources, including MUL. APIN (see below) and various individual tablets, contain schemes for determining whether a year should contain an intercalary month, often based upon the date of conjunction of the moon and the Pleiades.[17] It is not at all clear, however, whether these schemes were ever intended to be used to determine actual intercalations, and, if they were, whether they were ever put into practice.[18] The question of whether or not to intercalate is frequently discussed in the correspondence of scholars with the Neo-Assyrian kings in the seventh century BCE. Some of these scholars were clearly basing their recommendations upon astronomical criteria; nevertheless, the final decision over whether or not to intercalate rested with the

[13] R. A. Parker and W. H. Dubberstein, *Babylonian Chronology 626* BC–AD *75* (Providence, RI: Brown University Press, 1956); R. Englund, "Administrative Timekeeping in Mesopotamia," *Journal of the Economic and Social History of the Orient* 31 (1988), 121–85; J. P. Britton, "Calendars, Intercalations and Year-Lengths in Babylonian Astronomy," in J. M. Steele (ed.), *Calendars and Years: Astronomy and Time in the Ancient Near East* (Oxford: Oxbow Books, 2007), pp. 115–32; J. M. Steele, "Making Sense of Time: Observational and Theoretical Calendars," in K. Radner and E. Robson (eds.), *The Oxford Handbook of Cuneiform Culture* (Oxford: Oxford University Press, 2011), pp. 470–85.

[14] P.-A. Beaulieu, "The Impact of Month-lengths on the Neo-Babylonian Cultic Calendar," *Zeitschrift für Assyriologie* 83 (1993), 66–87; Britton, "Calendars, Intercalations and Year-Lengths." The procedure as just described may represent an idealization of actual practice. In the Neo-Assyrian period, for example, we find considerable debate over when the month began; see, for example, S. Stern, "The Babylonian Month and the New Moon: Sighting and Prediction," *Journal for the History of Astronomy* 39 (2008), 19–42.

[15] There is considerable evidence that intercalation was not performed in the Middle Assyrian calendar. See Y. Bloch, "Middle Assyrian Lunar Calendar and Chronology," in J. Ben-Dov, W. Horowitz, and J. M. Steele (eds.), *Living the Lunar Calendar* (Oxford: Oxbow Books, 2012), pp. 19–61.

[16] Steele, "Making Sense of Time"; J. M. Steele, "Living with a Lunar Calendar in Mesopotamia and China," in Ben-Dov, Horowitz, and Steele (eds.), *Living the Lunar Calendar*, pp. 373–87.

[17] H. Hunger and E. Reiner, "A Scheme for Intercalary Months from Babylonia," *Wiener Zeitschrift für die Kunde des Morgenlandes* 67 (1975), pp. 21–8.

[18] D. Brown, *Mesopotamian Planetary Astronomy-Astrology* (Groningen: Styx, 2000), pp. 118–22.

king. Early in the Persian period (ca. 480 BCE), a fixed 19-year intercalation cycle was adopted.[19] Under the preceding Neo-Babylonian Empire, the king continued to have an official role in proclaiming the intercalary months, which had been indicated by the intercalation cycle, but the limited evidence we have from the early Persian period suggests that the responsibility for proclaiming intercalations had by then passed to the temple.[20]

The second calendar found in Mesopotamia is a schematic calendar where the year contains twelve 30-day months. This calendar appears in at least two contexts: (i) in literary works such as the Babylonian creation epic *Enūma Eliš* and a number of hymns, where the 360-day year appears as the ideal state of the universe as it was created;[21] and (ii) in certain astronomical and economic documents, where it was probably used to simplify calculation.[22]

THE SECOND AND EARLY FIRST MILLENNIUM BCE

Only a very small number of cuneiform tablets containing texts relating to the astral sciences are preserved from the second and early first millennium BCE. These include a handful of second millennium tablets containing celestial omens from Babylon and Nippur plus several more from sites on the periphery of Mesopotamia such as the ancient Hittite capital of Hattuša,[23] several star lists including a Middle Assyrian copy of the so-called "Astrolabe B," and an Old Babylonian tablet containing a scheme for the variation of the length of day and night for each month of the year. By contrast, a fairly substantial number of works that were composed during the second and early first millennium are known from late copies dating to the Neo-Assyrian and Late Babylonian periods. By and large the evidence we have for the astral sciences from original early sources is not out of keeping with what is known from the later copying tradition, although it must always be remembered that there may have been other astronomical texts of different kinds where the original tablets have not been discovered and which did not become part of the later tradition of copies. Any attempt to characterize early Babylonian astronomy must therefore be treated with some caution. Nevertheless, three texts from this early period – all of which are known in multiple copies from the later period – seem

19 Britton, "Calendars, Intercalations and Year-Lengths."

20 The evidence is found in a small group of letters discussed most recently by K. Kleber, *Tempel und Palast: Die Beziehungen zwischen dem König und dem Eanna-Tempel im spätbabylonischen Uruk* (Münster: Ugarit-Verlag, 2008), pp. 267–8.

21 Brown, *Mesopotamian Planetary Astronomy-Astrology*.

22 L. Brack-Bernsen, "The 360-Day Year in Mesopotamia," in Steele (ed.), Calendars and Years, pp. 83–100. See further my comments in Steele, "Making Sense of Time."

23 U. Koch-Westenholz, *Mesopotamian Astrology: An Introduction to Babylonian and Assyrian Celestial Divination* (Copenhagen: Museum Tusculanum Press, 1995), pp. 36–51.

representative of second and early first-millennium BCE astronomy: the "Three Stars Each," which list three stars and the length of daylight for each month of the year (often misleadingly called "Astrolabes" in modern scholarship because several exemplars are written on circular tablets), the celestial omen series *Enūma Anu Enlil*, and the two-tablet compendium of astronomy and omens MUL.APIN.

The "Three Stars Each" set out a scheme in which three stars (or star groups) are assigned to each month of the ideal year.[24] The stars are divided into three "paths" named after the gods Anu, Enlil, and Ea. These paths extend across the sky in bands of declination: the path of Anu stretches across the center of the sky extending to declinations of about ± 17°, with the path of Enlil to the north and the path of Ea to the south.[25] The underlying principle of the "Three Stars Each" scheme is that each month one star in each of the three paths will rise heliacally. In practice, however, several of the stars included in these lists will not rise heliacally in accordance with the scheme: a few of the stars listed are circumpolar stars which never rise heliacally, and the planets Venus and Jupiter are included in the paths of Anu and Enlil respectively, but because the planets move relative to the fixed stars, they will not have their heliacal rising in the same month each year.

Several of the "Three Stars Each" texts also include a scheme for the length of daylight for each month of the year. The scheme is a simple zigzag function with maximum M = 4,0 UŠ in Month III, minimum m = 2,0 UŠ in Month IX, and monthly difference 20 UŠ. In this day-length scheme, therefore, it is assumed that the spring equinox takes place in the final month of the year. Similar zigzag functions for the length of daylight are found in an Old Babylonian tablet and in tablet 14 of *Enūma Anu Enlil* (see below); the same scheme but shifted by one month so that the spring equinox is placed in Month I is found in MUL.APIN (see below). The ratio of longest to shortest daylight of 4,0 UŠ: 2,0 UŠ or 2:1 is a gross exaggeration of the actual variation between the length of daylight at the winter and summer solstices for any site in either Assyria or Babylonia. Various attempts have been made to explain the inaccuracy of this parameter in terms of peculiarities in the design of waterclocks or by assuming that time was somehow measured as the rising arc of the Sun along the horizon, but ample textual evidence makes it certain that the astronomical scribes did indeed interpret the 2:1 ratio as time measured in equal units.[26] As I will discuss below, the length of daylight was mainly used as a generating function used to calculate the variation of other astronomical phenomena such as the duration of the visibility of the moon, and it seems that the

[24] W. Horowitz, *The Three Stars Each: The Astrolabes and Related Texts* (Horn: Berger & Söhne, 2015).

[25] J. Schaumberger, *Sternkunde und Sterndienst in Babel 3. Ergänzungscheft zum Ersten und Zweitren Buch* (Münster: Aschendorffsche Verlagsbuchhandlung, 1935), pp. 321–2.

[26] D. Brown, J. Fermor, and C. B. F. Walker, "The Water Clock in Mesopotamia," *Archiv für Orientforschung* 47 (1999), 130–48.

inaccuracy of the 2:1 ratio was outweighed by its simplicity and its utility in generating other functions of greater accuracy.

The celestial omen series *Enūma Anu Enlil* is one of several collections of omens that were compiled sometime in the second or early first millennium BCE, and which were extensively copied, commented upon, and sometimes revised in the Neo-Assyrian and Late Babylonian periods. *Enūma Anu Enlil* contains omens drawn from the appearance of astronomical and meteorological phenomena in the sky. Broadly, the series can be divided into omens dealing with the moon, lunar eclipses, the sun, solar eclipses, weather, stars, and the planets. The largest group of omens concern the moon, in particular the appearance of the new moon at the beginning of the month, lunar halos, and eclipses of the moon.[27] The lunar eclipse omens concern the date and time of the eclipse, its duration, appearance (magnitude, color, and direction of motion of the shadow), and various phenomena associated with the eclipse (visibility of stars and planets, wind, etc.).[28] The omens concerning the sun are drawn from the appearance of the sun on the first morning of a new month, the appearance (especially the color) of the sun's disk, and solar eclipses.[29] The stellar and planetary sections of *Enūma Anu Enlil* are much more poorly preserved than the lunar and solar sections. In general, these sections contain omens derived from the appearance of stars and planets (color, brightness, dates of first and last visibility, etc.), the motion of planets through constellations, and the motion of planets relative to one another.[30]

Tablet 14 of *Enūma Anu Enlil* differs in format and content from the other tablets of the series. Instead of containing omens written in the standard protasis–apodosis format, tablet 14 contains a series of four numerical tables.[31] The first two tables contain the duration of visibility of the moon for each night of a thirty-day equinoctial month according to two different schemes. The first table, said to be from the "tradition of Nippur," expresses the duration in UŠ, whereas the second expresses the duration in *mina* where 1 *mina* = 60 UŠ. Both tables model the change in the duration of lunar visibility using a zigzag function with M = 3,0 UŠ and m = 0 UŠ (or their equivalent in *mina*). In the first table, however, the zigzag function is modified such that the lunar visibility doubles every day over the first five

[27] For a survey of the contents of this group, see L. Verderame, "Enūma Anu Enlil Tablets 1–13," in J. M. Steele and A. Imhausen (eds.), *Under One Sky: Astronomy and Mathematics in the Ancient Near East* (Münster: Ugarit-Verlag, 2002), pp. 447–57. For an edition of the first six tablets of this section of *Enūma Anu Enlil*, see L. Verderame, *Le Tavole I–VI della serie astrologica Enūma Anu Enlil* (Rome: Grafica Cristal, 2002).

[28] F. Rochberg-Halton, *Aspects of Babylonian Celestial Divination: The Lunar Eclipse Tablets of Enūma Anu Enlil* (Horn: Berger & Söhne, 1988).

[29] W. H. van Soldt, *Solar Omens of Enuma Anu Enlil: Tablets 23 (24)–29 (30)* (Leiden: Nederlands Historische-Archaeologisch Instituut te Istanbul, 1995).

[30] E. Reiner and D. Pingree, *Babylonian Planetary Omens*, 4 vols. (Malibu, CA: Undena, 1975–81 [vols. 1–2] and Leiden: Brill, 1998–2005 [vols. 3–4]).

[31] F. H. N. al-Rawi and A. R. George, "Enūma Anu Enlil XIV and Other Early Astronomical Tables," *Archiv für Orientforschung* 38–39 (1991–2), 55–73.

days of the month and halves every day over the last five days of the month. The reason for modifying the zigzag function in the first table is not understood; a Late Babylonian commentary text uses various mathematical tricks to relate the numbers in the first two tables, but it is clear that the author of this commentary is attempting to justify the difference between the tables but does not know why the difference exists.[32] The third table on *Enūma Anu Enlil* 14 contains the length of day and night for the fifteenth and thirtieth day of each month in the ideal calendar expressed in *mina*. The variation in the length of daylight is modeled using a zigzag function whose maximum and minimum are equivalent to 4,0 UŠ and 2,0 UŠ. Finally, the fourth table gives the duration of lunar visibility for the first and fifteenth day of each month in the ideal calendar expressed in UŠ. The monthly change in duration of lunar visibility is modeled by a zigzag function and is connected to the monthly change in the length of night by the assumption that the daily change in the duration of the moon's visibility is equal to one-fifteenth of the length of night on the fifteenth of the month.

MUL.APIN is a two tablet series known from many Neo-Assyrian and Late Babylonian copies.[33] It contains several lists of stars, a small set of celestial omens, mathematical schemes concerning intercalation, the variation in the length of daylight, the duration of lunar visibility, and the length of shadow cast by a gnomon at different times of day on the dates of the solstices and equinoxes, and the periods of visibility of the planets. Some of the material contained in MUL.APIN is related to the texts discussed above: the first star list in MUL.APIN, for example, contains lists of stars in each of the three paths of Anu, Enlil, and Ea, which relate to the "Three Stars Each" texts, and the length of daylight and the duration of visibility scheme are related to the tables on *Enūma Anu Enlil* 14, differing only by placing the spring equinox in Month I instead of Month XII. Other parts of MUL.APIN contain material that is not found in either *Enūma Anu Enlil* or the "Three Stars Each," such as the shadow-length scheme and the list of *ziqpu* ("culminating") stars. Nevertheless, these parts of MUL.APIN are clearly part of the same astronomical tradition as the material in these other texts. The shadow-length scheme, for example, is based upon the common zigzag scheme for the length of daylight found in the "Three Stars Each," *Enūma Anu Enlil* 14, and elsewhere in MUL.APIN.[34]

Notwithstanding the caution expressed above about the possible bias in our understanding of early Babylonian and Assyrian astral science caused by the relatively small number of original sources preserved, it appears that in

[32] J. M. Steele and L. Brack-Bernsen, "A Commentary Text to *Enūma Anu Enlil* 14," in M. Ross (ed.), *From the Banks of the Euphrates: Studies in Honor of Alice Louise Slotsky* (Winona Lake, IN: Eisenbrauns, 2008), pp. 257–66.

[33] H. Hunger and J. M. Steele, *The Babylonian Astronomical Compendium MUL.APIN* (Abingdon: Routledge, 2018).

[34] J. M. Steele, "Shadow-length Schemes in Babylonian Astronomy," *SCIAMVS* 14 (2013), 3–39.

the second and early first millennium BCE two threads within astronomical writings can be distinguished: celestial omens and simple astronomical schemes. These schemes usually relate astronomical phenomena such as the duration of lunar visibility or the heliacal rising of a star to the date in the ideal calendar. Where these schemes concern numerical quantities their variation is usually represented by a simple zigzag function with an integer period. Most of these schemes are interrelated through the use of the ideal calendar and in several cases through their dependence on the 2:1 ratio for the length of daylight at summer and winter solstice. For example, the lunar visibility schemes in *Enūma Anu Enlil* 14 and MUL.APIN are both generated by assuming that the daily change in lunar visibility is given by one-fifteenth of the length of night on the fifteenth of the month. Similarly, in the shadow-length scheme in MUL.APIN, the time after sunrise when the shadow reaches one cubit in length is derived from the length of daylight.[35] The interconnection of these mathematical schemes, and especially the role played by the length of daylight as a function from which other functions are generated, is probably responsible for the use of a grossly inaccurate ratio for the length of daylight at the summer and winter solstices.

The purpose of early Babylonian astronomical texts such as MUL.APIN and the "Three Stars Each" remains a matter of debate. For example, David Brown has argued that the various astronomical schemes in these texts were intended to provide a model of the ideal state of the universe against which observations could be judged and agreements and discrepancies from the ideal interpreted as omens.[36] By contrast, Lis Brack-Bernsen has argued that the schemes in MUL.APIN were used to predict astronomical phenomena, and has shown how the different schemes in MUL.APIN could be combined to produce more accurate predictions.[37] On the evidence currently known it is not possible to rule out either interpretation; indeed, seeking a single purpose for these texts is probably the wrong question to ask as they may very well have been used in several different ways by the Babylonian scribes.[38]

THE NEO-ASSYRIAN PERIOD

Several hundred cuneiform tablets dating to the eighth and seventh centuries BCE containing texts of astral science are preserved from sites in Assyria. The majority of these texts fall into one of three groups: copies of earlier works such as *Enūma Anu Enlil*, MUL.APIN, and related works; individual

[35] Steele, "Shadow-length Schemes."

[36] Brown, *Mesopotamian Planetary Astronomy-Astrology*, pp. 153–6.

[37] L. Brack-Bernsen, "The "Days in Excess" from MUL.APIN: On the "First Intercalation" and "Water Clock" Schemes from MUL.APIN," *Centaurus* 47 (2005), 1–29.

[38] Steele, "Making Sense of Time," p. 474.

texts dealing with astronomical topics such as schemes to determine intercalation and lists of stars; and the correspondence between the Neo-Assyrian kings and scholars employed to provide advice on a range of topics including celestial divination, astronomical observation, and the calendar.

Among the star lists known from the Neo-Assyrian period, one list stands out in importance: a list of *ziqpu* ("culminating") stars. This text, known from a tablet from Nineveh and a Late Babylonian duplicate from Uruk, contains a list of 26 stars or star groups which culminate in order. Accompanying the star names are distances between the stars "on the ground" and "on the sky" (the meaning of these phrases is unclear), where the distances "on the sky" are 64,800 times the distances "on the ground," and time intervals given by weights of water in a waterclock.[39] The distances "on the ground" are given in UŠ and correspond to differences in right ascension; for reasons that are not understood, the total circuit of the *ziqpu* stars in this text adds up to 364° (we would expect a total of 360°). The *ziqpu* stars were used for timing events at night; references to their use for marking the time of nighttime events are found in both astronomical and non-astronomical contexts during the Neo-Assyrian period. As the text implies, the differences in UŠ between the *ziqpu* stars (i.e. their differences in right ascension measured in UŠ) correspond to UŠ as a unit of time. The 26-star list known from Nineveh and its Uruk copy may be an atypical list as a 25-star version of the *ziqpu*-star list which omits one star from the 26-star list is widely attested in the Late Babylonian period.[40]

The correspondence between the Neo-Assyrian kings Esarhaddon and Assurbanipal and their scholarly advisors provides a unique insight into the practice of astronomy during this period.[41] The main topic of the correspondence concerns the interpretation of ominous events by the scholars and the advice they give to the king based upon their interpretations. Whilst astronomical events are not the only topic discussed in this correspondence, celestial divination is certainly one of the main subject matters. It is important to remember that these letters were not written as "astronomical texts," but were part of scholarly advice being provided to the kings. The different scholars who wrote to the king displayed differing levels of astronomical knowledge. What was important was the advice they gave the king (whether or not he chose to follow it). In some cases, displays of astronomical knowledge may have served to add authority to the recommendation the scholar was making. The scholar Bel-ušezib, for example, clearly used his letters about celestial omens as a forum to make political points to the king, even recommending specific courses of military action.

[39] Hunger and Pingree, *Astral Sciences*, pp. 84–8.

[40] J. M. Steele, "Late Babylonian *Ziqpu*-Star Lists: Written or Remembered Traditions of Knowledge?," in D. Bawanypeck and A. Imhausen (eds.), *Traditions of Written Knowledge in Ancient Egypt and Mesopotamia* (Münster: Ugarit-Verlag, 2014), pp. 123–51.

[41] H. Hunger, *Astrological Reports to Assyrian Kings* (Helsinki: Helsinki University Press, 1992); S. Parpola, *Letters from Assyrian and Babylonian Scholars* (Helsinki: Helsinki University Press, 1993).

The content of the letters and reports shows that the scholars were well versed in the traditions of celestial omens, both from the series *Enūma Anu Enlil* and from other sources, both textual and oral. The most frequent celestial omens discussed in the correspondence have to do with the appearance of the moon at the beginning of the month and associated calendrical omens; other omens discussed include eclipses, planets entering constellations, meteors and comets, and what we would consider weather phenomena. Discussion of the omens in the correspondence implies that they were observed, and in many cases, though by no means all, accounts of the observation are reported as well as the ominous interpretation.[42] Some of these reports show that some of the scholars were interested in making precise and accurate astronomical observations. The correspondence also shows that some of these scholars were also attempting to make predictions of upcoming astronomical events in preparation for their observation.[43] For example, several letters containing attempts to predict eclipses, the length of the month, or the visibilities of the planets are preserved. It is not certain what techniques were used to make these predictions, but it does not seem unlikely that they involved the use of simple period relations akin to the goal-year type astronomy of the Late Babylonian period.

The relationship between Neo-Assyrian astronomy and Late Babylonian astronomy is not yet fully understood. The Neo-Assyrian period overlaps with the early part of the Late Babylonian period. Furthermore, some of the scholars writing to the Neo-Assyrian kings were based in Babylonian cities such as Babylon. Further research into the question of whether these scholars writing to the Assyrian king were the same scholars responsible for the astronomical texts (which are of a quite different kind to the material found in the Neo-Assyrian correspondence) written in Babylon at the same time is still required.

THE LATE BABYLONIAN PERIOD

The Late Babylonian period provides the most abundant and diverse collection of cuneiform texts dealing with the astral sciences.[44] More than 4,000 tablets containing texts of astronomy and astrology have been recovered from sites in Babylonia. The contents of the Late Babylonian texts include reports of astronomical observations, collections of predicted astronomical phenomena, procedures for calculating lunar and planetary phenomena

[42] Hunger and Pingree, *Astral Sciences*, pp. 116–38.
[43] Brown, *Mesopotamian Planetary Astronomy-Astrology*, pp. 197–204.
[44] For a classification of the main groups of Late Babylonian astronomical texts, see A. Sachs, "A Classification of the Babylonian Astronomical Tablets of the Seleucid Period," *Journal of Cuneiform Studies* 2 (1948), 271–90 and H. Hunger, "Non-Mathematical Astronomical Texts and their Relationships," in N. M. Swerdlow (ed.), *Ancient Astronomy and Celestial Divination* (Cambridge, MA: MIT Press, 1999), pp. 77–96.

using mathematical schemes and the results of such calculations, star catalogs, horoscopes containing astronomical data relating to the date of birth of an individual which can be used to predict the person's life, astrological treatises, texts of astral medicine, and copies of earlier works such as MUL. APIN and *Enūma Anu Enlil* as well as commentaries and new compositions which further develop the material in these early texts.

Among the Late Babylonian astronomical tablets the largest group are the so-called "Astronomical Diaries," which contain reports of night-by-night observations.[45] The earliest published Diary dates to 652 BCE and only a few scattered examples are known from the sixth and fifth centuries BCE. Many more Diaries are preserved from the fourth century to the early part of the first century BCE. The lack of early examples is almost certainly due to the accidents of preservation and recovery, and it is likely that a more or less complete series of Diaries existed throughout the last six or seven centuries BCE. It is not known when the Diaries first started being written. It is possible that the Diary tradition began during the reign of Nabonassar in the middle of the eighth century BCE,[46] but a later date during the late seventh or early sixth century seems more likely.

A typical Astronomical Diary covers half a year and is divided into sections for each month. Each monthly section begins with a statement of the length (29 or 30 days) of the preceding month followed by a statement of the time interval between sunset and moonset on the evening that begins the month. This time interval, denoted NA, is one of six time intervals defined by the rising and setting of the sun and moon (the so-called "Lunar Six") measured each month. Four of the other intervals are measured around the full moon in the middle of the month (ŠÚ, NA, ME, and GE_6) and the final interval (KUR) is measured on the morning of last visibility of the moon towards the end of the month. These time intervals are measured in UŠ and NINDA (where 1 UŠ = 60 NINDA). The lunar six could be calculated in advance (see below), and when bad weather prevented observation of the sun or moon the calculated value for the lunar six was recorded in the Diary in place of a measured value. Other lunar phenomena, in particular the moon's passages by certain reference stars and lunar and solar eclipses, are systematically reported in the Diaries. Observed eclipses are reported in considerable detail; typically an eclipse record includes a report of the time of the eclipse relative to sunrise or sunset, the duration of the phases of the eclipse, the magnitude of the eclipse, the direction of the path of the shadow across the eclipsed body, the position of the moon, the visibility of planets, and the

[45] A. J. Sachs and H. Hunger, *Astronomical Diaries and Related Texts from Babylonia. Volumes I–III* (Vienna: Österreichishe Akademie der Wissenschaften, 1988–96).
[46] A. Sachs, "Babylonian Observational Astronomy," *Philosophical Transactions of the Royal Society of London, Series A* 276 (1974), 43–50.

direction of terrestrial winds.[47] Eclipses could also be predicted in advance (see below), and brief reports of predictions of lunar eclipses that were calculated to take place during the day and solar eclipses which were predicted to take place at night, as well as predictions of eclipses which were not seen in Babylon because they were lunar eclipses where the moon entered only the penumbral shadow or solar eclipses whose path passed entirely to the north or the south of Babylon, are also found in the Diaries. The Diaries also contain records of observations of planetary phenomena. Two types of planetary observations are recorded: observations of the planets passing certain stars, and observations of the synodic phenomena of the planets (first and last visibilities, stations, and acronychal risings). Occasionally, reports of other astronomical phenomena such as comets and meteors are also recorded in the Diaries. The dates of solstices, equinoxes, and the first visibility, acronychal rising, and last visibility of Sirius are also recorded in the Diaries. At least during the Seleucid period, this data was calculated according to a simple 19-year cycle and not observed.[48]

At the end of every month, a summary of planetary data is given listing any synodic phenomena that took place during that month and (in later Diaries) reporting the zodiacal signs within which the planets were located during the month and the dates on which planets crossed from one sign to the next sign. The dates of entries of the planets into zodiacal signs were derived from the observations of the passages of the planets by the Normal Stars.[49] Also given in the end-of-month summaries are reports of the height of the river Euphrates at Babylon, the value of six commodities (barley, dates, cress, mustard, sesame, and wool) in the Babylonian markets, and brief accounts of important events in the life of the city.

Two systems for recording the position of the moon or the planets in the sky are found in the Astronomical Diaries.[50] In the first system, the moon or the planet is said to be a number of cubits and fingers above or below and (usually for lunar positions only) in front of or behind a star. These distances correspond roughly to differences in celestial latitude and longitude.[51] From at least 400 BCE onwards, a group of 28 stars distributed irregularly around the zodiacal belt (known as "Normal Stars") served as reference points for tracking the motion of the moon and planets, supplemented very occasionally by a small number of other stars.[52] Two catalogs of the Normal Stars are preserved in cuneiform sources which give the positions of the stars in the

[47] P. J. Huber and S. De Meis, *Babylonian Eclipse Observations from 750 BC to 1 BC* (Milan: IsIAO-Mimesis, 2004).

[48] Neugebauer, *HAMA*, pp. 357–65.

[49] P. J. Huber, "Ueber den Nullpunkt der babylonischen Ekliptik," *Centaurus* 5 (1958), 192–208; Jones, "A Study of Babylonian Observations of Planets."

[50] Steele, "Celestial Measurement."

[51] Jones, "A Study of Babylonian Observations of Planets."

[52] Jones, "A Study of Babylonian Observations of Planets."

signs of the zodiac.[53] The zodiac was the second system within which celestial positions were given in the Astronomical Diaries. The zodiac was developed towards the end of the fifth century BCE,[54] almost certainly by comparison with the ideal year of twelve 30-day months making a year of 360 days. The zodiac divides the paths of the sun, moon, and planets into twelve equal parts each of which contain 30 degrees. The twelve signs of the zodiac were named after constellations in the corresponding part of the sky. After the zodiac's invention the position of a planet at the time of one of its synodic phenomena was usually given with respect to the sign of the zodiac; the position of the moon during an eclipse was also usually specified by the sign of the zodiac.

The astronomical data recorded in the Diaries provided the source material for another group of texts which contain compilations of lunar and planetary phenomena.[55] These compilation texts usually cover several years and contain collections of a specific type of astronomical data. The most frequently attested type of compilation text contains accounts of lunar eclipse observations and predictions (by contrast only one text is known containing solar eclipse data). These tablets may be arranged either as straightforward lists or in a table format where the data in each column is separated from the previous column by a period of 223 months which corresponds to either 18 years or 18 years plus one month in the Babylonian calendar.[56] Irrespective of their arrangement, the eclipse texts generally contain entries for all successive eclipse possibilities (i.e. syzygies at which an eclipse is either observed or predicted) separated by six or occasionally five months from the preceding eclipse possibility. Particularly notable among the eclipse texts are three fragments from a large compilation of lunar eclipse records arranged in 18-year cycles which when complete covered more than four centuries from 747 BCE to 315 BCE.

The arrangement of several eclipse texts in 18-year cycles reflects the Babylonian discovery of the Saros eclipse period. One Saros of 223 synodic months contains an almost whole number of draconitic and anomalistic months (223 synodic months ≈ 242 draconitic months ≈ 239 anomalistic months). The near whole number of draconitic and anomalistic months in the Saros implies that after 223 synodic months there is an almost exact return in lunar latitude and in lunar velocity. The near return in lunar

53 N. Roughton, J. M. Steele, and C. B. F. Walker, "A Late Babylonian Normal and *Ziqpu* Star Text," *Archive for History of Exact Sciences* 58 (2004), 537–72. The star catalogs and other evidence demonstrate that the Babylonian zodiac was sidereal; see Huber, "Ueber den Nullpunkt."

54 J. P. Britton, "Studies in Babylonian Lunar Theory: Part III. The Introduction of the Uniform Zodiac," *Archive for History of Exact Sciences* 64 (2010), 617–63.

55 H. Hunger, *Astronomical Diaries and Related Texts from Babylonia. Volume V: Lunar and Planetary Texts* (Vienna: Österreichishe Akademie der Wissenschaften, 2001).

56 C. B. F. Walker, "Achaemenid Chronology and the Babylonian Sources," in J. Curtis (ed.), *Mesopotamia and Iran in the Persian Period: Conquest and Imperialism 539–331* BC (London: British Museum, 1997), pp. 17–25.

latitude means that if an eclipse took place at a given syzygy then at the syzygy 223 synodic months later the moon will again be in position to be eclipsed. Furthermore, the magnitude of a lunar eclipse will be similar for eclipses separated by one Saros. However, because 223 mean synodic months is equal to approximately 6585 1/3 days, the time of a lunar eclipse will change by approximately 1/3 of a day after a Saros. Thus, if a lunar eclipse is observed in a given month, knowledge of the Saros can be used to predict that another eclipse will take place 223 months later, but roughly eight hours later in the day (which will often move the eclipse into the hours of daylight making it unobservable at the same geographical longitude). The repetition of solar eclipses after a Saros is less exact because of parallax, but the Saros can be used to identify months of eclipse possibilities, even if it cannot be used to determine whether a solar eclipse will be seen at a given location.

The Babylonian astronomers combined knowledge of the Saros period with the fact that eclipse possibilities generally occur after six months but occasionally take place after a five-month interval to develop a method for predicting all lunar and solar eclipse possibilities for many centuries.[57] This method was almost certainly developed first for lunar eclipses and subsequently applied by analogy to solar eclipses. Because the Saros period of 223 months is an eclipse cycle there must be a whole number of lunar eclipse possibilities within those 223 months. These eclipse possibilities must all be separated from one another by either six or five months, with six months being far more common. Simple mathematics shows that if there are a eclipses separated by six-month intervals and b eclipses separated at five-month intervals, $6a + 5b = 223$, and so a must be 33 and b must be 5. In other words, within a Saros of 223 months there are 38 lunar eclipse possibilities of which 33 are separated from the previous eclipse possibility by six months and 5 are separated by five months. It is natural (and astronomically sensible) to distribute the 5 five-month intervals as evenly as possible within the 38 eclipse possibilities. As a result, within a Saros the eclipse possibilities fall into five groups containing, respectively, eight, seven, eight, seven, and eight eclipse possibilities. Within each group, the first eclipse is placed at a five-month interval and the other eclipses are at six-month intervals. Because the Saros period is an eclipse cycle, the eclipse possibilities within one Saros, and hence the five groups of eclipse possibilities, repeat for the next Saros and so on, producing a scheme for predicting eclipse possibilities more or less indefinitely into the future (in practice the scheme will break down after a few centuries because the Saros period is not exactly a whole number of draconitic months). In order to use this scheme, it simply needs to be aligned with the record of observed eclipses in the first Saros cycle (or ideally the first two or three cycles), after which it can be used to predict all future eclipses.

[57] J. M. Steele, "Eclipse Prediction in Mesopotamia," *Archive for History of Exact Science* 54 (2000), 421–54.

The eclipse observations and predictions recorded in the various eclipse compilations, as well as in the Astronomical Diaries and other texts, all agree with a scheme of this kind that stretched from 747 BCE down to the first century BCE, with three small realignments of the beginning of the scheme during the third and second centuries BCE. It is not known exactly when the Saros scheme for predicting eclipses was developed. The Neo-Assyrian scholars of the seventh century BCE knew about the five- and six-month intervals between eclipse possibilities, and it is quite possible that this knowledge already existed earlier. We can be certain, however, that the Saros scheme was in use in Babylon at the latest by the end of the seventh century BCE.

The Saros period was also exploited to predict the lunar six. Around the beginning of the sixth century BCE the Babylonian astronomers formulated a set of simple but very accurate procedures for calculating the lunar six intervals from linear combinations of observations of those intervals made 223 and 229 months earlier.[58] The basic rules are as follows (following Brack-Bernsen, I use $\mathrm{NA_N}$ to indicate NA at the beginning of the month and NA to indicate NA in the middle of the month):

$$\mathrm{NA}_{\mathrm{N}i} = \mathrm{NA}_{\mathrm{N}i-223} - 1/3(\check{\mathrm{S}}\acute{\mathrm{U}} + \mathrm{NA})_{i-229}$$
$$\check{\mathrm{S}}\acute{\mathrm{U}}_i = \check{\mathrm{S}}\acute{\mathrm{U}}_{i-223} + 1/3(\check{\mathrm{S}}\acute{\mathrm{U}} + \mathrm{NA})_{i-223}$$
$$\mathrm{NA}_i = \mathrm{NA}_{i-223} - 1/3(\check{\mathrm{S}}\acute{\mathrm{U}} + \mathrm{NA})_{i-223}$$
$$\mathrm{ME}_i = \mathrm{ME}_{i-223} + 1/3(\mathrm{ME} + \mathrm{GE}_6)_{i-223}$$
$$\mathrm{GE}_{6i} = \mathrm{GE}_{6i-223} - 1/3(\mathrm{ME} + \mathrm{GE}_6)_{i-223}$$
$$\mathrm{KUR}_i = \mathrm{KUR}_{6i-223} + 1/3(\mathrm{ME} + \mathrm{GE}_6)_{i-229}$$

Corrections need to be applied to these rules to take into account varying month lengths and the minimum value of $\mathrm{NA_N}$ and KUR required for lunar visibility.[59] These methods of calculating the lunar six intervals also provided the means by which the length of the month could be predicted, either by consideration of the calculated $\mathrm{NA_N}$ or by consideration of the corrections that had to be applied during its calculation.[60]

The identification of characteristic periods and the arrangement of observations into compilations based upon those periods extended also to the

[58] L. Brack-Bernsen, "Goal-Year Tablets: Lunar Data and Prediction," in Swerdlow (ed.), *Ancient Astronomy and Celestial Divination*, pp. 149–77; L. Brack-Bernsen and H. Hunger, "TU 11: A Collection of Rules for the Prediction of Lunar Phases and of Month Lengths," *SCIAMVS* 3 (2002), 3–90. Evidence for the use of these procedures in the early sixth century is discussed by P. J. Huber and J. P. Britton, "A Lunar Six Text from 591 BC," *Wiener Zeitschrift für die Kunde des Morgenlandes* 97 (2007), 213–17 and P. J. Huber and J. M. Steele, "Babylonian Lunar Six Tablets," *SCIAMVS* 8 (2007), 3–36.

[59] L. Brack-Bernsen, "Prediction of Days and Pattern of the Babylonian Lunar Six," *Archiv für Orientforschung* 52 (2011), 156–78.

[60] L. Brack-Bernsen, "Predictions of Lunar Phenomena in Cuneiform Sources," in Steele and Imhausen (eds.), *Under One Sky*, pp. 5–19.

planets. For example, several collections of Venus records are arranged into 8-year cycles and a Jupiter text is arranged by 12-year cycles. Just as with the lunar Saros cycle, these planetary cycles could be and were used to predict future planetary phenomena. Several sets of planetary periods are attested in cuneiform sources but the most commonly used were the following:[61]

Mercury: 46 years
Venus: 8 years
Mars: 79 years and 47 years
Jupiter: 71 years and 83 years
Saturn: 59 years

These periods are used in a group of texts known today as "Goal-Year Texts."[62] The Goal-Year Texts are divided into individual sections for each planet (Mars and Jupiter each have two sections), eclipses of the sun and moon, and lunar six data. Each section contains data taken from one period before the "goal year" of the text. For example, a Goal-Year Text for year 100 of the Seleucid Era will contain Mercury data from year 56, Venus data from year 92, Mars data from years 21 and 53, Jupiter data from years 29 and 17, and Saturn data from year 41. Lunar data is taken from one Saros earlier, year 82 in the present example.

The utility of the Goal-Year Texts lies in the fact that the date in the Babylonian calendar of the synodic and sidereal phenomena of a planet and its position in the zodiac repeat almost exactly after one period. For example, if Saturn's first visibility took place in Leo on the 10th of month II in a given year, 59 years later Saturn will again be in Leo and its first visibility will again take place on about the 10th of month II. Two periods are used for Mars and Jupiter because in these cases the first period is more accurate for the dates of synodic phenomena and the second period is better for dates of sidereal phenomena. In order to make a prediction for a coming year, therefore, all that is necessary is to go back the appropriate number of years according to the characteristic periods of each planet and copy out the observed data from the earlier year. This data then becomes the prediction for the coming year. The Goal-Year Texts collect together all of the historical data that are needed to predict the planetary and lunar phenomena for the target "goal" year. Two corrections may need to be applied when making these predictions. First, because the planetary periods are not exactly whole numbers of lunar months, small positive or negative corrections of a few days may need to be made to the dates of the predicted phenomena. Several texts listing various small corrections are known, and evidence from analysis of the data in

[61] J. M. Steele, "Goal-Year Periods and their Use in Predicting Planetary Phenomena," in Selz and Wagensonner (eds.), *Empirical Dimension*, pp. 101–10.

[62] H. Hunger, *Astronomical Diaries and Related Texts from Babylonia. Volume VI: Goal Year Texts* (Vienna: Österreichishe Akademie der Wissenschaften, 2006).

Goal-Year Text and the resulting predictions demonstrates that appropriate corrections of a few days were usually applied.[63] Secondly, on occasions a one-month correction to the date of the prediction may be required to take into account intercalation. Because intercalation in the late period followed a strict 19-year cycle it is easy to deduce when this one-month correction will be needed, and evidence from comparison of the Goal-Year Texts with the resulting predictions indicates that the correction was applied in every case where it was needed.[64]

The predictions that were made from the Goal-Year Texts were recorded in texts known today as Almanacs and Normal Star Almanacs.[65] These texts contain predicted astronomical data for a coming year. For each month the texts contain the length of the preceding month; the dates of the synodic phenomena of the planets together with the zodiacal sign in which the planet will be located on that date; predictions of lunar and solar eclipses; and the date of equinoxes, solstices, the first visibility, acronychal rising, and last visibility of Sirius. In addition, the Normal Star Almanacs contain predicted lunar six data and the dates of passages of the planets by the Normal Stars with the distance of the planet above or below the star. By contrast, the Almanacs give the dates of the day on which the moon set for the first time after sunrise (designated NA) and the day of last lunar visibility (KUR) in place of the lunar six, and the dates on which the planets pass from one sign of the zodiac into another sign of the zodiac. The dates of NA and KUR in the Almanacs are simply the dates of the corresponding calculated NA and KUR values in the Normal Star Almanacs. The dates on which the planets enter the signs of the zodiac are, as in the Diaries, calculated from the dates of the passages of the planets by certain Normal Stars. Thus, all of the data in the Almanacs can be either copied from the Normal Star Almanacs or derived from the data in the Normal Star Almanacs, which, with the exception of the dates of the solstices, equinoxes, and Sirius phenomena, was itself calculated from the data in the Goal-Year Texts. Just as in the Diaries, the solstice, equinox, and Sirius data were generated by a simple 19-year scheme.

Together, the Astronomical Diaries, lunar and planetary compilations, Goal-Year Texts, Almanacs, and the Normal Star Almanacs form a complete and integrated astronomical system. Observations are recorded in the Astronomical Diaries. Observations of those astronomical phenomena that are capable of being predicted are extracted from the Diaries and copied into compilations and Goal-Year Texts. The data in the Goal-Year Texts are then used to make the predictions in the Almanacs and Normal Star Almanacs.

[63] J. M. K. Gray and J. M. Steele, "Studies on Babylonian Goal-Year Astronomy I: A Comparison Between Planetary Data in Goal-Year Texts, Almanacs and Normal Star Almanacs," *Archive for History of Exact Sciences* 62 (2008), 553–600.

[64] J. M. K. Gray and J. M. Steele, "Studies on Babylonian Goal-Year Astronomy II: The Babylonian Calendar and Goal-Year Methods of Prediction," *Archive for History of Exact Sciences* 63 (2009), 611–33.

[65] H. Hunger, *Astronomical Diaries and Related Texts from Babylonia. Volume VII* (Vienna: Österreichishe Akademie der Wissenschaften, 2014).

Finally, those predictions are used to guide future observations and were often inserted into the Diaries where observation was not possible (either because of bad weather or, in the case of eclipses, because the predicted event was not visible at Babylon). Thus, we have here a closed system of astronomy that, in order to function, required no input of data or methods from outside these groups of astronomical texts. Nevertheless, another major astronomical system existed alongside the astronomy of the Diaries and related texts. This other system, usually referred to by modern scholars as "mathematical astronomy," operated in parallel with the Diary tradition for the last four centuries BCE (the original of mathematical astronomy can be traced back to at least the early fifth century BCE).

Texts concerned with mathematical astronomy may be divided into two types: tabular texts, most of which are the so-called "Ephemerides," and texts which outline how the data in the Ephemerides is to be calculated, usually referred to as "Procedure Texts."[66] Babylonian mathematical astronomy used a set of mathematical tools (zigzag functions and step functions) to calculate various phenomena of the moon and the planets. Whereas in the goal-year astronomy discussed above predictions of future astronomical events were made from observations of the same event in the past, in the mathematical astronomy calculations of astronomical phenomena were made using an arithmetic procedure that did not rely directly upon past observations (apart from some putative observations which set the starting point for the system). Crucial to the development of mathematical astronomy was the concept of the zodiac for its ability to locate celestial bodies in the sky using a uniform system of measurement.

The phenomena calculated in the texts of mathematical astronomy bear similarities in their range with those found in the Normal Star Almanacs. For the planets, the most commonly calculated phenomena are the synodic phenomena.[67] Secondary schemes allowed the daily position of the planet in the zodiac and (at least for some planets) its latitude to be calculated.[68] For the moon, the dates, position, latitude, and velocity of the moon at conjunction with and opposition to the sun are calculated, from which was deduced the date and circumstances of eclipses of the sun and moon, the number of days within each month, and the lunar six time intervals. Unlike the Normal Star Almanacs, passages of the planets by the Normal Stars are not calculated, although they could in principle have been determined from

[66] Neugebauer, *ACT* contains a detailed study of the mathematical astronomical texts and provides editions of all texts known at that time. Several dozen tabular texts have been published (mainly by Aaboe and Steele) subsequently; see most recently J. M. Steele, "Newly Identified Lunar and Planetary Tables from Babylon in the British Museum," *SCIAMVS* 11 (2010), 211–39 which also contains references to earlier publications. The Procedure Texts have recently been studied by M. Ossendrijver, *Babylonian Mathematical Astronomy: Procedure Texts* (New York: Springer, 2012).

[67] Neugebauer, *ACT*, p. 279.

[68] P. J. Huber, "Zur täglischen Bewegung des Jupiter nach babylonischen Texten," *Zeitschrift für Assyriologie* 52 (1957), 265–303; Steele, "Planetary Latitudes."

the calculations of the daily positions of the planets and the star catalogs. A second difference between the mathematical astronomical texts and the Normal Star Almanacs is that all known Ephemerides contain calculated data for only one planet or for the moon; data for different planets or for the moon and a planet are never combined in the Ephemerides. The single exception to this rule is a tablet from Uruk, which contains eclipses and planetary phenomena calculated for eleven years.[69]

Underlying all Babylonian mathematical astronomy are period relations, which equate, for example, a number of occurrences of a phenomenon, a number of years or months, and a number of revolutions around the zodiac. For example, for Jupiter it is assumed that 391 occurrences of a specific synodic phenomenon (for example, first visibility) take place in 427 years during which Jupiter has traveled around the zodiac exactly 36 times. In the mean, therefore, first visibilities of Jupiter should occur every 427/391 years = 13 months + 15;14 days and these should be separated in longitude by 36/391 revolutions of the zodiac = 33;8,44,52, . . . °. However, because the sun, the moon, and the planets do not move with uniform velocity, the true spacing between phenomena in both time and longitude may be somewhat greater or less than these mean values. This problem was solved by the use of step functions or zigzag functions whose parameters were chosen such that the period of the function was identical with the underlying period relation.

Several different systems of calculation are known for each of the five planets. These can be grouped into two categories: those that use step functions to calculate the synodic arc ($\Delta\lambda$), which are usually referred to as "System A" type schemes, and those that use zigzag functions to calculate the synodic arc, usually referred to as "System B" type schemes. System A type schemes are known for all of the planets; System B schemes were apparently only used for the outer planets. Generally, in both System A and System B schemes the synodic time (Δt) is determined from the synodic arc by the addition of a constant (c) that is characteristic for each planet ($\Delta t = \Delta\lambda + c$).

For the outer planets, the System A schemes calculate all of the phenomena using the same step function, with the exception of the retrograde phases of Mars. The position and the date of these two phenomena of Mars were calculated as satellite phases of Mars' first station using simple schemes which depend upon Mars' longitude at first station (five schemes are currently known).[70] For the inner planets, different step functions are used to calculate each of the synodic phenomena. The most well-attested schemes for Mercury

[69] J. M. Steele, "A 3405: An Unusual Astronomical Text from Uruk," *Archive for History of Exact Sciences* 55 (2000), 104–35.

[70] Neugebauer, *HAMA*, pp. 458–60; A. Aaboe, "A Late-Babylonian Procedure Text for Mars, and Some Remarks on Retrograde Arcs," in D. A. King and G. Saliba (eds.), *From Deferent to Equant: A Volume of Studies in the History of Science in the Ancient and Medieval Near East in Honor of E. S. Kennedy* (New York: New York Academy of Sciences, 1987), pp. 1–14; J. M. Steele, "A New Scheme from Uruk for the Retrograde Arc of Mars," *Journal of Cuneiform Studies* 57 (2005), 129–33.

are unusual in that one scheme uses step functions to calculate the morning and evening first visibilities and then calculates the last visibilities as satellites of the first visibilities, whereas the other scheme calculates the evening and morning last visibilities using step functions and treats the first visibilities as satellite phases.[71] It remains unclear why the two schemes were never used together to calculate all four first and last visibilities using step functions. Because Mercury always remains close to the sun, at certain times of year Mercury will never reach a sufficiently high altitude to experience a period of morning or evening visibility. As a result, Mercury will sometimes skip an appearance. This phenomenon was known to the Babylonians who noted in both the Diaries and related texts certain first and last visibilities that would be omitted. The same phenomenon is accounted for in the Ephemerides with the first and last visibilities marked as omitted in the same way.[72]

System B schemes are attested for Mars, Jupiter, and Saturn. These schemes calculate the synodic arc and the synodic time using zigzag functions. One again, the relationship $\Delta t = \Delta\lambda + c$ holds, although in practice separate zigzag functions were used to calculate the synodic arc and the synodic time; the zigzag function for synodic time is offset from that of the synodic arc by c (modulo 12 or 24 months as necessary). The zigzag functions employed in the mathematical astronomical texts differ from the simple zigzag functions found in MUL.APIN and *Enūma Anu Enlil* tablet 14 in that they have non-integer periods. As a result, the value of the zigzag function will not reach the maximum or minimum each time through the zigzag. This represents a major conceptual innovation in Babylonian astronomy.

Babylonian lunar theory is considerably more complicated than the planetary theory. Again, there are two main systems of calculation: System A in which the longitude of the moon at syzygy is calculated by a step function, and System B in which it is calculated using a zigzag function. The basic structure and aims are the same for both systems: a series of functions model various quantities calculated for every syzygy. Some of these functions correspond directly to the output of the theory (for example the longitude of the moon at syzygy), while other functions are combined to calculate other functions that are the output of the theory (for example, functions which represent the contribution of lunar anomaly to the varying length of the synodic month, the contribution of solar anomaly to the varying length of the synodic month, and the monthly change in the time of sunset are combined to calculate the time of syzygy). Separating complicated astronomical variables such as the length of the synodic month into contributions relating to different astronomical causes is probably the most significant innovation in the history of mathematical astronomy. It is worth noting that many of the individual contributions cannot be observed directly,[73] and so

[71] Neugebauer, *HAMA*, pp. 401–3.
[72] Neugebauer, *HAMA*, pp. 403–4.
[73] L. Brack-Bernsen, *Zur Entstehung der babylonischen Mondtheorie: Beobachtung und theoretische Berechnung von Mondphasen* (Stuttgart: Franz Steiner Verlag, 1997), pp. 47–54.

must have been derived from observation indirectly, which implies a deep theoretical understanding of the phenomena.[74]

Although the System A and System B lunar theories have the same structure and aims, in detail they differ at almost every level. In System A, the longitude of the moon at syzygy is calculated using a simple two-zone step function. The length of daylight and the contribution of solar anomaly to the variable length of the synodic month are calculated directly from the moon's longitude. The moon's latitude is calculated from its longitude and an assumption that the moon's node moves with uniform motion backwards through the zodiac. From the latitude, the magnitude of potential lunar and solar eclipses is calculated (the calculation of the magnitude of solar eclipses does not take into account parallax or geographical location). A second set of functions accounts for the effects of lunar anomaly. First, a zigzag function is used to calculate the contribution of lunar anomaly to the variation in the length of the Saros (223 months). From this function two further functions are derived: the instantaneous lunar velocity in degrees per day and the contribution of lunar anomaly to the variation in the length of the synodic month. Next, the solar and lunar contributions to the variable length of the month are combined to calculate the true length of the synodic month. Finally, the length of the synodic month is used to calculate the time of syzygy, and then the duration of visibility of the moon around the new (which defines the length of the calendar month) and full moon. Overall, System A exhibits a tight and rigorous theoretical structure, which enabled the system to run unchanged from at least the last part of the fourth century BCE down to the end of the first century BCE.

System B, in contrast to System A, does not have such a tight theoretical structure: some functions seem to be based upon different parameters to other functions, while others may even be theoretically incompatible.[75] Generally, in System B the various functions are calculated independently, except for the final outputs such as the time of syzygy, where, as in System A, functions representing the contributions of lunar and solar anomaly are combined.

System A and System B coexisted during the last three centuries BCE, alongside the observations and the goal-year type predictions found in the Astronomical Diaries and related texts. There was no transition from observation to goal-year astronomy to mathematical astronomy in Babylonia, nor does there appear to have been a different status for these different astronomies: so far as we can tell all three were undertaken by the same individuals.

[74] Brack-Bernsen, *Zur Entstehung der babylonischen Mondtheorie*; J. P. Britton, "Studies in Babylonian Lunar Theory: Part I. Empirical Elements for Modeling Lunar and Solar Anomalies," *Archive for History of Exact Sciences* 61 (2007), 83–145; J. P. Britton, "Studies in Babylonian Lunar Theory: Part II. Treatments of Lunar Anomaly," *Archive for History of Exact Sciences* 63 (3009), 367–431.

[75] A. Aaboe, "On Columns H and J in Babylonian Lunar Theory of System B," in Steele and Imhausen (eds.), *Under One Sky*, pp. 1–4, but see the comments by J. P. Britton, "On Corrections for Solar Anomaly in Babylonian Lunar Theories," *Centaurus* 45 (2003), 46–58.

The invention of the zodiac, which was fundamental to the development of mathematical astronomy, also enabled the creation of new techniques within astrology. Traditional celestial divination of the kind found in *Enūma Anu Enlil*, where the apodosis generally referred to the king and the land as a whole, became increasingly redundant during the second half of the first millennium BCE as Babylonia was incorporated into the Persian and Greek empires and there was no longer a Babylonian king. Although this traditional form of celestial divination did not disappear, its decline opened space for the development of personal astrology. This new form of astrology took several forms, but the most significant was the practice of predicting the life of an individual based upon the position of the sun, moon, and planets (and sometimes the astronomical phenomena) at or near the time of birth. This practice is usually referred to as horoscopic astrology. A small number of horoscopes are preserved, dating from the late fifth to the first century BCE,[76] as well as several texts which set out systems for predicting the life of a native from astronomical data.[77] Most important among the methods of horoscopic astrology was consideration of the positions of the sun, moon, and planets in the zodiac, taking into account not only their individual positions, but also their relative geometrical arrangements and their presence in "secret places" (corresponding to the Greek hypsoma).[78] The astronomical data used to construct horoscopes seems to have been taken from both the Almanacs and the Ephemerides.[79]

Another development within astrology was the use of the zodiac in astral medicine. For example, a group of texts associate days of the year with positions in the zodiac (ultimately derived from a simple scheme for the moon's mean motion based upon the schematic years of 360 days), which are then linked with various stones, plants, animals, and cultic places related to healing.[80] Other texts include lists of diseases associated with signs of the zodiac.[81]

[76] F. Rochberg, *Babylonian Horoscopes* (Philadelphia, PA: American Philosophical Society, 1998).

[77] See, for example, the texts edited in A. Sachs, "Babylonian Horoscopes," *Journal of Cuneiform Studies* 6 (1952), 271–90, F. Rochberg-Halton, "TCL 6, 13: Mixed Traditions in Late Babylonian Astrology," *Zeitschrift für Assyriologie* 77 (1987), 207–28, and E. Reiner, "Early Zodialogia and Related Matters," in A. R. George and I. L. Finkel (eds),*Wisdom, Gods and Literature: Studies in Assyriology in Honour of W. G. Lambert* (Winona Lake, IN: Eisenbrauns, 2000), pp. 421–7.

[78] F. Rochberg-Halton, "Elements of the Babylonian Contribution to Hellenistic Astrology," *Journal of the American Oriental Society* 108 (1988), 51–62.

[79] F. Rochberg-Halton, "Babylonian Horoscopes and Their Sources," *Orientalia* 58 (1989), 102–23; Steele, "A 3405."

[80] L. Brack-Bernsen and J. M. Steele, "Babylonian Mathematics: Two Mathematical Astronomical-Astrological Texts," in C. Burnett, J. P. Hogendijk, K. Plofker, and M. Yano (eds.), *Studies in the History of the Exact Sciences in Honour of David Pingree* (Leiden: Brill, 2004), pp. 95–125; N. Heeßel, "Stein, Pflanze und Holz: Ein neuer Text zur 'medizinischen Astrologie'," *Orientalia* 74 (2005), 1–22; J. M. Steele, "Astronomy and Culture in Late Babylonian Uruk," in C. L. N. Ruggles (ed.), *Archaeoastronomy and Ethnoastronomy: Building Bridges Between Cultures* (Cambridge: Cambridge University Press, 2011), pp. 331–41.

[81] M. J. Geller, *Melothesia in Babylonia: Medicine, Magic, and Astrology in the Ancient Near East* (Boston, MA: De Gruyter, 2014).

The practice of astronomy during the Late Babylonian period was a multi-faceted endeavor, encompassing the observation of celestial phenomena, the development of techniques to predict future astronomical phenomena based upon past observations, the development of mathematical astronomy, and the application of astronomy to a range of new astrological techniques. We still only partially understand the interconnections and interdependencies of the astral sciences in Babylonia, and of the uses to which astronomy was put. On this latter question, we can point to the application of astronomy in regulating the calendar, in astral medicine, in astrology, and in providing advance warning for the preparation for rituals.[82] But it seems apparent that astronomy was also practiced as a purely intellectual endeavor by some scribes, and it is important not to ignore that fact in seeking to understand Babylonian astronomy as a social activity.

[82] See, for example, Steele, "Astronomy and Culture."

Part II

EGYPT

5

THE CULTURAL CONTEXT OF (MATHEMATICAL) EXPERTS IN ANCIENT EGYPT

Annette Imhausen

INTRODUCTION

The cultural location of mathematics in ancient Egypt was the realm of administration. This can be documented with extant sources from the very beginnings of Egyptian mathematics (i.e. the invention of a number notation) until at least the New Kingdom (ca. 1550–1069 BCE). On a basic unsophisticated level, Egyptian mathematics may be described as the skillful handling of numerical values (and quantities indicated in various metrological systems) encountered in various problems.

The following chapter gives an introduction into the context of the production of mathematical knowledge, including its aims, social settings, and the professional identities of the experts that produced the written evidence we study today to learn about Egyptian mathematics. The structure of this overview is chronological. The culture of ancient Egypt lasted for more than 3,000 years, and while much of its output may at a first glance seem "typically Egyptian" to us in comparison to the output of other cultures, there are significant changes that make it necessary to look at the material of a specific Egyptian period in order to draw more profound conclusions. As with other ancient cultures, the source material is not spread evenly over time and the various disciplines. Sources that inform us about the ancient Egyptian experts are of various kinds: the most detailed information about specific practices comes from handbooks of the individual disciplines (like the Rhind mathematical papyrus for mathematics or the Edwin Smith surgical papyrus for medicine).[1] Further evidence about the practices of individual sciences may come from letters or texts written by the professionals in their work life, like the accounts compiled by mathematical experts. Information about these experts' self-perception can be gained from a corpus of literary texts in which the experts feature prominently.

[1] See chapter 6 on Egyptian medicine and chapter 8 on Egyptian mathematics.

Apart from this evidence, further information is provided by the titles that they used to describe their job and/or their rank within the bureaucracy. And, finally, the tombs, which hold ample evidence of the lives of those buried in them, show depictions and texts of experts describing their work.

A significant part of our information on ancient Egyptian science comes from chance finds. The deserts of Egypt, where tombs and temples were located, provided excellent conditions for the preservation of artifacts, hence practically all of the ancient papyri that we have (not only Egyptian but also Greek) originate from Egypt. However, settlements and towns, the locations where life happened, were located then as now in the proximity of the Nile that provided the necessary water. Therefore, ancient Egyptian settlements are often buried under their modern successors and cannot be excavated. In addition, their characteristic feature, the proximity to water or presence of humidity, renders the survival of any organic artifacts from periods several millennia before our present time unlikely. Therefore, it is not surprising that the available evidence is severely biased towards religious and funerary themes, topics that are likely to be located in tombs and temples of the desert.[2] It is from only a few chance finds that we have detailed information about some aspects of daily life and sciences like mathematics or medicine.

LITERACY, NUMERACY AND THE EVOLUTION OF EGYPTIAN EXPERTS

At the beginning of any development of more complicated or advanced mathematical or scientific techniques stands the invention of script and number systems. For Mesopotamia, the origins of literacy and numeracy have been shown to be closely linked, resulting from accounting needs.[3] Although some of the evidence from ancient Egypt supports a similar claim, the situation there seems to be more complex as far as the uses of writing and numeracy are concerned.[4] The available source material comes, on the one hand, as is often the case in ancient Egypt, from funerary contexts and, on

[2] On the use of tombs for our knowledge about life in ancient Egypt cf. Janet Richards, *Society and Death in Ancient Egypt: Mortuary Landscapes of the Middle Kingdom* (Cambridge: Cambridge University Press, 2005) and Steven Snape, *Ancient Egyptian Tombs: The Culture of Life and Death* (Oxford: John Wiley and Sons, 2011).

[3] Eleanor Robson, "Literacy, Numeracy, and the State in Early Mesopotamia," in K. Lomas, R. D. Whitehouse, and J. B. Wilkins (eds.), *Literacy and the State in the Ancient Mediterranean* (London: Accordia Research Institute, 2007), pp. 37–50.

[4] On the development of writing and its significance for pharaonic culture, cf. John Baines, "The Earliest Egyptian Writing: Development, Context, Purpose," in Stephen D. Houston (ed.), *The First Writing: Script Invention as History and Process* (Cambridge: Cambridge University Press, 2004), pp. 150–89, John Baines, *Visual and Written Culture in Ancient Egypt* (Oxford: Oxford University Press, 2007), and Nicholas Postgate, Tao Wang, and Toby Wilkinson, "The Evidence for Early Writing: Utilitarian or Ceremonial?," *Antiquity* 69 (1995), 459–80.

the other hand, from the context of temples. It is estimated that writing was invented in Egypt around 3000 BCE. This was the time of Egypt's transition from multiple predynastic settlements into a unified state, governed and administered centrally.[5] From the beginning, the use of writing seems to have been restricted to the elite. It was used by the early kings and their executive body to represent royal power and also to administer all kinds of products.[6] It is not surprising, therefore, to find evidence for numbers along with the first evidence for writing. Two cultural centers of predynastic times were the Nagada culture of Upper Egypt (named after the largest known predynastic site, Nagada) and the Maadi culture of Lower Egypt. From the first half of the fourth millennium an increasing social stratification becomes apparent within cemeteries of cultural centers such as Nagada.[7]

The analysis of the cemeteries demonstrates the existence of an elite with control over goods and materials. The objects found in elite burials prove the existence of specialized workshops of potters, stone workers, and metalworkers, which seem to have evolved from the Nagada II period (3500–3200 BCE)[8] on.

Artifacts from this context include the earliest evidence of writing and the use of numbers. Labels inscribed with hieroglyphs were found in royal or elite tombs at Abydos, Nagada, and Saqqara. The earliest written objects that have been found so far come from the tomb Uj at Abydos, a major cultural center from predynastic times on. Cemetery U at Abydos (Nagada IIIa, 3200–3100 BCE) consists of pit tombs (usually designated by numbers) and constructed brick tombs (designated by letters).[9] The tomb Uj consists of several rooms, which contained many objects, among which a significant number show evidence of early writing.

[5] For an overview of this part of Egyptian history, cf. Robert J. Wenke, "The Evolution of Early Egyptian Civilization: Issues and Evidence," *Journal of World Prehistory* 5 (1991), 279–329, Toby Wilkinson, *State Formation in Egypt: Chronology and Society*, BAR International Series, 651 (Oxford: Tempus Reparatum, 1996), and Kathryn A. Bard, "The Emergence of the Egyptian State (c. 3200–2686 BC)," in Ian Shaw (ed.), *The Oxford History of Ancient Egypt* (Oxford: Oxford University Press, 2000), pp. 61–88.

[6] Kathryn A. Bard,"Origins of Egyptian Writing," in Renee Friedman and Barbara Adams (eds.), *The Followers of Horus: Studies Dedicated to Michael Allen Hoffman* (Oxford: Oxbow Books, 1992), p. 297.

[7] Predynastic developments have been traced by the evolution of their products. With the emergence of the importance of a mortuary cult, graves become indicators of status, power, and wealth. Thus the evolving social stratification can be followed by evidence from cemeteries. The same evidence also reflects the expansion of the Nagada culture. By the end of the Predynastic period (Nagada III, 3200–3000 BCE) the ceramics of the Nagada culture had spread as far as the Northern Delta. For a detailed description cf. Kathryn A. Bard, "The Egyptian Predynastic: A Review of the Evidence," *Journal of Field Archaeology* 21/3 (1994), 265–88.

[8] All dates are taken from Shaw (ed.), *Oxford History of Ancient Egypt*.

[9] For cemetery U and tomb Uj, cf. Günter Dreyer, "Nachuntersuchungen im frühzeitlichen Königsfriedhof. 5./6. Vorbericht," *Mitteilungen des Deutschen Archäologischen Instituts, Abteilung Kairo* 49 (1993), 23–62 and Günter Dreyer, *Umm el-Quaab I: Das prädynastische Königsgrab Uj und seine frühen Schriftzeugnisse*, Archäologische Veröffentlichungen, 86 (Mainz: Philipp von Zabern 1998).

The evolution of writing is likely to have originated at least partly from the needs of daily life economic activities in the settlements.[10] Apart from the administrative uses that writing served, it was also recognized as an instrument to represent power, and therefore seems to have been monopolized by the elite for their own representative uses. It is the hieroglyphic writing that mostly serves this purpose. The administrative side of power was assigned to a group of people who defined themselves through their ability to read and write. The realm of reading and writing in ancient Egypt naturally included the ability to manipulate numbers. The ancient Egyptian designation for this group is "scribe," written by a hieroglyph showing a scribe's kit, which is attested from the late First Dynasty.[11]

The Egyptian number system as it was used throughout the pharaonic period was fully developed at the time of king Narmer, as is documented by his ritual macehead, which was found at the temple of the god Horus at Hierakonpolis, the most important predynastic site in the south of Egypt.[12] The object originated from a ceremonial context, and the scenes represent king Narmer receiving a tribute consisting of bulls, goats, and captive prisoners. Although there are discussions about what exactly is depicted on the macehead, it is agreed that the numbers used in the tribute are too large to represent actual numbers of a real tribute: 400,000 bulls, 1,422,000 goats, and 120,000 captives are presented to King Narmer. The symbols and method used to express these rather high numbers are the same as those employed later in hieroglyphic inscriptions. Thus, this macehead gives evidence for a fully developed number system before the First Dynasty. It also demonstrates that the number system was not only used for the obvious administrative purposes, but was also important in the representational spheres of Egyptian culture – the tribute received by King Narmer is thought to be symbolic, and meant to impress with its large numbers of items.

LICENSED BY ROYALTY: EGYPTIAN EXPERTS DURING THE OLD KINGDOM

Despite the existence of written texts from the Old Kingdom (2686–2160 BCE) on, the rate of literacy in ancient Egypt is difficult to assess. It certainly varied from period to period and probably varied even more from region to

[10] For a concise overview of Egyptian writing see Richard B. Parkinson and Stephen Quirke, *Papyrus* (London: British Museum Press, 1995). The story of the modern decipherment has been told in detail in Richard B. Parkinson, *Cracking Codes: The Rosetta Stone and Decipherment* (London: British Museum Press, 1999), pp. 12–45.

[11] Jochem Kahl, *Das System der ägyptischen Hieroglyphenschrift in der 0.–3. Dynastie* (Wiesbaden: Harrassowitz, 1994), p. 833.

[12] Cf. Nicholas B. Millet, "The Narmer Macehead and Related Objects," *Journal of the American Research Center in Egypt* 27 (1990), 53–9 and Nicholas B. Millet, "The Narmer Macehead and Related Objects: Correction," *Journal of the American Research Center in Egypt* 28 (1991), 223–5.

region (estimates vary from <1 percent to 15 percent).[13] A general estimate, difficult to achieve at all because of the patchy evidence, is not likely to represent the actual situation for a specific place like Middle Kingdom Lahun (for which the 15 percent may be more accurate) or New Kingdom Deir el-Medina – the two places from which the majority of extant written evidence on papyrus or ostraca originated. These estimates of literacy in ancient Egypt indicate that the ability to read and write was something that was treasured by a small group, to whom the option was open of becoming part of the royal or temple administration. They wrote the handbooks on mathematics, medicine, and other subjects that are our main source of information for scholarly knowledge in ancient Egypt.

No such handbooks originating from the Old Kingdom have survived. However, a variety of sources, like depictions in tombs, archaeological finds, and two extant archives demonstrate the development of metrological systems, and, with the remains of the Giza pyramids still in place today, it seems unlikely that mathematical techniques to plan a building project were not in place by that time.[14] A diagram on a piece of limestone (Cairo JE 50036) found in Saqqara indicates the planning of the construction of a curved structure of some kind (possibly a saddleback construction),[15] and at Meidum a brick wall shows guidelines for the construction of the sloping walls of a mastaba.[16] At least indirect evidence for the use of mathematics in administration can be drawn from the Abusir papyri, which originate from the mortuary temples of two kings of the Fifth Dynasty at Abusir.[17] These texts indicate the assessment of cattle at regular intervals; with regard to formal layouts, they show the arrangement of numeric and other data in tabular format.

[13] Cf. James P. Allen, "Language, Scripts and Literacy," in Alan B. Lloyd (ed.), *A Companion to Ancient Egypt* (Oxford: Blackwell, 2010), pp. 641–62,; John Baines, "Literacy and Ancient Egyptian Society," *Man*, New Series, 18 (1983), 572–99,; John Baines and Christopher Eyre, "Four Notes on Literacy," *Göttinger Miszellen*, 61 (1983), 81–5,; Leonard H. Lesko, "Literacy," in Donald Redford (ed.), *The Oxford Encyclopedia of Ancient Egypt* ed. Donald Redford (Oxford: Oxford University Press, 2001), vol. 2, 297–9; and the webpages Digital Egypt, s.v. "literacy" (http://www.digitalegypt.ucl.ac.uk/literature/literacy.html, accessed 14 September 2011).

[14] For an overview of mathematics and architecture cf. Corinna Rossi, *Architecture and Mathematics in Ancient Egypt* (Cambridge: Cambridge University Press, 2004). For an overview of capacity measures (not only Old Kingdom) cf. Tanja Pommerening, *Die altägyptischen Hohlmaße*, Studien zur altägyptischen Kultur, Beiheft, 10 (Hamburg: Buske, 2005). On early metrology, cf. also Jim Ritter, "Metrology and the Prehistory of Fractions," in Paul Benoit, Karine Chemla, and Jim Ritter (eds.), *Histoire de fractions, fractions d'histoire* (Basel: Birkhäuser, 1992), pp. 19–35.

[15] Cf. Nigel C. Strudwick, *Texts from the Pyramid Age* (Writings from the Ancient World, 16; Atlanta, GA: Society of Biblical Literature, 2005), p. 153, no. 78. Cf. also Rossi, *Architecture and Mathematics*, p. 115.

[16] Cf. Rossi, *Architecture and Mathematics*, pp. 188–92.

[17] For a translation of the Abusir papyri, cf. Paule Posener-Kriéger, *Les archives du temple funéraire de Néferirkarê-Kakaï (Les papyrus d'Abousir). Traduction et commentaire*, Bibliothèque d'étude, 65 (Cairo: Institut français d'archéologie orientale, 1976). Cf. also the now published Gebelein papyri: Paule Posener-Kriéger, "Les papyrus de Gebelein," *Revue d'Egyptologie* 27 (1975), 211–21 and Paule Posener-Kriéger and Sara Demichelis, *I papiri di Gebelein: Scavi Farina 1935*, Studi del Museo Egizio di Torino, 1 (Turin: Ministero per i Beni e le Attività Culturali, 2004).

How mathematical techniques developed, or what they were exactly at this time, we cannot say. However, some experts of the Old Kingdom have left descriptions of their lives and careers within their tombs that allow an assessment of the cultural environment in which they worked.[18] There is evidence that the sons of experts followed their fathers in their careers; thus, the inscription of Akhethotep, an expert of the Fifth Dynasty, mentions a tribute by his son "for having educated him to the satisfaction of the king".[19] The description of the career of Harkhuf of the Sixth Dynasty begins with an expedition to Yam where he is sent with his father, followed by a second expedition to Yam where he is sent on his own.[20] From these autobiographic texts, we also grasp that the tasks of an expert in the royal service were many and varied.

The option of becoming an expert was obviously limited to those persons who were able to read and write (which included the ability to handle numerical values), which presumably was taught by fathers to their sons. This prerequisite is also expressed in the basic Egyptian designation of these experts, "scribe." Hence, being a member of a scribal family increased an individual's likelihood of becoming a scribe. This very basic prerequisite is also expressed in a form of self-depiction chosen by high officials from the Old Kingdom on, the scribal statue, depicting an official in the process of reading or writing a document.[21] Other pictorial representations in tombs of high officials often include scenes of the tomb owner inspecting the harvest. This would also presuppose the skillful handling of numbers, that is, quantities of goods, measured in a variety of metrological systems. At the same time, lower ranks of scribes are depicted preparing the respective documentation, and, judging by depictions of scribes being beaten, the documentation apparently was not always considered to be adequate.

Such succession in office is also the framework of one of the teachings, often attributed to the Old Kingdom (possibly the latter part of the Sixth Dynasty[22]), although the surviving manuscripts date to the Middle and New Kingdoms only. This *Teaching of Ptahhotep* begins by the statement of the mayor of the city, the vizier Ptahhotep, who asks the permission of the king to train a successor ("a staff of old age"[23]), because he himself has become old: "May this servant be ordered to make a staff of old age, so as to tell him the words of those who heard, the ways of the ancestors, who have listened to

[18] For a selection of texts from the Old Kingdom in English translation, see Strudwick, *Texts from the Pyramid Age*.

[19] Translation from Strudwick, *Texts from the Pyramid Age*, p. 261.

[20] Miriam Lichtheim, *Ancient Egyptian Literature, vol. 1: The Old and Middle Kingdoms* (Berkeley, CA: University of California Press, 2006), pp. 23–7.

[21] On this type of statue see Gerry D. Scott, "The History and Development of the Ancient Egyptian Scribal Statue" (dissertation, Yale University, 1989).

[22] Based on the date of the earliest surviving manuscripts and their linguistic features, others propose this to be a Middle Kingdom composition: cf. William Kelly Simpson, *The Literature of Ancient Egypt: An Anthology of Stories, Instructions, Stelae, Autobiographies, and Poetry* (New Haven, CT: Yale University Press, 2003), p. 129.

[23] Lichtheim, *Ancient Egyptian Literature*, vol. 1, 63.

the gods."[24] The teaching speaks *expressis verbis*, at the beginning as well as at the end[25], of the necessity of the king's approval to appoint a successor. The maxims taught in the *Teaching of Ptahhotep* include aspects of both professional life (e.g. the correct behavior towards superiors and inferiors within the administrative system, advice to accept one's position in life, and a warning against exploiting power or acting greedily) and private life (the advice to take a wife and treat her well, for example).

During the Old Kingdom, the social and cultural setting of the scribes was closely connected to the king and his needs. The king was the representative of mankind towards the gods; it was his duty to ensure that life on earth followed the rules of Maat – the cosmological order of the gods. Only if the king guaranteed the preservation of this order on earth, which also included carrying out rituals regularly, would the world continue to exist, i.e. the sun would continue to rise, harvests would take place, etc. Scribes during the Old Kingdom were to serve their king in this task. Whatever the king ordered was meant to be done in order to ensure the continuation of Maat on earth. In return – as is demonstrated, for example, in the biographical inscription of the scribe Weni of the Sixth Dynasty – the king would reward his officials for their successes.[26] These rewards were not limited to advancements in their careers, but also came in the form of various goods, especially expensive objects for their future graves. The setting of ancient Egyptian scholarly activities can be traced through the autobiographic inscriptions in the tombs of these officials. During the Old Kingdom, the success of scribes was judged according to their rank within the royal or temple administration and (linked to this) according to their relationship with the king.

THE FIRST INTERMEDIATE PERIOD: INCREASING POSSIBILITIES FOR INDIVIDUAL EXPERTS

During the First Intermediate Period (2160–2055 BCE), the central administration of Egypt controlled by a single king broke down. Two centers of power existed and fought for dominance: the Heracleopolites in the north (with their center in Heracleopolis south of the Fayyum) and a Theban family (who may have been relatives of the nomarchs of Elephantine) in the south. The tombs of the nomarchs of this period seem to indicate that the responsibility to care for the provisions of the inhabitants of their provinces fell to them. Thus, Ankhtifi, the nomarch at Hierakonpolis (and a supporter of the Heracleopolites) writes in the autobiography in his tomb:

[24] Ibid., p. 63.
[25] "As you succeed me, sound in your body, the king content with all that was done, may you obtain (many) years of life!" Ibid., p. 76.
[26] For a translation of this autobiography, see ibid., pp. 18–23.

> I am the vanguard of men and the rearguard of men. One who finds the solution where it is lacking. A leader of the land through active conduct. Strong in speech, collected in thought, on the day of joining the three nomes. For I am a champion without peer, who spoke out when the people were silent, on the day of fear when Upper Egypt was silent.[27]

Another official, Merer, from Edfu records in his stela:

> Never did I hand a person over to a potentate, so that my name might be good with all men. I never lied against any person – an abomination to Anubis. And when fear had arisen in another town, this town was praised. I acquired cattle, acquired people, acquired fields, acquired copper. I nourished my brothers and sisters.
>
> I buried the dead and nourished the living, wherever I went in this drought which had occurred. I closed off all their fields and mounds in town and countryside, not letting their water inundate for someone else, as does a worthy citizen so that his family may swim. When it happened that Upper Egyptian barley was given to the town, I transported it many times. I gave a heap of white Upper Egyptian barley and a heap of *ḫmi*-barley, and measured out for every man according to his wish.[28]

During the beginning of the First Intermediate Period, climactic changes seem to have led to famines. Thus, Egyptologists have seen this period as a dark one, associated with social and political instabilities. Some years of the First Intermediate Period were almost certainly difficult. Its beginning, with famines caused by climatic changes and the failure of the former king to maintain control over all of Egypt, must have been a drastic time for the Egyptian population. These problems were then mastered – at least if we believe their autobiographies – by local nomarchs, or even officials on a lower level, as can be seen from the letters of Heqanakhte, a priest whose work led him to be away from his household, which he then managed by writing letters.[29] These letters detail the work that he wants to have done (as well as indicating the respective members of the household to do it), and one letter also indicates the rations that the members of the Heqanakhte household should receive according to Heqanakhte.[30]

However, more recent research has revealed that these times may not have been so hard for all of the First Intermediate Period. This is indicated by

[27] Ibid., p. 86.
[28] Ibid., p. 87.
[29] For a recent edition of these letters cf. James P. Allen, *The Heqanakht Papyri* (New York: Metropolitan Museum, 2002).
[30] For a translation see ibid., pp. 16–17.

records of water levels and the occurrence of new types of burial goods, made for the Egyptian middle class, which first appears in this period. Before, all but the wealthiest officials were simply buried with their former belongings; now, specifically made grave goods were available, indicating that there seems to have been enough demand for objects of this kind. In addition, no massive royal building projects had to be completed in these years, which may have resulted in some relief for the population.

During the First Intermediate Period, while the central administrative framework was lacking, the individual nomarchs presumably continued their administrative roles towards the population of their towns or regions. However, they no longer perceived their duties as a service to the king, but rather saw themselves as installed through the power of the gods, as can be seen, for example, at the beginning of the autobiography of Ankhtifi:

> Horus brought me to the nome of Edfu for life, prosperity, health, to reestablish it, and I did (it).[31]

The individual success of a nomarch, as it is expressed in the autobiographies, was measured through his ability to ensure social and economic stability within his own region and through his conduct towards the weak members of society. In the work of an official, the same mathematical knowledge must have played an important role in order to assess (for example) the available grain rations. Through this ability, previously used in the service of the king, the nomarchs were now able to master the new responsibilities that had fallen to them.

STATE ORGANIZATION OF EDUCATION: MIDDLE KINGDOM EVIDENCE (AND SOME FROM THE SECOND INTERMEDIATE PERIOD)

After the reunification of Egypt under Mentuhotep Nebhepetre (Mentuhotep II), a second period of interior stability and external activities followed, during which the borders of Egypt were expanded to the south into Nubia with fortresses at Buhen, Semna, and elsewhere.

The earliest evidence of mathematical texts originates from the Middle Kingdom (2055–1650 BCE). While it cannot be guaranteed that there were no such texts during earlier periods, it may not be a complete fluke that the extant texts date to this period. From their form and content it has been assumed that they come from an educational context, and again, while this cannot be guaranteed, the assumption seems reasonable. The mathematical texts teach the mathematics that a scribe would need to use in his daily work,

[31] Lichtheim, *Ancient Egyptian Literature*, vol. 1, 85.

e.g. calculating volumes of granaries of various shapes, the amounts of rations to be distributed, and the amount of produce that has to be delivered by certain professionals. A remarkable formal feature is the style of the texts, and specifically the grammar used. Mathematical knowledge was presented in the form of collections of problems and their solutions, indicated as step-by-step instructions (which one might refer to as algorithms).[32]

Similar to the area of mathematics, the most informative sources on Egyptian medicine are the medical texts. They are presented in a style similar to collections of cases, and they also use specific grammatical forms. The grammatical form of the *sḏm.ḫr.f* is used in both mathematics and medicine, and serves to indicate the necessity of the action described in this verb. The mathematical and medical procedure texts provide step-by-step instructions about how to proceed in specific cases. The individual cases are organized, at least in some of the papyri, in an intentional sequence.

It is probably no coincidence that this type of text appears during the Middle Kingdom. The reorganization of Egypt under a single king came with a strict reorganization of the administration that was to serve him; in this context the emergence of this type of text seems plausible.

However, perhaps as a consequence from times in which the king did not serve as a reference frame for the experts, a new self-conception of their position emerged. This can be grasped in the literature of the Middle Kingdom. While scribes during the Old Kingdom estimated their own worth according to their role in the king's administration, and their closeness to the king, Middle Kingdom texts indicate that scribes were aware of the power that came with the execution of their duties. The *Tale of the Eloquent Peasant* describes the abuse of this power by an official and the subsequent complaints by the farmer who was mistreated.[33] The peasant appeals to the superior about the injustice that he has suffered and complains about the corruption among the experts who are meant to act truthfully. In this context, several tasks related to mathematics are mentioned explicitly: measuring of grain; weighing with scales; and filling granaries, which, as is depicted frequently in various tombs, was also supervised by experts. In this context, it is interesting to note that there are usually more scribes present than one would think were actually needed – presumably a precaution

[32] For an introduction to Egyptian mathematical texts and their interpretation, cf. the works of Jim Ritter (Jim Ritter, "Measure for Measure: Mathematics in Egypt and Mesopotamia," in Michel Serres (ed.), *A History of Scientific Thought: Elements of a History of Science* (Oxford and Cambridge, MA: Blackwell, 1995), pp. 44–72; Jim Ritter, "Egyptian Mathematics," in Helaine Selin (ed.), *Mathematics Across Cultures: The History of Non-Western Mathematics* (Dordrecht: Kluwer, 2000), pp. 115–36; and Jim Ritter, "Reading Strasbourg 368: A Thrice Told Tale," in Karine Chemla (ed.), *History of Science, History of Text*, Boston Studies in the Philosophy of Science, 238 (Dordrecht: Springer, 2004), pp. 177–200.

[33] A translation of the story, which is well worth reading, can be found in Richard B. Parkinson, *The Tale of Sinuhe and Other Ancient Egyptian Poems 1940–1640* BC (Oxford: Oxford University Press, 2009), pp. 54–88.

against corruption as described in the petition of the peasant. The king, although being mentioned in the petition, is only a figure in a distant palace; the actual power lies with the officials who can act either truthfully or dishonestly. This power originates from the practice of executing measuring and administrative duties.

Similarly, a subtle change can also be noted within the genre of teachings, such as in the *Loyalist Teaching*. The frame is similar to that of the *Teaching of Ptahhotep*; it is a text written for a successor to hand down guidelines for correct conduct in office. Apart from the ideas of modesty and diligence, respect towards superiors is detailed, especially to the king, who is referred to at length in the opening sections of the teaching:

> The king is sustenance; his speech is Plenty. The man he makes is someone who will always exist. He is the heir of every god, the protector of his creators. They strike his opponents for him. Now, his Majesty is in his palace – he is an Atum of joining necks: his protective might is behind the man who promotes his power.[34]

Thus, the teaching reminds the expert to be loyal towards the king; this duty, which had been self-evident in the *Teaching of Ptahhotep* now has to be solicited and justified. The king is authorized by the gods, and it will have positive effects if an official is loyal towards his king. The teaching also refers explicitly to the professional conduct of the scribe in office:

> The man who fixes the taxes in proportion to the barley is [a just] man in God's eyes. The riches of the unjust man cannot stay; his children cannot benefit from any remainder of his. The man who afflicts is making the end of his own life: there are no children of his close to him.[35]

Other passages of the *Teaching* detail the necessity of individual professions like that of a cattle-herdsman and a field-worker and the consequences for them if the scribe treats them well or badly.

The texts of the Middle Kingdom indicate a subtle change in the self-perception of the scribes. The framework of their professional existence remains the service of the king; however, this framework is detailed explicitly, and reasons for this relationship are given with reference to the gods. Likewise, the oversight of the correct conduct of the scribe in the quote above is also affected by the god. In addition, the consequences of scribal behavior for the subordinates are explicitly registered, including the misconduct of scribes and its consequences.

[34] Ibid., p. 239.
[35] Ibid., p. 241.

During the Second Intermediate Period (1650–1550 BCE), Egypt was ruled by the so-called Hyksos, possibly Canaanites from Palestine. It is during this time that the Rhind mathematical papyrus was copied, as is noted in its title section:

> Behold, this roll was written in year 33, month 4 of the inundation season, [… under the majesty of the king of Upper] and Lower Egypt Aauserra, endowed with life, according to the model of writing of former times made in the time of the king of Upper and Lower Egypt […]. It was the scribe Ahmose who wrote this copy.

The Rhind mathematical papyrus is both in quality and quantity the best document that is extant among all hieratic mathematical texts, and it is noteworthy that it originates from the Second Intermediate Period. Obviously, mathematical experts were around and had the facility to compose mathematical compendia. In terms of the subtle change in the self-perception of the scribes mentioned above, there are some problems within the collection of the Rhind papyrus that reflect this change, e.g. problem 67, referring to the calculation related to the work of a herdsman, which is somewhat more elaborate in its rhetorical style than the standard problems:

> Example of calculating the work-produce of a herdsman. Behold now, this herdsman came to the numbering of cattle with 70 oxen: said this accountant of cattle to this herdsman: How few are the head of oxen which you have brought. Where are your numerous head of oxen? This herdsman said to him: Two-thirds of one-third of the cattle, which you had entrusted to me, I have brought to you. Count for me and you will find me complete.

Similarly, problem 61b (the calculation of two-thirds of a fraction) and problem 79 (the calculation of a number of items in a house) are in content or style variations on the "standard" problems found in mathematical texts. These problems may hint at a development from a mathematical papyrus designed and used solely in the training of scribes for their administrative duties to that of a collection of mathematical knowledge, which is considered an important element of scribal culture. The comparison of the two titles of the mathematical papyri that are extant also confirms this development. The Middle Kingdom Lahun fragment UC32162 has the title "Method of calculating the matters of accounting." The Rhind mathematical papyrus in contrast claims in its title to provide the "Method of calculating for entering into affairs, and for knowing all that exists, [every] mystery, […] every secret."

SCRIBAL CULTURE IN NEW KINGDOM EGYPT

The Seventeenth Dynasty rulers from Thebes under Kamose began an insurgence against the Hyksos, and, under Ahmose, the unification of Egypt was achieved. This marks the beginning of the New Kingdom (1550–1069 BCE). This is the last period of pharaonic Egypt, in which, at least temporarily, Egypt held an influential political position. Numerous artifacts from the New Kingdom, some from the tombs of high-ranking New Kingdom experts, are still extant. They demonstrate the level of artwork during this period, e.g. the tomb of Nebamun, the "Scribe and Grain accountant in the Granary of Divine Offerings of Amun."[36] The paintings of the tomb include several scenes concerning the administration of goods, e.g. agricultural scenes showing the harvesting of grain and the viewing of the produce of the estates, in which animals of the estates are led before Nebamun by officials who also bring the corresponding documentation.[37] The areas of numerical responsibilities of the experts apparently remained unchanged. It is somewhat puzzling, therefore, that practically no mathematical texts are extant from this period. While there is a continuity in the domain of medicine – some of the "best" medical texts that are extant originate from the New Kingdom – there is an almost complete break as far as mathematical texts are concerned.[38]

However, there is still reason to assume that mathematics continued to play a significant role in the life of an Egyptian scribe. Instead of school books of individual subjects, a variety of texts, which were presumably circulated among New Kingdom scribes, is extant. They include compositions describing the superiority of the scribal profession over any other profession, eulogies to scribal teachers, and model letters. This corpus of texts is referred to as the *Late Egyptian Miscellanies*.[39] At least some of these texts include references to mathematical practices or mathematical

[36] For a publication of the tomb paintings, see Richard B. Parkinson, *The Painted Tomb-chapel of Nebamun* (London: British Museum Press, 2008).

[37] For a detailed description of the viewing of the produce of the estates, cf. ibid., pp. 92–109; for the agricultural scenes, cf. ibid., pp. 110–19.

[38] Only two very fragmentary sources exist: both are ostraca (i.e. stone or pottery sherds used for writing) with a few incomplete lines of text.

[39] The first study of these texts was Adolf Erman, *Die ägyptischen Schülerhandschriften*, Abhandlungen der preussischen Akademie der Wissenschaften Jahrgang 1925, Philosophisch-historische Klasse, 2 (Berlin: Verlag der Akademie der Wissenschaften, 1925). The *editio princeps* remain the works by Alan H. Gardiner (hieroglyphic transcription) and Ricardo A. Caminos (translation and commentary): Alan H. Gardiner, *Late-Egyptian miscellanies*, Bibliotheca Aegyptiaca, 7 (Brussels: Édition de la Fondation égyptologique reine Élizabeth, 1937) and Ricardo A. Caminos, *Late-Egyptian Miscellanies* (London: Oxford University Press 1954). Recent English translations of some of the texts can be found in Miriam Lichtheim, *Ancient Egyptian Literature, vol. 2: The New Kingdom* (Berkeley, CA: University of California Press, 2006) and Simpson, *Literature of Ancient Egypt*; a recent German translation is provided by Nikolaus Tacke, *Verspunkte als Gliederungsmittel in ramessidischen Schülerhandschriften*, Studien zur Archäologie und Geschichte Altägyptens, 22 (Heidelberg: Heidelberger Orientverlag, 2001).

education. The following examples taken from this corpus illustrate the reputation of mathematics in the Egyptian scribal culture.

The theme of scribal superiority above all other professions is the topic of the following excerpt, section 4,2–5,7 of Papyrus Lansing, which was titled *All callings are bad except that of the scribe* by its first translator, Ricardo Caminos:[40]

> Look for yourself with your own eye. The professions are set before you. The washerman spends the whole day going up and down, all his [body] is weak <through> whitening the clothes of his neighbors every day and washing their linen. The potter is smeared with earth like a person one of whose folk has died. His hands and his feet are full of clay, he is like one who is in the mire. The sandal-maker mixes tan; his odor is conspicuous; his hands are red with madder like one who is smeared with his (own) blood and looks behind him for the kite, like a wounded man whose live flesh is exposed. The flower binder prepares floral offerings and brightens ring-stands; he spends a night of toil, like one <upon> whose body the sun shines. The merchants fare downstream and upstream, and are as busy as brass, carrying goods <from> one town to another and supplying him that has not, although the tax-people (immediately) carry the (earned) gold, the most precious of all minerals. The ships' crews of every (commercial) house have received their load(s) so that they may depart from Egypt to Djahy. Each man's god is with him. Not one of them (dares) say: "We shall see Egypt again." A carpenter, the one who is in the shipyard, carries the timber and stacks it. If he renders today his produce of yesterday, woe to his limbs! The shipwright stands behind him to say to him evil things. His outworker who is in the fields, that is tougher than any profession. He spends the whole day laden with his tools, tied down to his toolbox. He goes back to his house in the evening laden with the tool box and the timber, his drinking mug and his whetstones. But the scribe, it is he that reckons the produce of all those. Take note of this.

Reference to scribal work is made twice within this section, first when describing how the profit of the merchants is taken away by the tax-people (scribes!), and again at the end of this passage when the profession of the scribe is compared to the aforementioned professions: "But the scribe, it is he that reckons the produce of all those." Both of these references are with respect to the mathematical abilities of the scribe, who has to calculate the taxes of the merchants before carrying them off and who also calculates the output of the other professions, presumably to determine their taxes.

[40] The translation is taken from Caminos, *Late-Egyptian Miscellanies*, pp. 384–5 with some modifications (cf. Tacke, *Verspunkte*, pp. 92–4). The text is also translated in Lichtheim, *Ancient Egyptian Literature*, vol. 2, 169–70.

The satirical letter of Papyrus Anastasi I, another New Kingdom document, even includes direct references to mathematical problems.[41] This text is a fictional letter from the context of a learned debate between two scribes, Hori (the author of the letter) and Amenemope/Mapu (the addressee of the letter). As a response to a letter from Mapu, Hori proposes a scholarly competition covering various aspects of scribal knowledge, among which are several mathematical problems: the calculation of bricks needed to construct a ramp; the number of workers needed to transport an obelisk; the number of workers needed to move sand when a colossal statue has to be erected in a given time; and the calculation of rations for a military excursion. Although these problems are phrased like their earlier counterparts of the mathematical texts, the numerical information given in Papyrus Anastasi I does not suffice to actually solve these problems. Their intention may have been to remind the numerate reader of his mathematical education. To us, the text is a source that provides us with an idea of the variety of numerate tasks that a scribe had to master. In addition, it also informs us about areas of their profession that scribes thought of as meaningful and valuable. Thus, although there is practically no evidence for mathematical texts, administrative documents and evidence from literary texts leave no doubt that mathematics continued to play an important role in the scribal profession during the New Kingdom.

Again, the image can be complemented by an example from the teachings, *The Teaching of Amenemope*, another text to serve as a guideline for a junior scribe (like the Old Kingdom *Teaching of Ptahhotep* and the *Loyalist Teaching* of the Middle Kingdom cited earlier). A comparison of this teaching with its earlier counterparts may be used to reveal some changes in the self-conception of an Egyptian expert.

While the king and his administration still provide the formal framework for the scribal profession, it is the scribe who executes the royal power, as is indicated in the prologue, when Amenemope, the author of the teaching, introduces himself:[42]

> Made by the overseer of the fields, experienced in his office,
> The offspring of a scribe of Egypt,
> The overseer of grains who controls the measure
> Who sets the harvest-dues for his lord,
> Who registers the islands of new land,
> In the great name of his majesty,
> Who records the markers on the borders of the fields,
> Who acts for the king in his listing of taxes,

[41] For an English translation of the mathematical passage see Annette Imhausen, "Egyptian Mathematics," in Victor J. Katz (ed.), *The Mathematics of Egypt, Mesopotamia, China, India, and Islam: A Sourcebook* (Princeton, NJ: Princeton University Press, 2007), pp. 9–12.

[42] Translation from Lichtheim, *Ancient Egyptian Literature*, vol. 2, 148–9.

Who makes the land register of Egypt;
The scribe who determines the offerings for all the gods.

The authority of the scribe, *formally* provided by his being in the service of the king, *de facto* originates from his numerical and metrological expertise, which enable him to execute the tasks mentioned in the quote above. Further references to numerical and metrological duties occur throughout the thirty chapters of instructions. It is explicitly mentioned that a scribe must not "falsify the temple rations" (chapter 5), "move the markers on the borders of fields, not shift the position of the measuring-cord," "be greedy for a cubit of land, nor encroach on the boundaries of a widow" (chapter 6), "move the scales nor alter the weights, nor diminish the fractions of the measure," "desire a measure of the fields, nor neglect those of the treasury," "make for himself deficient weights" (chapter 16), "disguise the measure, so as to falsify its fractions," and "force it (the measure) to overflow, nor let its belly be empty" (chapter 17).[43] During the New Kingdom, the awareness of the scribes about their abilities and the authority connected to them is reflected even in the instructions. Although the scribe remains formally in his position to serve the king, his authority is based on his professional knowledge, especially his numerical and metrological (i.e. mathematical) abilities.

TRADITION, TRANSMISSION, DEVELOPMENT: DEMOTIC MATHEMATICS AS A MIRROR OF EGYPTIAN CULTURE DURING THE GRECO-ROMAN PERIOD

Neither the Third Intermediate Period (1069–664 BCE) nor the Late Period (664–332 BCE) add to the Egyptian corpus of mathematical texts. However, a second group of mathematical papyri, and several ostraca, have survived that date from the Greco-Roman Periods (332 BCE–395 CE).[44] It is from this corpus that a definite influence between the two mathematical cultures of Mesopotamia and Egypt can be traced for the first time.[45] The example par

[43] Translations from Lichtheim, *Ancient Egyptian Literature*, vol. 2, 146–63.

[44] For the demotic mathematical papyri, see Richard A. Parker, "A Demotic Mathematical Papyrus Fragment," *Journal of Near Eastern Studies* 18 (1959), 275–9; Richard A. Parker, *Demotic Mathematical Papyri* (Providence, RI: Brown University Press, 1972); and Richard A. Parker, "A Mathematical Exercise – P. dem. Heidelberg 663," *Journal of Egyptian Archaeology* 61 (1975), 189–96. For the ostraka, see Giulia Belli and Barbara Costa, "Una tabellina aritmetica per uso elementare scritta in demotico," *Egitto e Vicino Oriente* 4 (1981), 195–200; Edda Bresciani, Sergio Pernigotti, and Maria C. Betrò, *Ostraka demotici da Narmuti. I. (nn.1–33)* (Pisa: Giardini, 1983), no. 9; Didier Devauchelle, "Remarques sur les méthodes d'enseignement du démotique," in Heinz J. Thissen and Karl-Theodor Zauzich (eds.), *Grammata Demotika* (Würzburg: G. Zauzich Verlag, 1984), pp. 47–59, no. 3; and Sten V. Wångstedt, "Aus der Ostrakasammlung zu Uppsala. 3," *Orientalia Suecana* 7 (1958), 70–1.

[45] For a comparison of their earlier mathematics see, for example, Ritter, "Measure for Measure."

excellence cited to demonstrate this is a group of problems known as the *pole-against-a-wall-problems*.[46] From Mesopotamia, examples of this type of problem are extant from the Old Babylonian and the Seleucid periods; from Egypt, examples originate from the Greco-Roman Periods only. Egyptian examples can be found in problems 24–31 of the Cairo demotic mathematical papyrus (Cairo DMP).[47] A comparison of problem 24 of the Cairo DMP with its contemporary Mesopotamian version in problem 12 of the Seleucid BM 34568 reveals obvious parallels.[48] The situation described in the setting of the problem again involves a wall and a pole/reed, which is leaning against it. However, in the wording of the problem both the length of the reed and the height of the wall are explicitly mentioned and both of them are to be calculated in the two parts of the problem. The situation found at the beginning is also similar, with the foot of the reed moved a certain given distance away from the wall. As a result, the top of the reed is then the same height as the wall (i.e. the reed is longer than the height of the wall). As Duncan Melville has pointed out, the procedure can be divided into two parts, the calculation of the length of the reed and the calculation of the height of the wall. The solution to the first part "is really rather elegant, and I suspect that anyone stumbling through the algorithm without additional explanation would have little feeling for why the obtained result is correct."[49]

Within the history of Mesopotamian mathematical problems, this type of problem is known from Old Babylonian examples, e.g. problem 9 of BM 85196.[50] Exactly the same problem with an identical procedure/algorithm can be found in another demotic pole-against-the-wall-problem, that is problem 27 of the Cairo DMP. Apart from the wording of two steps, indicated as squaring in the Mesopotamian text, and as a multiplication of a number with itself in the demotic text, the two procedures are identical. Therefore, we assume that the type of problem and its procedure were at some point transmitted from Mesopotamia to Egypt.

Further evidence for a transmission from Mesopotamia to Egypt is supplied from other problem types. The calculation of the circle, distinctly different in the hieratic Egyptian and Old Babylonian Mesopotamian texts underwent a significant change in Egypt from hieratic to demotic problems;

[46] E.g. Parker, *Demotic Mathematical Papyri*, p. 6 and Jens Høyrup, *Lengths, Widths, Surfaces: A Portrait of Old Babylonian Algebra and its Kin* (New York/Berlin/Heidelberg: Springer, 2002), pp. 405–6. For a detailed discussion of this group of problems and their attestations see Duncan Melville, "Poles and Walls in Mesopotamia and Egypt," *Historia Mathematica* 31 (2004), 148–62.

[47] For a translation, see Parker, *Demotic Mathematical Papyri*, pp. 35–40.

[48] For a translation of the respective problem texts see Parker, *Demotic Mathematical Papyri*, pp. 35–6 and Høyrup, *Lengths, Widths, Surfaces*, p. 394.

[49] Melville, "Poles and Walls," pp. 153–4.

[50] Høyrup, *Lengths, Widths, Surfaces*, p. 275. A comparison of the text of the two Mesopotamian examples reveals the fundamental change in terminology. A similar observation can be made when comparing Middle Egyptian with Demotic mathematical texts.

in the latter (e.g. in problem 30 of Cairo DMP) we find a treatment of the circle that clearly resembles the Old Babylonian treatment (i.e. the explicit mentioning of the circumference, which does not appear in hieratic Egyptian problems, as well as a constant of numerical value 3 to indicate the relation between diameter and circumference). While hieratic Egyptian geometric problems are mostly straightforward calculations of the area of a simple geometric object (circle, triangle, rectangle), Old Babylonian and demotic Egyptian problems include examples of calculating more complex geometric shapes, e.g. circle segments (see for example the Old Babylonian tablet BM15285 and problem 36 of Cairo DMP). And the formula used in the Old Babylonian calculation of the area of a trapezium (Yale Babylonian Collection 7290)[51] – later known in the context of Roman land surveyors as the *agrimensor formula* – is also used in the calculation of the area of a rectangle in the demotic BM 10520.[52] Finally, while multiplication tables constitute a significant number of the Mesopotamian mathematical texts, hieratic Egyptian tables include only those for fraction reckoning and the conversion of metrological units – not a single multiplication table. But the demotic BM 10520 has a multiplication table for 64 from 1 to 16.[53]

The extant hieratic mathematical papyri that have a known context (i.e. only the Lahun mathematical fragments) come from a town, and were probably used in the education of scribes for their administrative work. The examples of various other sources referring to scribal professions cited above from the Old to the New Kingdom confirm this setting for ancient Egyptian mathematics. From the Old Kingdom to the New Kingdom, a subtle development in the self-perception of the scribes can be traced. Whereas the scribes of the Old Kingdom understood themselves as executors of royal affairs and perceived the success of their work life according to the expressions of gratitude they received from the king, later scribes seem to have developed a self-esteem that is based on their professional abilities in carrying out the administrative tasks for the king. This professional knowledge also begins to find expression in some of the written texts composed by scribes. Thus, the Rhind mathematical papyrus, the latest of the hieratic mathematical papyri that we have, not only boasts a rather spectacular title, but also includes some problems that are different from the regular problems preparing scribes for the mathematics of their administrative duties. The demotic mathematical papyri, which date from more than a thousand years later than the hieratic material, seem to originate from the context of

[51] The tablet is published in Otto Neugebauer and Abraham Sachs, *Mathematical Cuneiform Texts*, American Oriental Series, 29 (New Haven, CT: American Oriental Society, 1945), p. 44. On the use of the agrimensor formula in earlier Mesopotamian sources, cf. Eleanor Robson, *Mesopotamian Mathematics 2100–1600* BC*: Technical Constants in Bureaucracy and Education* (Oxford: Clarendon Press, 1999), pp. 142–3.

[52] Problems 64 and 65 in Parker, *Demotic Mathematical Papyri*.

[53] Ibid., pp. 64–5.

temples. These temples may have been the places in Greco-Roman Egypt where indigenous Egyptian knowledge was preserved in an otherwise Greek/Roman environment.[54] The number of such problems found in these texts shows that the importance of problems without a direct practical application had, by then, gained further importance in collections of mathematical knowledge.

54 According to Friedhelm Hoffmann, *Ägypten – Kultur und Lebenswelt in griechisch-römischer Zeit: Eine Darstellung nach den demotischen Quellen* (Berlin: Akademie Verlag, 2000), p. 48, since the conquest of Alexander the Great, administration was generally conducted in Greek; only the lowest local levels of state administration, which were in direct contact with the Egyptian population, used demotic. During the Roman period, demotic was completely displaced from public administration.

6

EGYPTIAN MEDICINE

John Nunn

Ancient Egyptian medicine has many features of great importance for the history of medicine. First, it is of great antiquity; evidence of organized ancient Egyptian medicine first appeared in the third dynasty (2686–2613 BCE). Secondly, there are copious sources of information, including two- and three-dimensional art, as well as lengthy texts relating to doctors and to persons with disabilities. Medical papyri dating from the Middle Kingdom (2040–1795 BCE), usually written in hieratic script, were intended to guide doctors in the treatment of their patients. In addition, there are lengthy ancient Egyptian herbals, which describe the plants upon which much of the treatment was based. Finally, the dry climate of Egypt, as well as the practice of mummification, ensured the preservation of very many bodies, both healthy and diseased.

Egyptian medicine remained remarkably self-contained and conservative, until foreign influence increased in the Late Period (664–332 BCE), during which there were two separate Persian occupations. Greek medicine came to Egypt during the Ptolemaic Period (305–30 BCE), and anatomical research by Herophilus (b. 325, died ca. 255 BCE) flourished in Alexandria. Galen worked in Alexandria in the Roman Period (30 BCE–323 CE).

THE MEDICAL PAPYRI

The medical papyri are of inestimable value in understanding ancient Egyptian medicine. Their general format would accord with their being used for general reference. It is unclear whether they would have been in the hands of individual practitioners or a central repository; the great length of the surviving papyri and the paucity of surviving copies suggest the latter. Survival of these papyri was very much a matter of chance and there must be many that are lost.[1] The most important are considered below, where the dates relate to surviving copies.

[1] Unfortunately the actual find site of medical papyri is often unknown.

THE EDWIN SMITH SURGICAL PAPYRUS

The Edwin Smith surgical papyrus (1550 BCE) was purchased in Luxor in 1862 by Edwin Smith himself, an American Egyptologist and collector of antiquities who lived in Egypt during the second half of the nineteenth century. It is now held in the New York Academy of Sciences. J. H. Breasted prepared a monumental multi-volume work on the papyrus, which included a photographic reproduction of the surviving original, hieroglyphic reproduction, translation, and extensive commentaries on the forty-eight cases which have survived in the papyrus as it was found.[2]

Most cases are related to injury and are listed in vertical order according to the site of the injury, starting with a head wound and descending to finish with a strained spinal vertebra. It seems likely that further material relating to injuries of the lower parts of the body has not survived. The systematic approach to case reporting, the (usually) logical approach to prognosis and treatment, and the relative absence of magical spells in the Edwin Smith papyrus is noteworthy.

THE EBERS PAPYRUS

The longest and most widely ranging papyrus is the Ebers (1500 BCE), which was purchased in Luxor by Edwin Smith in 1862. It was subsequently purchased in 1872 by Georg Ebers, by whose name it is now known. It is kept in the University of Leipzig. A full reproduction, along with the classical translation in German and an extensive commentary, is given in the *Grundriss der Medizin der alten Ägypter* (1958), subsequently translated into English by P. Ghalioungh ui (1987).[3]

It comprises 877 sections (some subdivided), arranged in a rather haphazard order and in variable style. However, certain groups of sections are identifiable as devoted to particular medical problems, for example, the book of the stomach (sections 188–207), diseases of the heart (217–20), diseases of the eyes (336–41), treatment of bites (432–6), treatment of the liver (477–81), dentistry (739–49), and diseases of women (783–953). The symptoms and findings at medical examination are often recognizable in modern western experience of medicine, but the basis of the recommended treatment is usually difficult to understand today. This is in stark contrast to the Edwin Smith papyrus. However, in the absence of effective remedies, the

[2] J. H. Breasted, *The Edwin Smith Surgical Papyrus*, 2 vols. (Chicago, IL: University of Chicago Press, 1930).

[3] H. von Deines, H. Grapow, and W. Westendorf, *Grundriss der Medizin der altern Ägypter*, 5 vols. (Berlin: Äkademie Verlag, 1958–62); P. Ghalioungh ui, *The Ebers Papyrus* (Cairo: Academy of Scientific Research and Technology, 1987).

placebo benefit of ancient Egyptian remedies, incantations, and spells should not be discounted.

Ten further medical papyri are listed as follows:[4]

Kahun	1820 BCE	University College, London	Gynecological
Ramesseum III–V	1700 BCE	Oxford Univ., UK	General
Hearst	1450 BCE	California	General
Chester Beatty VI	1200 BCE	British Museum	Rectal diseases
Berlin	1200 BCE	Berlin	General
London	1300 BCE	British Museum	Mainly magical
Brooklyn snake	300 BCE	Brooklyn	Snake bite
Carlsberg VIII	1300 BCE	Copenhagen	Gynecological
London & Leiden	250 CE	British Museum and Leiden	General
Crocodopolis	150 CE	Vienna	General

THE PATTERN OF DISEASE

The expectation of life in ancient Egypt is difficult to determine, but the best estimates are about thirty years for the Predynastic Period (up to 3100 BCE) and thirty-six thereafter.[5] This contrasts with seventy-two for modern Egypt. The average nutritional state in Pharaonic times was probably amongst the best in countries of the ancient world, and there are many illustrations of obesity. However, famines were known to have occurred, probably as a result of failure of the Nile to rise.

A major problem was parasitic diseases including *schistosomiasis* (bilharzias), *dracunculiasis* (guinea worm), *filariasis*, and *strongyloidiasis*, evidence for which has been found in mummies.[6] Effective treatments were not available. Bacterial and viral diseases also occurred and several mummies show classical spinal tuberculosis, the best example being the priest Nesparehan.[7] The incidence of tropical diseases in ancient Egypt was reviewed, with special reference to intestinal worms, malaria, plague, leprosy, and tetanus, by J. F. Nunn and E. Tapp.[8] In contrast, and probably as a result of the short expectation of life, as well as the incidence of certain carcinogens being lower than today, evidence of cancer in ancient Egypt is remarkably scarce.

[4] These papyri are briefly described in J. F. Nunn, *Ancient Egyptian Medicine* (London: British Museum Press, 1996).

[5] Ibid.

[6] A. R. David, *The Manchester Museum Project* (Manchester: Manchester University Press, 1979).

[7] Reported by M. A. Ruffer, *Potts'che Krankheit an einer ägyptischer Mumie aus der Zeit der 21 Dynastie* (Giessen: Verlag von Alfred Töpelmann, 1910).

[8] J. F. Nunn and E. Tapp, "Tropical Diseases in Ancient Egypt," *Transactions of the Royal Society of Tropical Medicine and Hygiene* 94 (2000), 147–53.

Statues, reliefs, and skeletons give convincing evidence of deformities. These include dwarfism, achondroplasia, hydrocephalus, and club-foot.[9] Hernias and/or hydrocoeles were clearly shown in reliefs. Statues of Akhenaten, and also the Queen of Punt, portrayed on the walls of Hatshepsut's temple at Dier el Bahri, remain as diagnostic problems; a variety of medical explanations have been proffered for their apparently abnormal physiology, though consensus for a precise diagnosis has yet to be reached.

The concentration of the ancient Egyptians on their digestive systems and the book of the stomach (Ebers papyrus 188–208) leave no doubt of their recognition of, and concern with, constipation, for which many remedies were recommended.

Migraine was described in Ebers 250 as "suffering in half the head (*ges-tep*)." The Hippocratic corpus also referred to pain in half the head (*hemicrania*) – from which the word "migraine" is derived.

HEALTH CARE PROFESSIONALS

There is no doubt that the Egyptian word *swnw* means something very close to "doctor of medicine." The first known *swnw*, Hesy-ra, lived in the third dynasty (2686–2613 BCE), and the word *swnw* survived as *saein* in Coptic. In the doctors known to us, the title *swnw* was frequently embellished by *wr*, meaning "great," or "chief of" (doctors). *Imy-r swnw* means literally "overseer of doctors." Royal connections are not unusual in doctors known to posterity, as in *swnw per aa*, "doctor of the palace." Their specialty was also indicated as, for example, in *swnw irty*, "doctor of the two eyes or ophthalmologist." Ir-en-akhty, a *swnw per aa* in the First Intermediate Period (2181–2040 BCE) is remarkable for the number of his specialties, including ophthalmology, gastroenterology, and "herdsman of the anus." The feminine termination of *swnw* appears in the stela of the lady Peseshet (fifth–sixth dynasty, 2494–2181 BCE). Her title was *imy-r swnwt*, "supervisor of the female physicians." Whether or not Peseshet was herself a *swnwt* is uncertain, but it would indeed appear that female doctors (*swnwt*) did exist. No other Egyptian female doctor is known until at least the Ptolemaic Period (305–30 BCE). The author has tabulated 150 known *swnw*.[10] Nothing is known of the range of social status to which working *swnw* belonged. Some were of such exalted status that it seems doubtful whether they were actually medical practitioners; they may have been "honorum."

There is also evidence of a variety of other medical practitioners. The wooden stela of Hesy-ra, the first known Egyptian doctor mentioned

[9] On achondroplasia, see Seneb in Nunn, *Ancient Egyptian Medicine*.
[10] See ibid.

above, indicates that he was also an *ibeh*, a word with a tusk as the determinative, which is translated as dentist. He must have been too illustrious to engage in dental practice, but there were others who carried the title of *ibeh*. Three of the five *ibeh* of the Old Kingdom also carried the title of *swnw*. There are no known ancient Egyptian words for gynecologist, obstetrician, or midwife, and no firm evidence for the role of the *swnw* in childbirth. Indeed, the medical papyri are silent on childbirth. Surviving reliefs portraying childbirth tend to show goddesses in attendance.

EDUCATION OF HEALTH CARE PROFESSIONALS

There is a serious lack of information on the status, family connections, and training of the *swnw*. We are really reliant on inscriptions rather than descriptions of these aspects of the *swnw*. There was a powerful ancient Egyptian tradition for sons to follow in the calling of their fathers, but there is no inscription which states categorically that a *swnw* had followed directly in the profession of his father. However, F. Jonckhere reported two families, one of which contained two doctors and another three, but without filiation being specified.[11] Furthermore, the stela of Iuny in the Ashmolean Museum shows two known *wr swnw*, Huy and Khay. Khay was described, almost certainly in relation to Huy, as "his son who causes his name to live."

The method of training doctors is not made clear in the available literature. It has been suggested that the *Per Ankh* (the house of life) might have been something analogous to a medical school. However, it seems more likely that instruction of the *swnw* was as an apprentice, possibly in the hands of the father or a near relative. The medical papyri with their innumerable remedies indicate that there was much for the student to learn.

MEDICAL KNOWLEDGE

CONCEPTS OF ANATOMY AND PHYSIOLOGY

Embalmers, who were responsible for preparing the body for burial during the process of mummification, must have had considerable anatomical knowledge, particularly as a result of removing liver, lungs, stomach, and intestines for storage in their respective canopic jars (the vessels buried alongside the body in which the extracted organs were preserved). A most remarkable achievement was the removal of the brain through the nose, which required perforation of the ethmoid bone. Surprisingly, the brain was

[11] F. Jonckheere, *Les Médecins de L'Égypte Pharaonique* (Brussels: Fondation égyptologique de la reine Élisabeth, 1958).

one of the few organs that the ancient Egyptians did not attempt to preserve, removing it to prevent it from decaying within the body after burial.

Ancient Egyptian anatomical terminology was extensively studied by J. H. Walker.[12] It appears likely that the ancient Egyptians knew of the existence of the following internal organs: brain; lungs; heart; spleen; gall bladder; stomach; intestines; colon; rectum; uterus; kidneys; and bladder. However, they had a very important anatomical concept with no obvious parallel in modern practice. The *metw* (plural of the word *met*) was used to cover a diverse range of long narrow structures, connecting different parts of the body, including blood vessels, ducts (including ureters and spermatic cords), tendons, and possibly nerves. *Metw* are listed in two medical papyri: Ebers 854 and 856, and Berlin 163.

The entire Pharaonic Period was long before the discovery of the circulation of the blood. However, Ebers (854a) shows an appreciation that the beating of the peripheral pulses reflected the beating of the heart that "speaks from the vessels of all the limbs." There was frequent reference to the breath of life (*tjaw n ankh*) as of life-giving importance, entering the nose and thence, by way of the *metw*, to the heart and all the body (Ebers 855a).[13]

The ancient Egyptians clearly understood that food entered the stomach and the residue went down to the anus. The kidneys are not mentioned in the medical papyri, but it was known that the bladder contained urine. It is unclear whether the ancient Egyptians had any clear understanding of the nervous system, and it may be significant that the brain was never preserved at mummification, suggesting that it was considered unnecessary in the afterlife. However, the Edwin Smith papyrus indicates that transmission of sensation from the lower part of the body was lost in spinal injuries (case 31).

The essentials of reproduction were well known. Ebers 854i states "there are two *metw* to his testicles (*kherwy*: 'those two which are underneath'). It is they which give semen." However, the ovaries may well have been unknown. The Egyptian word for placenta was *mwt remt*, "the mother of mankind."

THE ROLE OF MAGIC AND RELIGION

Spells and incantations featured prominently in the medical papyri, and were often directed at the disease or disease-demon itself. The concept is introduced formally in the first three sections of the Ebers papyrus. In contrast, the Edwin Smith papyrus is relatively less concerned with magical elements of medical practice, with only one magical spell in the entire papyrus. Most incantations were to be recited in relation to another

[12] J. H. Walker, *Studies in Ancient Egyptian Anatomical Terminology* (Australian Centre for Egyptology Studies, 4; Warminster: Aris and Phillips, 1996).

[13] If we were to translate *ankh* as "oxygen" instead of "life," this is close to the truth, although it was at a time when arteries were thought to contain air and not blood.

remedy (more conventional by modern standards). Whatever the actual efficacy of the remedy, incantations must have provided a powerful (placebo) effect.

MEDICAL PRACTICES

The Edwin Smith papyrus offers crucial insight into medical practice in ancient Egypt. The standardized presentation of each case consists of Title, Examination, Diagnosis and Prognosis, and Treatment. The Title starts with the words "Instructions for a ... " followed by a brief description of the injury, for example: "Instructions for a split in his cheek" (case 1). The Examination starts with the words: "If you examine a man having ... " followed by a repetition of the title, and then further detail of the examination. Thus the examination of case 1 continues with: "If you examine a man having a split in his cheek and you find there is a swelling, raised and red on the outside of his split." Examination is followed by Diagnosis and Prognosis, which in case 1 is as follows: "You shall say concerning him: One having a split in his cheek. An ailment which I will treat." There are three options in treatment. The first is, as in case 1, "An ailment which I will treat" (this pronouncement occurs in thirty cases). The second is "An ailment with which I will contend" (occurs in eight cases), interpreted as palliative treatment, and thirdly "An ailment not to be treated" (occurs in fourteen cases), interpreted as an injury with a hopeless prognosis. The final section outlines the Treatment. In case 1 it reads: "You should bandage it with fresh meat on the first day. His treatment is sitting until his swelling is reduced. Afterwards you should treat it with grease, honey and a pad every day until he is well." There is only one magical spell in the entire papyrus, and the logic of the treatment is generally recognizable to a modern doctor.

Case 1, outlined above, is remarkably simple. Many other cases are far more complicated and the section on Diagnosis and Prognosis may be updated as the condition of the patient changes, usually for the worse. A celebrated example is case 7, which may indicate the development of tetanus in a case of head injury. The first Diagnosis and Prognosis says:

> You shall say concerning him: one having a gaping wound of the skull, extending to the bone and penetrating the *tepau* (meaning uncertain) of his skull. The "cord" of the mandible is contracted; he discharges blood from his two nostrils and from his two ears; and he suffers stiffness in his neck. An ailment with which I will contend.

The "cord" (*metu*) of the mandible probably refers to the jaw muscles and the passage suggests the onset of spasm of the jaw, which is a warning sign of tetanus. The second pronouncement of Diagnosis and Prognosis says:

> You shall then say concerning him: one having a gaping wound in his head, extending to the bone and penetrating the *tepau* of his skull. He has developed toothache; his mouth is bound; he suffers stiffness of his neck. An ailment not to be treated.

Many cases contain one or more Glosses to explain aspects of the case that the reader might find difficult to understand. For example, case 12 starts: "Instructions for a break in the column of his nose." This is supplemented by Gloss A, which says: "As for: 'the column of his nose,' it means the outer edge of his nose as far as its side(s) on the top of the nose, being the inside of his nose in the middle of his two nostrils."

SURGERY AND OTHER SPECIALTIES

There is evidence of the skillful management of wounds and other injuries in ancient Egypt, but there is a dearth of information on the practice of surgery outside the sphere of trauma. The instruments shown in the relief on the outer wall of the temple in Kom Ombo all date from the Roman Period. However, a relief from the tomb of Ankh-ma-hor undoubtedly shows a circumcision.[14] Skeletal remains show evidence of skillful reduction of fractures of long bones.

The Ebers papyrus has a section (863–77) with references to "knife treatment" for a range of medical disorders, but details are lacking. It seems unlikely that they ventured far beyond the opening of an abscess or the removal of a foreign body. The Egyptians were capable of superb needle work, but no mummy has shown evidence of *in vivo* suturing of skin. However, Edwin Smith surgical papyrus case 28 contains the words "you should then draw together the wound with stitching." More commonly it was written that "the lips of the gash should be drawn together with bandages," a skill which was well advanced in mummification. Many excellent copper needles have been found in Egypt, dating back to before the Bronze Age in Predynastic Egypt.[15]

[14] Nunn, *Ancient Egyptian Medicine*.

[15] Ibid. The author has since demonstrated that no particular skill was required to prepare copper from malachite, to form the eye of a needle, and then to work-harden the needle so that it could be used to suture skin; see J. F. Nunn and J. Rowling, "The Eye of the Needle in Predynastic Egypt," *Journal of Egyptian Archaeology* 87 (2001), 171–2.

DENTAL PRACTICE

The Edwin Smith papyrus case 25 carries a superb description of the successful reduction of a dislocated jaw:

> His mouth is open and does not close for him. You place your fingers on the back of the two rami of the mandible inside the mouth, your two claws (i.e., thumbs) under his chin and you cause them (the two mandibles) to fall so they lie in their [correct] place.

It is not known whether this procedure was undertaken by a *swnw* or an *ibeh*.

In spite of the wealth of skeletal remains, there is virtually no surviving archaeological evidence of work undertaken on teeth. R. J. Forshaw drew attention to radiographs of the teeth of Amenhotep III and Ramesses II showing extremely worn teeth, periapical abscesses, and advanced periodontal disease.[16] He stressed that, if operative dental treatment was available, surely it would be evident in these two pharaohs of great distinction. He drew the conclusion that, if operative dental care did exist at all, it was extremely limited.

In contrast, nonoperative care, largely directed towards diseases of the gums, using medication of doubtful efficacy is detailed in the Ebers papyrus (739–49). Skeletal remains in ancient Egypt show dental caries to be remarkably rare until the first millennium BCE. This has been attributed to the absence of sugar in the diet and the high cost of honey. The first millennium BCE saw the introduction of sugar into the diet, when dental caries became common.

GYNECOLOGY AND OBSTETRICS

Ebers papyrus 783–834 contains a range of remedies for women's disorders of a medical rather than a surgical nature. The actual nature of the disorder is often far from clear. For example, Ebers papyrus 823 comprises a remedy to contract the uterus, but it is not clear whether this is intended to hasten birth, to expel the placenta, or to assist the return of the uterus to a normal size after delivery. Ebers papyrus (828–33) contains remedies for problems related to menstrual periods.

The Kahun, Berlin, and Carlsberg papyri contain tests for determining fertility, pregnancy, and even the sex of the unborn child. In one such test, a mixture of emmer and barley seeds was moistened daily with urine of an pregnant woman. If both grew, she was indeed pregnant. If only the barley grew the child would be male; if the emmer grew the child would be

16 R. J. Forshaw, "The Practice of Dentistry in Ancient Egypt," *British Dental Journal* 206 (2009), 481–6; J. E. Harris, A. T. Storey, and P. V. Ponitz, "Dental Disease in the Royal Mummies," in J. E. Harris and E. F. Wente (eds.), *An X-ray Atlas of the Royal Mummies* (Chicago, IL: University of Chicago Press, 1980), pp. 328–46.

female; if neither grew she was not pregnant. This was put to the test by P. Ghaliounghui, S. Khalil, and A. R. Ammar. They found one or other seed germinated in twenty-eight cases out of forty pregnant women, but no growth failed to exclude pregnancy in thirty percent of cases. Prediction of the sex of the offspring was correct in seven cases, but incorrect in sixteen.[17]

OPHTHALMOLOGY

There are many representations of eye diseases in ancient Egypt, and eye injuries played an important role in Egyptian mythology. For example, the eye of the falcon god Horus was torn out by Seth, one of the sons of the earth god, but magically restored. Despite the existence of specialist eye doctors, the *swnw irty*, the papyri offer little information on their practice. The only mention of an eye problem in the Edwin Smith papyrus is case 11, which comprises a gaping wound of the eyebrow, for which closure of the wound by stitching is recommended. Part of the Ebers papyrus (336–431) is devoted to diseases of the eye, but the emphasis is on medical rather than surgical problems and treatment of acuity of vision is not considered.

DRUG TREATMENT

The pharmacopoeia of ancient Egypt was vast and, although therapeutically weak by modern standards, would have seemed reasonable only 200 years ago. Clearly there was significant reliance on what is now understood as the placebo effect. Drugs fell into three main groups: mineral, animal, and plant.[18] In each group there were preparations for both internal and external use.

There were at least twenty-seven drugs of mineral origin, including common salt, malachite, lapis lazuli, and gypsum. Many of the mineral drugs were insoluble in body fluids, and it is difficult to see a basis for their therapeutic efficiency.

A very wide range of animal products was used, including honey, urine, blood, bile, fat, liver, testicles, and even placenta, and in addition the excrement of cat, ass, birds, lizard, crocodile, fly, and man. It is difficult to see any therapeutic logic. Perhaps the intention was to impart some aspect of the donor creature.

The vast majority of ancient Egyptian drugs consisted of extracts of plants. A major difficulty is identifying the botanical species represented

17 P. Ghaliounghui, S. Khalil, and A. R. Ammar, "On an Ancient Egyptian Method of Diagnosing Pregnancy and Determining Foetal Sex," *Medical Historian* 7 (1963), 241–6.

18 L. Manniche, *An Ancient Egyptian Herbal* (London: British Museum Publications, 2006).

by an Egyptian word. However, there are certain items with definite pharmacological activity, including castor oil, cannabis, acacia, and opium (imported from Cyprus). Pharmacological properties currently attributed to identifiable ancient Egyptian botanical species have been summarized by the author.[19] Ancient Egyptian drugs were dispensed by volume and not by weight.

[19] See Nunn, *Ancient Egyptian Medicine*, Appendix C.

7

EGYPTIAN CALENDARS AND ASTRONOMY

Rolf Krauss

INTRODUCTION

The beginnings of astronomy and the calendar in Egypt can be traced back to the third millennium BCE. Evidence for the 365-day solar calendar, for a lunar day-count, and for the observation of stars to divide the night into hours dates to this early period; water clocks and shadow clocks came later. Throughout pharaonic history astronomy was linked to the notion of a celestial Hereafter and lore about stellar gods. In the first millennium BCE Egyptian culture came under the influence of Babylonian astrological astronomy; from the Hellenistic period down through Greco-Roman times, Egypt was a center of astrology.

The earliest evidence for Egyptian "sky-watchers" dates to the early second millennium BCE. Contemporary texts use the title "hour man," a descriptive term presumably referring to a person observing the hour-stars.[1] A millennium later the title "hour watcher" is documented. The existence of titles such as "chief of hour-watchers" shows that the profession was organized like others. A Late Period description of astronomer's duties included

> knowing the time of the rising and setting of stars, especially Sothis (Sirius), the progress of the sun towards north or south, the proper length of the hours of daytime and night, and the proper performance of rituals, as well as charms against scorpions.[2]

COSMOGRAPHY, GEOGRAPHY, AND ORIENTATION

According to Late Period sources the Egyptians conceived of the earth as a disk in the midst of the ocean, beneath a celestial ocean on which the heavenly

[1] A. H. Gardiner, *Ancient Egyptian Onomastica*, 2 vols. (Oxford: Oxford University Press, 1947), vol. 1, 61*–2*.

[2] P. Derchain, "Harkhébis, le psylle-astrologue," *Chronique d'Égypte* 64 (1989), 74–89.

bodies sailed. A parallel concept imagined a sky goddess whose body arched above the earth god and who gave birth to the stars and the sun.[3]

Royal funerary literature from about 1500 BCE describes the nocturnal journey of the sun god through the caves of the underworld from west to east. Dimensions of earth and sky are implied by the number of miles cited in the texts: the sun god covered a distance of about 15000km during his underworld journey, the path covered during an entire day being presumably twice as long.[4]

South was considered "forward," north was thus "behind"; the west was at the right and the east to the left.[5] The subdivisions SW, NW, NE, and SE supplemented the four cardinal points. The Nile, which flows in general from south to north, offered a main reference line for orientation; axes of temples often lie perpendicularly to the local direction of the Nile. Sacral buildings include elements of astronomical orientation; study of such phenomena was initiated in the late nineteenth century,[6] and revived in the 1960s.[7] Since then it has been successfully pursued in relation to specific buildings[8] and by measuring most accessible temples.[9]

The precise orientation of pyramids erected during the Fourth Dynasty (ca. 2600–2450 BCE) implies the existence of commensurate astronomical knowledge. There is no general agreement on how true north was determined. Scholars have proposed plausible methods which do not require advanced astronomical techniques,[10] and criticized less plausible ones.[11]

ASTRONOMY IN FUNERARY TEXTS

The Old Kingdom Pyramid Texts (PT),[12] Middle Kingdom Coffin Texts (CT),[13] and the Book of the Dead from the New Kingdom onwards form

[3] E. Hornung, "Himmelsvorstellungen," in W. Helck and W. Westendorf (eds.), *Lexikon der Ägyptologie*, 7 vols. (Wiesbaden: Harassowitz, 1975–92), vol. 2, 1211–18.

[4] K. Ferrari d'Occhieppo, "Die Gefilde der altägyptischen Unterwelt: Spiegelbild der Sonnenbahnen im Jahreslauf," *Zeitschrift für ägyptische Sprache und Altertumskunde* 123 (1996), 103–10.

[5] G. Posener, "Sur l'orientation et l'ordre des points cardinaux chez les Égyptiens," in *Nachrichten der Göttinger Akademie der Wissenschaften* (1965), pp. 69–78.

[6] J. N. Lockyer, *The Dawn of Astronomy* (London: Cassell, 1894).

[7] G. S. Hawkins, "Astronomical Alinements in Britain, Egypt and Peru," in F. R. Hodson (ed.), *The Place of Astronomy in the Ancient World* (London: Oxford University Press, 1974), pp. 164–5.

[8] L. Gabolde, "Mise au point sur l'orientation du temple d'Amon-Rê à Karnak en direction du lever du soleil au solstice d'hiver," *Cahiers de Karnak* 13 (2010), 234–56.

[9] J. A. Belmonte, "Orientation of Egyptian Temples: An Overview," in Clive L. N. Ruggles (ed.), *Handbook of Archaeoastronomy and Ethnoastronomy* (Berlin: Springer, 2014), pp. 1501–18.

[10] Z. Zába, "L'orientation astronomique dans l'ancienne Égypte et la précession de l'axe du monde," *Archiv orientální. Supplementa* 2 (1953), 9–74.

[11] A.-A. Maravelia, "The Stellar Horizon of Khufu," in S. Bickel and A. Loprieno (eds.), *Basel Egyptology Prize 1: Junior Research in Egyptian History, Archaeology, and Philology* (Aegyptiaca Helvetica 17; Basel: Schwabe & Co., 2003), 55–74.

[12] R. O. Faulkner, *The Ancient Egyptian Pyramid Texts* (Oxford: Clarendon Press, 1969).

[13] R. O. Faulkner, *The Ancient Egyptian Coffin Texts*, 3 vols. (Warminster: Aris & Philipps, 1973–8).

a historically related corpus covering the period from about 2400 BCE down into Roman times. These texts, which presume a celestial Hereafter, provide insights into early astronomical concepts.[14]

The night sky appears to be divided by the "winding canal" which corresponds roughly to the non-Egyptian concept of the ecliptical belt. The waxing moon crosses the "winding canal" as "He-who-looks-forward," facing forward in the direction of its ecliptical movement, while the waning moon is "He-who-looks-backward." In the morning the sun god boards the solar barque, moored at one "bank" of the "winding canal" (PT 548, 569).

The "two heavens" comprise the fixed stars north and south of the "winding canal." The northern fixed stars – the "imperishable ones" – row the solar barque during the day; they expressly include stars that rise and set. As blessed souls they are described as "yonder people of whom it has been said: You have not died the death" (CT 72).

The southern fixed stars – the "unwearying ones" – row the nocturnal solar barque. Their movement conforms to the annual rhythm of ascending in the east, then descending in the west for a period of invisibility; they, too, embody blessed souls.

The dichotomy mirrors astronomical facts. The stars in the ecliptical belt and south of it are invisible between heliacal setting and subsequent rising when the sun moves through their vicinity. This behavior was interpreted as dying and reviving, or as "renewing and rejuvenation at the due season" (PT § 883). The fixed stars north of the ecliptical belt, because their heliacal rising precedes setting, are visible every night and are thus "imperishable." It follows that the circumpolar stars to which the Big Dipper (Mesekhtyu) belongs represent a subgroup of the "imperishable ones."

CIVIL CALENDAR

The earliest evidence for reckoning time year by year dates to Dynasty I (ca. 3000 BCE), but the form of the calendar is not known.

Djoser's reign (ca. 2650 BCE) provides the earliest contemporaneous evidence for the civil calendar of 365 days divided into three seasons (akhet = inundation, peret = sowing, and shemu = harvest), each having four months of thirty days, plus five intercalary (epagomenal) days at the end of the year.

From ancient times the twelve months were not numbered consecutively but divided into three groups of four each, by season. A few month names have survived from about 1800 BCE. Varying series are documented half

[14] Rolf Krauss, *Astronomische Konzepte und Jenseitsvorstellungen in den Pyramidentexten* (Wiesbaden: Harrassowitz, 1997).

a millenium later from which the Greek, Aramaic, and Coptic names of the civil months derive. The entire year comprised thirty-six decades (ten day periods) plus five days. A truncated year, lacking the epagomenal days, was also in use.[15] An example is the temple budget year of 360 days implied in the statement that the daily income of a temple is 1/360 of the yearly revenue. A disregard of the five epagomenal days is shown by their often being left out in calendrical schemes and in day-counts which span more than a year.

Because of the civil calendar's regularity, it was well suited for administrative purposes and for everyday use. The royal bureaucracy employed the civil calendar for taxation, for duty rosters of the work force, for issuing provisions, and also for dating decrees and commemorative stelae. Civil dates were cited during most of the pharaonic period according to the regnal year of the reigning king, either coinciding with the calendar year or determined by the day of his accession. Use of the civil calendar in non-royal contexts is attested, for example, in dating of letters and contracts; civil dates of birth and death were first recorded for members of the elite and sacred animals in the Late Period.

Following the Greco-Roman astronomer Claudius Ptolemy, astronomers continued to employ the civil calendar down to the Renaissance. Double dates in his writings correlate the Egyptian civil and the Julian calendar. Ptolemy's data substantiate the conclusion implied by pharaonic sources that the Egyptian calendar day began before sunrise at dawn.[16] The Egyptian civil calendar was the precursor of the Julian calendar and thus of the modern Gregorian calendar. Neugebauer asserted that the Egyptian civil calendar was "the only intelligent calendar which existed in human history."

The civil calendar represents a solar year. Since the heliacal rising of Sothis marked the ideal beginning of the year, the civil calendar was actually a stellar year according to modern terminology. Since there was no "leap year," the date of Sothis' rising and with it the civil calendar and its seasons fell one day behind the tropical year every four years. The concept of a "Sothic cycle" for a complete shift of 1,460 years (that is, 365 x 4) is first attested in Hellenistic times.[17] Reckoning backwards on the basis of the few documented dates of heliacal risings, it would seem that the civil calendar was introduced, and the first historic Sothic cycle begun, about 2750 BCE. But the reliability of such reasoning is contested, because the heliacal rise of Sothis is not mentioned explicitly before ca. 2000 BCE.[18]

[15] Kurt Sethe, "Die Zeitrechnung der alten Ägypter im Verhältnis zu der der anderen Völker," in *Nachrichten der Gesellschaft der Wissenschaften zu Göttingen* (1919–20), pp. 302–7.

[16] A. J. Spalinger, *Five Views on Egypt* (Göttingen: Seminar für Ägyptologie und Koptologie, 2006), pp. 51–85.

[17] F. K. Ginzel, *Handbuch der mathematischen und technischen Chronologie*, 3 vols. (Leipzig: Hinrichs, 1906–14), vol. 1, 191.

[18] For earlier Sothic dates, though included in later text copies, see A. J. Spalinger, *Three Studies on Egyptian Feasts and their Chronological Implications* (Baltimore, MD: Halgo, 1992), pp. 29–30.

The date on which the star rises differs by about one day for each degree of latitude, so that Sothis rises up to nine days later on the Mediterranean coast than at the first cataract, Egypt's southern border in antiquity. Memphis in northern Egypt is the attested geographical reference site for the calendrical heliacal risings of Sothis in late antiquity.[19] The location of the reference site in earlier pharaonic times is subject to discussion.

There are a few remarks in pharaonic period texts about calendric disorder, and these appear to be literary topoi characterizing disorder in general (*Chaosbeschreibung*).[20] The calendric shift of religious festivals is only mentioned in the Decree of Canopus as a disadvantage resulting from the lack of a leap day. With this decree Ptolemy III intended the introduction of a Sothic fixed year by adding a leap day every four years; the reform was in principle abortive.[21] Augustus succeeded in reforming the civil calendar by adding a leap day: this so-called Alexandrian calendar remained in use in Egypt until the Arab conquest (640 CE); it still serves today as the liturgical calendar of the Coptic and Ethiopian churches.

LUNAR CALENDARS

Texts assert that the moon is "conceived" on the first lunar day, is "born" or "rejuvenated" on the second, and "grows old" after the fifteenth. These statements imply that the first lunar day coincides with invisibility of the moon. Egyptologists generally agree that the day *after* the last sighting of the waning moon was counted as the first day of a lunar month. Corroboration follows from late double dates equating lunar days with days in the civil calendar.[22]

Lunar days were designated by names or by ordinal numbers; names for each lunar day are documented from the Late Period.[23] Beginning around 2500 BCE private and royal lists record cultic feasts on lunar days 1, 2, 6, 7 (first quarter), 15 (full moon), and 23 (last quarter).[24]

Temple service schedules are attested according to civil or lunar months. Sources from the Middle Kingdom and from Ptolemaic-Roman times imply

[19] Sethe, " Die Zeitrechnung", p. 309.

[20] R. Weill, *Bases, méthodes et résultats de la chronologie égyptienne*, 2 vols. (Paris: Geuthner, 1926–8), vol. 1, 107–11.

[21] C. Bennett, *Alexandria and the Moon: An Investigation into the Lunar Macedonian Calendar of Ptolemaic Egypt* (Leuven: Peeters, 2011), pp. 179–86.

[22] R. A. Parker, *The Calendars of Ancient Egypt* (Chicago, IL: University of Chicago Press, 1950), pp. 9–23; R. Krauss, "Babylonian Crescent Observation and Ptolemaic-Roman Lunar Dates," *PalArch's Journal of Egypt/Egyptology* 9 (2012), 1–95, 23–4, 29–30. Access at <www.PalArch.nl>.

[23] Parker, *Calendars*, pp. 11–13.

[24] For the traditional view see H. Altenmüller, "Feste," in Helck and Westendorf (eds.), *Lexikon der Ägyptologie*, vol. 2, 172–3; for a critical examination of presumed lunar feasts, see A. J. Spalinger, *The Private Feast Lists of Ancient Egypt* (Wiesbaden: Harrassowitz, 1996).

that the temple service according to the lunar month began on the second lunar day (new crescent) and ended on the first (invisibility).[25]

Borchardt deduced a stellar-based lunar calendar from data in the Illahun archive (ca. 1850 BCE) and in Papyrus Ebers (ca. 1500 BCE).[26] One account from Illahun lists a series of months comprising a lunar year of 354 days, beginning, apparently, with the first lunar month after Sothis' rising. The same form of a lunar year is implied by various dates of the wagy-feast, which fell on a certain lunar day within a lunar month that shifted together with the dates for the rising of Sothis through the civil calendar. The calendar of Papyrus Ebers provides a synchronism of a series of presumably lunar months with civil months; it also notes the date of Sothis' rising. The calendar's structure is consistent with the calendaric data in the Illahun archive.

Parker modified Borchardt's ideas to create a model of a stellar-lunar year of twelve months which began on the first day of the lunar month after the rising of Sothis; the model accommodated the moveable *wagy*-feast. Occasionally, a thirteenth month was intercalated to ensure that the last month of the year witnessed the star's rising.[27] Such a stellar-lunar year would have remained synchronized with the seasons. Its use may have been restricted to temples; there are no unequivocal documentary traces later than Papyprus Ebers.

Parker based the contention that there was a lunar calendar running parallel to the civil calendar on the existence of civil-lunar double dates citing a lunar day and also a lunar month;[28] but his idea has been disputed.[29]

At issue is the interpretation of Papyrus Carlsberg 9, which preserves a lunar cycle consisting of 309 lunar months or 25 civil calendar years.[30] The papyrus dates from 144 CE at the earliest. The relation of the cyclical lunar dates to those astronomically computed points to the introduction of the scheme in early Hellenistic times. The cycle was based on the fact that a specific lunar day coincides schematically with the same civil date every 25 civil years. Furthermore, the cycle comprises nine "great" years of 13 lunar months each and 16 "small" years of 12 months, and thus corresponds in its structure to the presumed civil-based lunar calendar. There are civil-lunar double dates which are correct only if their lunar components were

[25] U. Luft, *Die chronologische Fixierung des ägyptischen Mittleren Reiches nach dem Tempelarchiv von Illahun* (Vienna: Österreichische Akademie der Wissenschaften, 1992), pp. 189–97; C. Bennett, "Egyptian Lunar Dates and Temple Service Months," *Bibliotheca Orientalis* 65 (2008), 525–54.

[26] L. Borchardt, *Die Mittel zur zeitlichen Festlegung von Punkten der ägyptischen Geschichte und ihre Anwendung* (Cairo: Selbstverlag, 1935), pp. 5–10, 24.

[27] Parker, *Calendars*, pp. 30–50.

[28] Ibid., pp. 24–29; L. Depuydt, *Civil Calendar and Lunar Calendar in Ancient Egypt* (Leuven: Peeters, 1997), pp. 137–215.

[29] A. J. Spalinger, "Ancient Egyptian Calendars: How Many Were There?," *Journal of the American Research Center in Egypt* 39 (2002), 241–50.

[30] Parker, *Calendars*, pp. 13–29.

determined by the cycle rather than by observation, indicating that the cycle and with it the civil-based lunar calendar were in actual use.[31] A double date of year 12 of Amasis or 559 BCE, referring to a cultic act,[32] points to the civil-based lunar calendar being possibly employed long before the introduction of the Carlsberg Cycle.

For some time after the Greek conquest the Macedonian lunar calendar was in effect alongside the Egyptian calendar. Earlier assumptions that the Carlsberg Cycle was adopted for the Macedonian year cannot be confirmed.[33]

HOURS OF THE DAY AND NIGHT

According to PT Spells 251 and 320 unspecified stars are related to the night hours. Neugebauer and Parker's interpretation of three types of star clocks known from later times[34] has been criticized and modified.[35] The earliest is the so-called diagonal star clock in the form of tables that are found on the lids of non-royal coffins of the Middle Kingdom (2100/2000 BCE), and also 900 years later in the Osireion temple of Seti I in Abydos. They consist of a grid of at least thirty-six columns, corresponding to the decades of a year. Read from top to bottom, a column lists twelve stars whose consecutive risings indicate twelve night hours. The designation decan or decanal star refers to an hour-star's function during a decade.

Any given star rises four minutes earlier with each passing day; after a decade the hour-stars rise forty minutes earlier, making the ascendance of the first hour-star unobservable. The situation was adjusted in the following decade by using decans two to twelve for hours one to eleven and adding a new decan for the twelfth hour. Since the consecutive adjustments result in a diagonal shift of the hour-stars through the grid, Egyptologists speak of "diagonal star clocks."

The prototype of a transit star clock known from the Osireion and the tomb of Ramesses IV dates to about 1800 BCE. The clock belongs to a text known as the "Book of Nut" or the "Fundamentals of the Course of the Stars,"[36] which was still known a millennium later when it was copied and

31 Krauss, "Babylonian Crescent Observation."

32 D. van Heel, "Abnormal Hieratic and Demotic Texts Collected by the Theban Choachytes in the Reign of Amasis" (PhD thesis, Leiden, 1996), pp. 93–9.

33 A. Jones, "On the Reconstructed Macedonian and Egyptian Lunar Calendars," *Zeitschrift für Papyrologie und Epigraphik* 119 (1997), 157–66.

34 O. Neugebauer and R. A. Parker, *Egyptian Astronomical Texts*, 3 vols. (Providence, RI: Brown University Press, 1960–9), vol. 1, passim.

35 For traditional and recent approaches see S. Symons, "Egyptian Star Clocks," in Ruggles (ed.), *Handbook*, pp. 1495–500.

36 A. von Lieven, *Grundriss des Laufs der Sterne. Das sogenannte Nutbuch*, 2 vols. (Copenhagen: Museum Tusculanum Press, 2007).

provided with commentaries. It describes the cycle of the decanal stars understood as a "year" of 360 days: after 70 days of death in the Duat (Underworld), a decanal star is reborn. It then spends 80 days in the eastern sky before it "works"; it then passes 120 days "working" – telling time by its transit starting with the twelfth hour. When it has finished "working" by marking the first hour, it passes 90 days in the western sky and then again it dies.

The so-called Ramesside star clock, which is found in three royal tombs, is based on a prototype from around 1500 BCE. It consists of twenty-four tables, one for each half-month, listing and depicting on a grid the positions of stars relative to the figure of a kneeling man. There are seven possible positions, the middle one corresponding to the meridian where the respective hour-star transits.

It is generally presumed that star clocks were a means for regulating cultic service at night. The "hours" of the star clocks will have been of uneven length even in one and the same night; the length varied also according to the season.

CLOCKS

In a text from about 1500 BCE the official Amenemhet mentions constructing an outflow water clock for measuring temporal night hours. The oldest preserved example of such a clock dates to ca. 1370 BCE, but the arrangement of scales and months shows that the prototype dates to the time of Amenemhet.[37]

Egyptian outflow water clocks are shaped like a flower pot with scales for each month down the inside to measure the fall of the water level. The scales are of different lengths corresponding to the seasonally determined length of the night; each scale is divided into twelve segments corresponding to temporal hours. The shape of the vessel would allow an almost uniform fall of the water-level, but modern study has determined that the speed of outflow depends much on the size of the outflow pipe and on the viscosity of the water, in turn a function of temperature. These factors will have resulted in an uneven pace of outflow.[38] Inscriptions on Hellenistic outflow clocks assert their use for regulating temple service at day or night when the sky was overcast.

[37] P. Mengoli, "La clessidra di Karnak: l'orologia ad acqua di Amenophis III," *Oriens Antiquus* 28 (1989), 227–71.

[38] L. von Mackensen, "Neue Ergebnisse zur ägyptischen Zeitmessung," *Alte Uhren: Zeitmessgeräte, wissenschaftliche Instrumente und Automaten* 1 (1978), 13–18; B. Cotterell, F. P. Dickson, and J. Kamminga, "Ancient Egyptian Water-Clocks: A Reappraisal," *Journal of Archaeological Science* 13 (1986), 31–50.

"Shadow tables" inscribed on votive cubits imply that "hours" of daylight were measured by the shadow thrown by a cubit, but details are not forthcoming.[39]

A shadow clock is described in the Osireion; there are also extant specimens.[40] Shadow clocks consist of two elements: a horizontal bar about 30cm long and a short vertical bar that throws a shadow – decreasing in the morning and increasing in the afternoon – on the longer bar. The shadow moves between markings to indicate the dividing points between successive hours. The relative distances between the markings do not correspond to trigonometric values of solar altitude, but rather to schematic arithmetical series. The scheme results in "hours" of uneven length, which differed from season to season like temporal hours. According to Neugebauer's interpretation, the Osireion clock represents a twelve-hour division of time between morning and evening dawn, leaving ten hours for the time between sunrise and sunset. Of them 4+4 hours were measurable by the sun's shadow. But the extant examples appear to divide the time between sunrise and sunset into twelve hours.

Since about 1200 BCE vertical-plane sun dials are attested which indicate schematic seasonal hours by the angular shadow of a gnomon.[41]

The "hours" produced by shadow clocks could be used for the timing of activities, but not for gauging other time pieces nor for measuring time in equal intervals.

Equal hours of the same length seem to be attested in the calendar of Papyrus Cairo 86637 (ca. 1300 BCE) in so far as the monthly sum of hours of day and night always adds up to twenty-four hours. The calendar also implies that daylight increases or decreases from month to month regularly by two hours, producing a ratio of 3:1 of the longest daylight to the shortest. These figures are incomprehensible, and the table is usually not accepted at face value. Equal hours are attested on a fragmentary table found at Tanis and dating to about 600 BCE citing hours of the day and night at fifteen-day intervals. The hours are listed as full hours and duodecimal fractions. It is debated whether the table is based on observation or on a scheme.[42]

The twenty-four-hour day represents the only contribution that the ancient Egyptians made to modern methods of timekeeping. In Europe, it caught on for everyday use with the invention of mechanical clocks in medieval times. In antiquity, a day with twenty-four equal (aequinoctial) hours was used only by astronomers like Ptolemy. The modern sexagesimal division of hours and minutes comes from Babylonia.[43]

[39] L. Borchardt, *Altägyptische Zeitmessung* (Berlin: Gruyter, 1920), pp. 27–8.
[40] M. Clagett, *Ancient Egyptian Science*, 3 vols. (Philadelphia, PA: American Philosophical Society, 1989–99), vol. 2, 83–95.
[41] S. Bickel and R. Gautschy, "Eine ramessidische Sonnenuhr im Tal der Könige," *Zeitschrift für ägyptische Sprache und Altertumskunde* 96 (2014), 3–14.
[42] Clagett, *Science*, vol. 2, 98–106.
[43] R. A. Parker, "Ancient Egyptian Astronomy," in Hodson (ed.), *The Place of Astronomy*, pp. 51–65, p. 57.

ASTRONOMICAL REPRESENTATIONS

The Egyptological term "celestial diagram" refers to schematic representations depicted on ceilings of temples and tombs, coffin boards, and water clocks; over eighty examples are known.[44] The earliest is found in a tomb prepared for the official Senenmut at Thebes (ca. 1450 BCE); the latest dates to the Roman Period.

The southern panel of the diagram regularly includes thirty-six decanal stars as a series of personified star-gods, including Orion (*Sah*) and Sirius (*Sothis*), followed by the five planets in the sequence Jupiter, Saturn, Mars, Mercury, and Venus. The three outer planets appear as falcons since they are forms of the falcon god Horus. Mars' name "Red Horus" shows that the Egyptians considered the planet's color distinctive, and his epithet "he-who-travels-backwards," reveals that they noticed the planet's markedly retrograde movement. Mercury is equated with the god Seth in the form of a chimerical animal. Venus is depicted as a heron labeled *Benu* (the phoenix, according to Herodotus), an identification traceable back to CT, paralleling an older identification of Horus with the same planet.

The northern panel incorporates a series of gods who originally represented the tutelary deities of the lunar days. It also includes the constellations of the northern sky but only Mesekhtyu (Big Dipper) is identifiable unequivocally. Traditionally, the identification of the northern constellations and also of the decanal stars has been considered unfeasible,[45] but today's computer-generated star maps allow a new approach.[46]

HEMEROLOGY

The principal function of menologies and hemerologies is to determine the propitiousness or lack of it for an entire month or for each day of the year. The Egyptian practice of selecting certain days as unpropitious spread from Egypt to the Hellenistic-Roman world; the unpropitious days were known in antiquity and medieval Europe as *dies aegyptiaci*.[47]

The earliest known Egyptian menology dates to about 1800 BCE. It singles out certain days of a civil month as good, bad, or half-and-half, and is to be used for each month of a year; the epagomenai are not considered. These last are included in a later menology covering each month of a year.[48]

[44] Neugebauer and Parker, *Egyptian Astronomical Texts*, vol. 3, 6–8.
[45] Ibid., pp. 183–94.
[46] J. Lull and J. A. Belmonte, "Egyptian Constellations," in Ruggles (ed.), *Handbook*, pp. 1477–87.
[47] E. Brunner-Traut, "Mythos im Alltag. Zum Loskalender im alten Ägypten," *Antaios* 12 (1970), 332–56.
[48] P. Vernus, "Omina calendériques et comptabilité d'offrandes sur une tablette hiératique de la XVIII[e] dynastie," *Revue d'Égyptologie* 33 (1981), 89–124.

The hemerology preserved in the Ramesside papyri Cairo 86637 and Sallier IV covers an entire year. It refers, inter alia, to star-god lore, which seems to be based on observations of Venus and Mercury as embodiments of the gods Horus and Seth.[49]

LATE PERIOD ZODIACS AND ASTROLOGY

Under Mesopotamian influence, astrological ideas became important in the Late Period. An early example is a papyrus in Vienna which includes the interpretation of lunar und stellar omens; eclipses and the varying color of the lunar disk and of the stars are listed.[50]

The Babylonian concept of the zodiac was introduced to Egypt after the Greek conquest. About two-dozen zodiacs are known which decorate temple ceilings and coffins. Most date to the Roman Period, or are only slightly older, like the celebrated example from Dendera Temple, now in the Louvre. The iconography of the zodiacal signs clearly derives from Babylonian prototypes. Traditional Egyptian features were incorporated, such as the northern constellations, and the iconography and names of sun, moon, and the planets. Furthermore, the traditional decans, now as stellar gods ruling over human fate, were correlated with 10°-segments of a zodiacal sign.[51] Some peculiar iconographical traits of the decans were preserved in astrology down to the Renaissance.[52]

Fully developed Hellenistic astrology is based on the zodiac and the personal horoscope which is also of Babylonian origin. Hellenistic astrological texts known from Egypt include horoscopes, numerical tables, and manuals on papyri and ostraca. For casting horoscopes manuals were used which included predictions and instructions for the use of the manual itself. The two types of predictions are universal and individual. Universal predictions are represented by the so-called Sothis texts, which foretell the fate of Egypt and neighboring countries from the positions of the planets in zodiacal signs at Sothis' first rise in the morning (heliacal rise) or last rise in the evening (acronychal rise).[53] For the casting of individual horoscopes, at least two different methods were employed, one referring to the sun and moon, the zodiacal signs, and the decans and also mentioning the planets, the other referring to

[49] R. Krauss, "The Eye of Horus and the Planet Venus: Astronomical and Mythological References," in J. M. Steele and A. Imhausen (eds.), *Acts of the Conference Under One Sky: Astronomy and Mathematics in the Ancient World* (Münster: Ugarit-Verlag, 2002), pp. 193–208.

[50] R. A. Parker, *A Vienna Demotic Papyrus on Eclipse and Lunar Omina* (Providence, RI: Brown University Press, 1959).

[51] C. Leitz, *Altägyptische Sternuhren* (Leuven: Peeters, 1995).

[52] W. Gundel, *Dekane und Dekansternbilder* (Glückstadt: Augustin, 1936), pp. 82–225.

[53] B. Bohleke, "In Terms of Fate: A Survey of the Indigenous Egyptian Contribution to Ancient Astrology in Light of Papyrus CtYBR inv. 1132(B)," *Studien zur altägyptischen Kultur* 23 (1996), 11–46, 28–9.

planets and their "houses." Circumstantial evidence implies that astrology was an important preoccupation of the Egyptian priesthood in the Roman period.[54]

The positions of the sun, moon, and five planets in the zodiacal signs or constellations provide the astronomical basis for Hellenistic horoscopes. Planetary and lunar tables (such as the famous Stobart tables of 140 BCE)[55] usually note only the dates of entry for planets and the moon in the signs of the zodiac. The entry times in these sign almanacs were computed, rather than determined by observation. The computational methods derive from Babylonian mathematical astronomy.[56]

There are also tables with no immediately recognizable use for an astrologer, such as a lunar table in Vienna citing computed differences between conjunctions.[57] A papyrus in Berlin describes a series of lunar eclipses between 85 and 74 BCE; the dates are mostly calculated, but some remarks seem to derive from observations.[58]

Whereas traditional Egyptian astronomy was centered on the calendar and measuring the night hours, Egyptian astrology in Hellenistic times represented positional astronomy of the luminaries, and the planets with the great circle of the ecliptic or zodiac as a basis for coordinates.

CONCLUDING REMARKS

Mathematical astronomy did not exist in ancient Egypt. The movements of the celestial bodies were only summarily described. Measurement of angles as practiced by the Babylonians was unknown. Egyptian astronomy was characterized by qualitative schemata; accordingly, Egyptologists' analysis of the sources can only be expected to produce qualitative results or, at best, very imprecise quantitative ones. In a narrow sense, Neugebauer's dictum that Egypt has no place in a history of scientific astronomy is correct. Nevertheless, the achievements of the ancient Egyptians should not be belittled, especially when the calendrical and astronomical insights of the third millennium BCE are reviewed.

The possibility cannot be dismissed that some relevant sources for a comprehensive evaluation of Egyptian astronomical understanding have

[54] A. Winkler, "On the Astrological Papyri from the Tebtunis Temple Library," in G. Widmer and D. Devauchelle (eds.), *Actes du IX^e congrès international des études démotiques Paris 31 août – 3 septembre 2005* (Cairo: IFAO, 2009), pp. 361–75.

[55] O. Neugebauer, "Egyptian Planetary Texts," *Transactions of the American Philosophical Society* N.S., 32.2 (1942), 209–50.

[56] A. Jones, "The Place of Astronomy in Roman Egypt," in Timothy D. Barnes (ed.), *The Sciences in Greco-Roman Society* (Edmonton: Academic Printing & Publishing, 1994), pp. 25–51.

[57] Neugebauer and Parker, *Egyptian Astronomical Texts*, vol. 3, 243–50; Bohleke, "In Terms of Fate."

[58] J. M. Steele, *Observations and Predictions of Eclipse Times by Early Astronomers* (Dordrecht: Kluwer, 2009), pp. 86–91.

been lost. Knowledge of eclipses is a case in point. Classical sources preserve the tradition that the Egyptians regularly observed eclipses; this assertion is not confirmed by data from Egypt. Furthermore, Ptolemy, who lived in Alexandria and evaluated all such data available to him, including Babylonian sources, did not cite a single pharaonic example. The classical tradition could be an invention deriving from an exaggerated respect for the erudition of the ancient Egyptians – not shared by Egyptologists.[59]

In the second half of the nineteenth century the Egyptologist Heinrich Brugsch laid the foundation for the study of ancient Egyptian calendars and astronomy. His successors in the twentieth century were Egyptologists Ludwig Borchardt and Richard A. Parker, and the historian of astronomy, Otto Neugebauer. The standard views of Parker on calendars and those of Neugebauer on star clocks have been criticized, but only partially revised since the 1980s,[60] then defended and doubted again; the discussion is ongoing. The recent scholarly discussion of water and shadow clocks is comparatively objective, since it is based on the construction of actual pieces. By contrast, the available data on calendars and also on star clocks are prone to different interpretations and subject to personal bias, which plays a role in revisionist views. Research on the astronomical content of Pyramid and Coffin Texts,[61] and into the sources for astrology in the Hellenistic period, seems to be far less prejudiced.

59 G. J. Toomer, "Mathematics and Astronomy," in J. R. Harris (ed.), *The Legacy of Egypt* (Oxford: Clarendon Press, 1974), pp. 27–54, p. 52.

60 Clagett, *Science*, vol. 2, 22.

61 P. Wallin, *Celestial Cycles: Astronomical Concepts of Regeneration in the Ancient Egyptian Coffin Texts* (Uppsala: Uppsala University Press, 2002). K. Goebs, *Crowns in Egyptian Funerary Literature. Royalty, Rebirth and Destruction* (Oxford: Griffith Institute, 2008); the book deals at length with Horus as Morning Star in PT and CT.

8

EGYPTIAN MATHEMATICS

Jens Høyrup

Ivor Grattan-Guinness in memoriam

The primary references of the term "Egyptian mathematics" are the computational techniques and the underlying mathematical knowledge attested in Pharaonic written sources. Secondary references are, on one hand, the corresponding techniques etc. as known from demotic sources and, on the other, the geometrical procedures used in Pharaonic and subsequent architecture and visual arts. Greek mathematics produced in Hellenistic Egypt is thus *not* included. Accordingly, all dates in the following are BCE when not specified to be CE.

THE SOURCES

The most important written sources for Pharaonic mathematics are the *Rhind Mathematical Papyrus* (henceforth RMP)[1] and the *Moscow Mathematical*

[1] Two editions with ample analysis exist: T. Eric Peet, *The Rhind Mathematical Papyrus, British Museum 10057 and 10058. Introduction, Transcription, Translation and Commentary* (London: University Press of Liverpool, 1923); Arnold Buffum Chace, *The Rhind Mathematical Papyrus. British Museum 10057 and 10058*, 2 vols. (Oberlin, OH: Mathematical Association of America, 1927–9), vol. 1 (with the assistance of Henry Parker Manning): *Free Translation and Commentary and a Bibliography of Egyptian Mathematics by R. C. Archibald*; vol. 2 (with Ludlow Bull and Henry Parker Manning): *Photographs, Transcription, Transliteration, Literal Translation and a Bibliography of Egyptian and Babylonian Mathematics (Supplement), by R. C. Archibald and the Mathematical Leather Roll in the British Museum, by S. R. K. Glanville* (the literal translation is indeed *very* literal; it is used in all quotations below from RMP). A recent facsimile edition with description and discussion and translation of large extracts is Gay Robins and Charles Shute, *The Rhind Mathematical Papyrus: An Ancient Egyptian Text* (London: British Museum Publications, 1987).Valuable general accounts of Egyptian mathematics which (by necessity) deal extensively with RMP are: Kurt Vogel, *Vorgriechische Mathematik.* Volume 1*: Vorgeschichte und Ägypten* (Mathematische Studienhefte, 1; Hanover: Hermann Schroedel / Paderborn: Ferdinand Schöningh, 1958); Richard J. Gillings, *Mathematics in the Time of the Pharaohs* (Cambridge, MA: MIT Press, 1972). After the manuscript for the present article was finished (essentially in 1997), Marshall Clagett has published *Ancient Egyptian Science. A Source Book.* Volume 3*: Ancient Egyptian Mathematics* (Memoirs of the American Philosophical Society, 232; Philadelphia, PA: American Philosophical Society, 1999). This volume reproduces the facsimile edition of Chace's volume 2 and includes a translation kept close to that of the same volume as well as an extensive analysis. Extensive analysis of RMP as well as other Middle Kingdom sources is also found in Annette Imhausen, *Ägyptische Algorithmen. Eine Untersuchung zu den mitteläggyptischen mathematischen Aufgabentexten* (Ägyptologische Abhandlungen, 65; Wiesbaden: Harrassowitz, 2003). Egyptian mathematics in its social setting is dealt

Papyrus (MMP).[2] To these can be added a few shorter papyri containing mathematical problems; a couple of listings of equivalent fractions; and a larger number of accounting papyri that apply the metrology and show how the basic arithmetical techniques were used. The RMP is a copy from a Middle Kingdom original (Amenemhet III, c. 1800), made during the Hyksos period; it is a teacher's or calculator's manual, containing several tables (to which we shall return) and some eighty problems with solution. The MMP seems to be a late Middle Kingdom copy from an earlier Middle Kingdom original; it is a collection of student's answers to problems, provided with the teacher's approval (at times justly refused). The other properly mathematical sources are from the Middle through New Kingdom.[3]

A source of particular character is the New Kingdom fictional "satirical letter" or Papyrus Anastasi I,[4] in which a scribe chides a colleague for his professional ignorance; it shows that a military scribe was supposed to be familiar with Palestinian geography and with the determination of the manpower, rations, and other requirements of construction work. Accounting papyri show that other categories of scribes had analogous tasks.

Some administrative records go back to the Old Kingdom (and a few documents containing numbers to the early dynastic period); even in this respect, however, the Middle through New Kingdom is much richer.

The mathematical sources of the Pharaonic period are written in hieratic script (evidently, some hieroglyphic documents contain numbers); part of the metrological terminology seems to have been created in hieratic and to have acquired hieroglyphic equivalents only at a later moment (when at all).

with in Annette Imhausen, *Mathematics in Ancient Egypt: A Contextual History* (Princeton & Oxford: Princeton University Press, 2016).

2 Edition with translation and extensive commentary: W. W. Struve, *Mathematischer Papyrus des Staatlichen Museums der Schönen Künste in Moskau* (Quellen und Studien zur Geschichte der Mathematik. Abteilung A: Quellen, 1. Band; Berlin: Julius Springer, 1930). The edition is reproduced in Marshall Clagett, *Ancient Egyptian Science*, vol. 3, which also contains an English translation and a commentary.

3 Marshall Clagett, following recent research, gives the following approximate dates (*Ancient Egyptian Science. A Source Book.* Volume I*: Knowledge and Order* (Memoirs of the American Philosophical Society, 184 A+B; Philadelphia, PA: American Philosophical Society, 1989), pp. 629–35):

Early dynastic period (dynasties 1–2): 3110–2665.
Old Kingdom (dynasties 3–8): 2664–2155.
First intermediate period (dynasties 9–10): 2154–2052.
Middle Kingdom (dynasties 11–13): 2040–1640.
Second intermediate period (Hyksos dynasties 15–16, Theban dynasty 17): 1640–1532.
New Kingdom (dynasties 18–20): 1550–1070.
Third intermediate period (dynasties 21–(initial) 25): 1070–712.
Late period (dynasties (final) 25–31, including the Assyrian hegemony during dynasty 26 and the Persian dynasties 27 and 31): 712–332.
Greco-Roman period: 332 BCE to 395 CE.

4 Full edition and translation in Alan H. Gardiner, *Egyptian Hieratic Texts.* Series I*: Literary Texts from the New Kingdom.* Part I*: The Papyrus Anastasi I and the Papyrus Koller, together with Parallel Texts* (Leipzig: J. C. Hinrichs'sche Buchhandlung, 1911).

From the demotic phase (late and Greco-Roman periods), several papyri containing mathematical problems and tables survive[5] – with the possible exception of one uncertain fragment all of Greco-Roman date.

NUMBERS AND METROLOGY

The Egyptian number system was decadic. Already in early dynastic times, individual signs for 1, 10, 100, 1,000, 10,000, 100,000, and 1,000,000 existed.[6] In hieroglyphic, other numbers were constructed additively, by mere repetition of these signs; in hieratic, these were contracted into individual signs for 1, 2, ..., 9, 10, 20, ..., 90, 100, 200, etc. From the Middle Kingdom onward, fractional numbers were expressed as sums of aliquot parts (including two-thirds); in order to keep close to the Egyptian notation we may transcribe them as follows: $3''$ $(={}^2/_3)$, $2'$ $(= {}^1/_2)$, $3'$ $(= {}^1/_3)$, $\overline{4}$ $(= {}^1/_4)$, $\overline{5}$, etc. ($3''$, $2'$, and $3'$ had special signs); the others were denoted by the sign *ro* ("mouth," here "part") or by a dot above the number (in hieroglyphic respectively hieratic writing), according to a canon that did not allow repetition of the same aliquot part but expressed for instance $\overline{5}\ \overline{5}$ as $3'$ $1\overline{5}$ (juxtaposition means addition). Essential metrologies, however, would operate with subunits instead of these fractions.

Many of the problems in the mathematical texts deal with the difficulties to which the non-decadic metrologies would give rise. Closest to decadic principles is the length system. The basic length unit was the "royal cubit" (*mḥ*, c. 52cm), subdivided into 7 "palms" of 4 "fingers" each (a "short cubit" of 6 palms was also in use). 100 royal cubits was a "rope" (*khet*). Land might be measured in *setat* (that is, square *khet*), divided by successive halvings into subunits with special names (down to ${}^1/_{32}$ *setat*); in surveying practice, the "cubit of land" (1 cubit versus 1 *khet*) and the "thousand of land" (1,000 cubit versus 1 *khet*) were mostly preferred.

The central capacity unit was the *ḥeḳat*, divided according to one system into 10 *henu* or 320 *ro* (the "part" again, but in a different use), according to another by successive halvings (down to ${}^1/_{64}$).[7] Multiples of the *ḥeḳat* were expressed by special signs or by non-standard use of the standard numerals. A special unit for bulky substances is the *khar*, equal to 20 *ḥeḳat* and to ${}^2/_3$ of a cubic royal cubit (probably a secondary normalization of originally distinct units).

[5] Edition with translation and extensive commentary in Richard A. Parker (ed.), *Demotic Mathematical Papyri* (Providence, RI and London: Brown University Press, 1972).

[6] The standard reference for Egyptian numerals and number words remains Kurt Sethe, *Von Zahlen und Zahlworten bei den Alten Ägyptern, und was für andere Völker und Sprachen daraus zu lernen ist* (Schriften der Wissenschaftlichen Gesellschaft in Straßburg, 25; Straßburg: Karl J. Trübner, 1916). In the second millennium, the sign for 1,000,000, and afterwards that for 100,000, went out of use; instead, multiplicative notations were used.

[7] The hieroglyphic signs for the successive halves of the *ḥeḳat* can be put together to the standard representation of the healing sacred eye of Horus, as pointed out by Peet, *The Rhind Mathematical Papyrus*, p. 26. However, the hieroglyphic writings do not antedate the Eighteenth Dynasty, whereas the hieratic forms go back to the third millennium; the mythological connotations of the system are

BASIC PATTERNS AND TECHNIQUES

Egyptian arithmetical thinking may be interpreted as based on two key principles: additivity and proportionality – the latter in the sense that any number might count another number; to this come the techniques of doubling and multiplying by 10. The multiplication of 75 by 53 might be performed thus:

/1	75
/2	150
/10	750
20	1500
/40	3000
Total	3975

Some texts reveal the underlying thought: If *1* (of the entity we count) is 75, then *2* (of it) is 150, etc. The multiplier 53, as we see, is split into components that can be obtained by successive doublings and decouplings (mostly, only doublings would be employed). Strokes mark addends that are actually used (53 = 1+2+10+40).

The corresponding division of 3975 by 75 would go by the same procedure, emptying 3975 by multiples of 75:

/1	75
/10	750
20	1500
/40	3000
/2	150
Total	3975

A separate phrase would state the result as 53 (= 1+10+40+2); strokes will of course have been inserted a posteriori in the scheme.

This remains simple only until fractions are introduced. An actual multiplication (of 8 3″ $\overline{6}$ 1$\overline{8}$ = $8^{8}/_{9}$ by itself) would run as follows (RMP 42):

1	8 3″ $\bar{6}$ 1$\bar{8}$
2	17 3″ $\bar{9}$
4	35 2′ 1$\bar{8}$
/8	71 $\bar{9}$
/3″	5 3″ $\bar{6}$ 1$\bar{8}$ 2$\bar{7}$
3′	2 3″ $\bar{6}$ 1$\bar{2}$ 3$\bar{6}$ 5$\bar{4}$
/$\bar{6}$	1 3′ 1$\bar{2}$ 2$\bar{4}$ 7$\bar{2}$ 10$\bar{8}$
/1$\bar{8}$	3′ $\bar{9}$ 2$\bar{7}$ 10$\bar{8}$ 32$\bar{4}$
Total	79 10$\bar{8}$ 32$\bar{4}$

It is no accident that 3″ of 8 3″ $\overline{6}$ $1\overline{8}$ is found before 3′. Even when only 3′ of the multiplicand is needed, 3″ is found first and 3′ then by halving. 3″ and 2′ were the basic fractions of the Middle Kingdom calculators; if at all possible, further divisions would be produced from these by successive halvings (the presence of $1\overline{8}$ illustrates that it was not always possible).

Beyond this, the calculation displays the main difficulties to which multiplication of fractions gives rise. The first doubling is obvious, since 3″ doubled is 1 3′; in the next, however, $\overline{9}$ has to be doubled, and the scribe has to know that this yields $\overline{6}$ $1\overline{8}$ (after which 3′ $\overline{6}$ is contracted to 2′). Finally, $\overline{9}$ 3″ $\overline{6}$ $1\overline{8}$ $2\overline{7}$ 3′ $1\overline{2}$ $2\overline{4}$ $7\overline{2}$ $10\overline{8}$ 3’ $\overline{9}$ $2\overline{7}$ $10\overline{8}$ $32\overline{4}$ has to be converted into 2 $10\overline{8}$ $32\overline{4}$.

For the former purpose, RMP contains a tabulation of 2÷*n*, for all odd values of *n* from 5 to 101. For the latter, a technique referred to as “red auxiliary numbers” was used. The fractions might be expressed as fractions of an adequate “reference magnitude” – in the present case probably 108 – in a scheme (red is rendered by italics):

$\bar{9}$	3″	$\bar{6}$	$1\bar{8}$	$2\bar{7}$	3′	$1\bar{2}$	$2\bar{4}$	$7\bar{2}$	$10\bar{8}$	3′	$\bar{9}$	$2\bar{7}$	$10\bar{8}$	$32\bar{4}$
12	*72*	*18*	*6*	*4*	*36*	*9*	*4 $\bar{2}$*	*1 $\bar{2}$*	*1*	*36*	*12*	*4*	*1*	*3′*

Since the sum of the red (i.e., italicized) numbers is 217 3′= 2·108+1+3′, the sum of the fractions is 2+$10\overline{8}$ +$32\overline{4}$. Structurally, this is equivalent to the use of a common denominator 108, and there are some hints that a notion of the fraction $^p/_q$ understood as *p* copies of $\overline{q}$ was not as strange to Egyptian calculators as the stylistic canon might make us believe – in RMP 81, the scribe erroneously writes $\overline{5}$ and 3 instead of $\overline{2}$ $\overline{8}$ (= $^5/_8$) and $\overline{4}$ $\overline{8}$ (= $^3/_8$).[8] Nonetheless, an interpretation of the underlying thought in terms of a reference magnitude[9] agrees so well with the global pattern of the texts that it is likely to be the primary explanation of the red auxiliaries.

/1	7
/$\overline{7}$	1
1	8
/2	16
2′	4
/$\overline{4}$	2
/$\overline{8}$	1
/1	2 $\overline{4}$ $\overline{8}$
/2	4 2′ $\overline{4}$
/4	9 2′

8 This was first pointed out by Kurt Vogel in *Die Grundlagen der ägyptischen Arithmetik* (first published Munich, 1929; reprint Wiesbaden: Martin Sändig, 1970), p. 43.

9 First proposed in Léon Rodet, “Les prétendus problèmes d’algèbre du manuel du calculateur égyptien (Papyrus Rhind),” *Journal asiatique*, septième série 18 (1881), 184–232, 390–559.

Most everyday practical computation above the level of counting is based on proportionality in one or the other way – since the Middle Ages often in the shape of the Rule of Three. The Egyptian approach may be illustrated by RMP 24, one of the problems treating of an abstract "quantity" or "heap" ${}^{c}\d{h}^{c}$ – the problem type by which the technique was trained: "A quantity, $\overline{7}$ of it added to it, becomes it: 19."

The computation looks as follows

/1	7	The doing as it occurs.	
/$\bar{7}$	1	The quantity	16 2′ $\bar{8}$
		$\bar{7}$	2 $\bar{4}$ $\bar{8}$
1	8	Total	19
/2	16		
2′	4		
/$\bar{4}$	2		
/$\bar{8}$	1		
/1	2 $\bar{4}$ $\bar{8}$		
/2	4 2′ $\bar{4}$		
/4	9 2′		

This may be explained as a "single false position": as a preliminary value for the heap we take 7; then the quantity together with its seventh part becomes 8. This is seen to be contained 2 $\overline{4}$ $\overline{8}$ times in 19 (an ordinary division); therefore the true value of the quantity is 2 $\overline{4}$ $\overline{8}$ times 7 – or, which is more convenient for the final proof, 7 times 2 $\overline{4}$ $\overline{8}$ = 16 2′ $\overline{8}$. The Egyptians, indeed, made ample use of the commutativity of multiplication, despite the obvious asymmetry of their algorithm; the frequent claim that their mathematical thought was purely additive is thus blatantly mistaken.

The principles of this computation were applied with flexibility: at times the preliminary value might be set to *1* (e.g. RMP 32); in combination with the commutativity of operations this might lead to something very close to the Babylonian division through multiplication by the reciprocal (e.g. RMP 63). The formulations, however, show that the Egyptian method is based on the usual principles and no borrowing from abroad.

The $2 \div n$ table of RMP is the largest extant piece of systematic Egyptian mathematics and may be considered its theoretical high point. Much effort has hence been dedicated to finding the principle(s) which underlie its construction – the same fraction may indeed be split in many different ways into aliquot parts ($^{2}/_{15}$ thus into $\overline{8}$ $\overline{120}$, $\overline{9}$ $\overline{45}$,

$1\overline{0}$ $3\overline{0}$, $1\overline{2}$ $2\overline{0}$, $1\overline{1}$ $3\overline{0}$ $11\overline{0}$, $1\overline{3}$ $2\overline{0}$ $15\overline{6}$, $1\overline{4}$ $3\overline{0}$ $3\overline{5}$, etc.). So much is certain that a standard existed in the later Middle Kingdom – the deviations from the RMP-norm are rare enough to count as aberrations. Kurt Vogel points to three principles (at times in mutual conflict) that seem to intervene:[10]

(i) The members of the sum should be few.
(ii) The first member should be as large as possible.
(iii) If more than two members are present, the largest denominator should be kept small.

(ii) might seem to suggest a search for a good first approximation – but (iii) shows that a good second approximation was not aimed at. The principles seem rather to have been of an aesthetic kind.

The technique that is used consists in dividing 2 into two parts p+r, where p is an aliquot part of n ($^{p}/_{n} = {}^{1}/_{m}$) and the remainder r is an aliquot part of 1 or the sum of such parts, $r = {}^{1}/_{s}+{}^{1}/_{t}+ \ldots$, whence $^{r}/_{n} = {}^{1}/_{(sn)}+{}^{1}/_{(tn)}+ \ldots$[11] This much is shown explicitly in the text, which lists p and the constituents of r and tells which part ($^{1}/_{m}$, $^{1}/_{(sn)}$, etc.) each one is of n. The essential trick, however, is of course to find an adequate splitting of 2. Here several ways were followed, perhaps reflecting the steps of the historical process that had engendered the table. If n is a multiple of 3 ($n = 3\ k$), the division is into 1 2′ and 2′, whence $2 \div n = {}^{1}/_{2\ k}+{}^{1}/_{6\ k}$. In many other cases, an adequate p was probably found by subdivision of 3" of n or 2' of n, as illustrated by the way the text explains 2÷13:

	1	13	*$\bar{8}$* 1 2′ $\bar{8}$	*5$\bar{2}$* $\bar{4}$	*10$\bar{4}$* $\bar{8}$
	2′	6 2′			
	$\bar{4}$	3 $\bar{4}$			
	/$\bar{8}$	1 2′ $\bar{8}$			
/4	5$\bar{2}$	$\bar{4}$			
/8	10$\bar{8}$	$\bar{8}$			

At first, 13 is subdivided by successive halvings until we get below 2; then 1 2′ $\overline{4}$ (= $\overline{8}$ of 13) is chosen as p, and the remainder is seen to consist of $\overline{4}$ (= $\overline{s}$) and $\overline{8}$ (= $\overline{t}$). n = 13 (considered as a "weak sign," i.e., as the representative of an aliquot

10 Vogel, *Vorgriechische Mathematik*, vol. 1, 42.
11 In two cases, n = 35 and n = 91, r/n turns out to be an aliquot part even though r iself is composite (2÷35 = $3\overline{0}$ +$4\overline{2}$, 2÷91 = $7\overline{0}$ +$13\overline{0}$).

part[12]) is then multiplied by 4 and 8, and we see that the *numbers* $\overline{4}$ and $\overline{8}$ are 5$\overline{2}$ and 10$\overline{4}$ *of 13*. The summary in the right column tells that 2 is $\overline{8}$ 5$\overline{2}$ 10$\overline{4}$ of 13.

In other cases *m* is stated directly, often as one of the abundant numbers 30 and 60. It cannot be excluded that these choices resulted from mere trial and error – values of *m* with a profusion of divisors are most likely to permit a nice splitting of the remainder *r* – but it seems more plausible that the Egyptians had discovered that 30 and 60 are often convenient choices and took this as their first guess;[13] however that may be, the procedure makes use of a reference magnitude or of splitting into smaller parts. We may look at 2÷73:

	73	$\overline{60}$ 1 $\overline{6}$ $\overline{20}$	$\overline{219}$ 3′	$\overline{292}$ $\overline{4}$	$\overline{365}$ $\overline{5}$
Find	\\$\overline{60}$	1 $\overline{6}$ $\overline{20}$			
\3	$\overline{219}$	3′			
\4	$\overline{292}$	$\overline{4}$			
\5	$\overline{365}$	$\overline{5}$			

This can be understood as follows (structurally equivalent interpretations are possible): 2 is split into 120 parts, each of which is then $\overline{60}$. 73 of these divided by 73 make $\overline{60}$; since 73 = 60+10+3, 73÷60 = 1 $\overline{6}$ $\overline{20}$ = *p*. The right-hand column tells that the remainder until 2 after the removal of *p* is 3′ $\overline{4}$ $\overline{5}$ (namely 47 of the 120 small parts of 2, grouped as 20+15+12; possible alternatives are 30+15+2 and 30+12+5 – the actual choice illustrates Vogel's rule (iii)); multiplication of 73 (considered "weak") by 3, 4, and 5 (left-hand column), respectively, shows that the remainder is $\overline{219}$ $\overline{292}$ $\overline{365}$ of 73 (middle column); all is summarized in the first line.

APPLIED ARITHMETIC

Beyond the abstract *ᶜḥᶜ*-problems, both RMP and MMP contain many arithmetical problems of practical or sham-practical character. Most important are distribution problems and the so-called *pesu*-problems.

Many of the distribution problems deal with equal partition – e.g., the distribution of *n* loaves among 10 persons, *n* = 1, 2, 6, 7, 8, 9 (RMP, 1–6); they illustrate why Plutarch and other Greek authors would link social equality to "arithmetical justice" (and hence reject the latter as morally unsound). Others follow the principle that the foreman and other officials get double share (RMP 65), or that the ratio between shares is given (RMP 63). Such problems are true to real life as revealed in administrative texts. RMP 40, on

12 This terminology is used in RMP 61B; cf. Peet, *The Rhind Mathematical Papyrus*, p. 104.

13 The repeated occurrence of 30 allows us to discard the hypothesis that the frequent choice of 60 was inspired by the Babylonian sexagesimal system.

the contrary, is wholly artificial: loaves are distributed in five shares (say, *a, b, c, d*, and *e*) in arithmetical proportion in such a way that

(i) $^{1}/_{7}$ of the sum of the three major shares equals the sum of the two minor ones;
(ii) $a+b+c+d+e = 100$.

The solution makes use of a simple false position: at first an arithmetical progression α, β, γ, δ, ε is constructed, starting with α = 1 and fulfilling (i); its sum is found to be 60, whence all members are multiplied by 1 3″. The first step is not explained, but since RMP 64 refers explicitly to and makes adequate use of the *average share* and the *excess* of one share over the other when determining the single members of an arithmetical progression from the sum and the difference, a simple algebraic solution (whether represented by words, by strokes on an ostrakon or by pebbles or other material tokens) will not have exceeded the conceptual capabilities of the Egyptian calculator, though apparently his standard discourse: if α is 1 and τ the difference, β = 1+τ, γ = 1+2τ, etc.; the sum of the three major shares is thus 3+(2+3+4)τ, which is 7 times α+β = 2+τ; thus 3+9τ = 14+7τ, 2τ = 11, τ = 5 2′.

Endowed with particular status – obviously because of the importance of bread and beer as staple food – are the *pesu* (*pfśw*) problems. *Pfśw* is derived from *pśj*, "cooking," and may be understood as "baking ratio." The *pesu* of a loaf is the number of similar loaves that may be made from one *ḥeḳat* of grain; similarly, the *pesu* of beer counts the number of jugs that are produced from one *ḥeḳat* of grain. In both cases, the baking ratio thus indicates the reciprocal grain content of the unit of consumption. *Pesu* problems may ask for the *pesu* given the number of units produced and the total amount of grain, or for the exchange of loaves with different *pesu* or of bread with beer. More complex problems deal with the dilution of beer, or with special brews made from several grain sorts or fruits.[14]

Together, these and other problems of applied arithmetic cover more or less the standard types of late medieval and early modern commercial arithmetic – proportional partition, exchange, alloying (only composite interest has no Egyptian counterpart); often the methods are familiar, although no technique similar to the double false position is ever applied; at times, however, unexpected steps demonstrate that ad hoc reasoning was no less important than automatic routines.

Among higher arithmetical problems, one recreational problem in RMP deals with the geometrical progression $7, 7^2, 7^3, 7^4, 7^5$, and finds the sum as 7·2801; nothing in the text tells whether the underlying reasoning is simply that $7+ \ldots +7^5 = 7\cdot(1+ \ldots +7^4) = 7\cdot 2801$, or a formula for the sum of a geometrical progression was known.[15]

[14] See Peet, *The Rhind Mathematical Papyrus*, pp. 112–21, and Struve, *Mathematischer Papyrus*, pp. 44–101.
[15] A third, somehow intermediate possibility, is suggested by Robins and Shute, *The Rhind Mathematical Papyrus*, pp. 56 f.

GEOMETRICAL COMPUTATION

Geometrical problems deal with slopes, areas, and volumes. The batter (*śd*) of pyramids is expressed as the retrocession in palms per cubit height, whereas that of a different (unidentified) structure is given as a pure-number ratio in RMP 60.

Already the metrology (cf. above section, “Numbers and Metrology”) shows that rectangular areas were found as length times breadth. The area of a triangle was determined as half the base multiplied by “the edge,” whose identity has been discussed; however, since RMP 51 takes the half of the base “for the giving of the rectangle of it,” there can be little doubt that the edge between the two parts into which an isosceles triangle is cut is meant – that is, the height.[16] The area of the trapezium was found correspondingly.

Area computation serves in a few cases as the basis for homogeneous second-degree problems. Thus in MMP 6, 7, and 17, the area of a right triangle and the ratio between the sides is given; doubling of the area and multiplication by the ratio yield the area of a corresponding square, whose square root (“corner”) is then one side of the triangle; similar considerations are used to solve problems about two squares whose sides have a given ratio.

The volume of a right parallelepiped was found by multiplication of the three dimensions measured in cubits, followed by a multiplication by 1 2′ in order to express it in *khar*. MMP 14 finds the volume of a truncated square pyramid (with height h and sides a and b of base and top, respectively) correctly as $^{h}/_{3}\cdot(a^2+ab+b^2)$. No cues are given as to how the formula was derived. It cannot be excluded that it is the result of a lucky generalization of the formula for the area of a triangle; nor is a heuristic argument based on dissection into simpler volumes to be excluded, however.[17]

The area of the circle was found as that of the square on $^{8}/_{9}$ of the diameter – 1.006 . . . times the true value. A diagram in RMP 48 suggests that this may be a computational approximation to the area of a geometrically approximating octagon, whose area is $^{63}/_{81}$ of the square in question (see Figure 8.1.). Volumes of circular cylinders were determined accordingly.

MMP 10 calculates the surface of a “basket” with “mouth” 4 2′ as $4^{1}/_{2}\cdot(^{8}/_{9}\cdot[^{8}/_{9}\cdot 9])$, with the argument that the “basket” is the half of an “egg” (Struve’s reading of a damaged word). The double factor $^{8}/_{9}$ leaves no doubt that explicit use is made of the formula for the circular area – no empirical measurement would be able to distinguish $(^{8}/_{9}\cdot[^{8}/_{9}\cdot 9]) = 7\ \ \overline{9}$ from 7 – and the conjectured “egg” seems to suggest that a hemisphere with diameter 4 2′ is intended, whose surface (in modern

16 This was already argued by Peet, *The Rhind Mathematical Papyrus*, pp. 91–3.
17 See, e.g., Robins and Shute, *The Rhind Mathematical Papyrus*, p. 49.

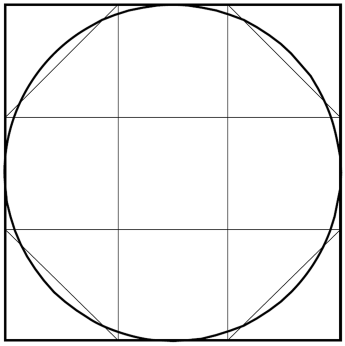

Figure 8.1. Method for finding the area of a circle.

terms) is then found correctly as $2\pi r^2$. This formula seems much more sophisticated that anything else found in the sources, for which reason the alternative interpretation of the "basket" as the curved surface of a semicylinder (with height = diameter = 4 2′) has been suggested. This does not fit an "egg" too well, but has the advantage to presuppose only that the Egyptians knew the relation between circular area and circumference – which agrees well with their explicit transformation of a triangle into a corresponding rectangle.[18]

GEOMETRICAL TECHNIQUES

Rules for geometrical computation evidently depend on techniques for mensuration. These will have been the responsibility of those "rope stretchers" (*harpedonaptai*) which the Greeks refer to.[19] Rope constructions were also used when the ground plans of prestige buildings were laid out. Architectural designs as well as pictorial art were constructed within square grids, following a strict canon (the "canonical system," coupled to metrology and already used in Early Dynastic iconography) for how many grid parts each part of a human body should occupy in the picture.[20]

[18] The two interpretations (due, respectively, to Struve and Peet) are confronted in grammatical detail in O. Neugebauer, *Vorgriechische Mathematik* (Berlin: Julius Springer, 1934), pp. 129–37.

[19] See Peet, *The Rhind Mathematical Papyrus*, p. 32, and Vogel, *Vorgriechische Mathematik*, vol. 1, 59 f.

[20] See Erik Iversen, *Canon and Proportion in Egyptian Art* (2nd edn; Warminster: Aris & Phillips, 1975) (first published 1955) and (on the influence of the system in later art) Erik Iversen, "The Canonical Tradition," in J. R. Harris (ed.), *The Legacy of Egypt* (2nd edn; Oxford: Oxford University Press, 1971), pp. 55–82. Corinna Rossi's *Architecture and Mathematics in Ancient Egypt* (Cambridge: Cambridge University Press, 2003) is a recommendable (and quite cautious) investigation of what can be said in general about the intertwinement of the two fields.

FABLES

Two remarks should be added concerning what we have *no* reason to ascribe to Pharaonic Egyptian geometry.

Firstly, ever since Moritz Cantor proposed that the rope stretchers *might* have used the 3-4-5-triangle to construct right triangles it has been a recurrent claim that they actually did so.[21] It has also been presumed that since several pyramids have the batter 3:4, the Egyptians will have known the properties of this triangle.[22] It must be emphasized that the sources do not contain the slightest hint pointing in this direction, and that the batter in question would be expressed as 5 $\overline{4}$ palms [per cubit height], which is not liable to have furthered any "Pythagorean" speculations.

Similarly, the attempts to find π (or the Golden Section) in the great pyramids founder on the observation that the approximate occurrence of such ratios in the construction are automatic consequences of the simple value for the batter; we should also remember that the Egyptians did not make use of π, that is, of the ratio between circular circumference and diameter, but of $\sqrt{\pi/4}$ (the ratio between the side of the squared circle and the diameter), which they approximated as $^8/_9$.[23]

ORIGINS AND DEVELOPMENT

The canonical system of proportions goes back to Dynasty 1; recordings of the Nile height in cubits, palms, and fingers (which probably served to fix the taxation level for the year) are roughly contemporary; biennial "countings" of the resources of the country begin with Dynasty 2; "chord stretching" at the foundation of prestige buildings is recorded during Dynasty 1.[24] Though nothing comparable to the bureaucratic precision of early Mesopotamian state formation was aimed at – the legitimization of the Pharaonic state rested on conquest and perhaps on the affirmation of cosmological stability, not on redistribution – practical mathematics proper was thus certainly present throughout the third millennium.

This, however, is not yet the mathematics of RMP and MMP. Until Dynasty 5 only metrological sub-units and the fractions 3″, 2′, and 3′ (and

[21] Thus, e.g., Alexander Badawy, *Ancient Egyptian Architectural Design. A Study of the Harmonic System* (Berkeley and Los Angeles, CA: University of California Press, 1965), pp. 3f. and passim.

[22] See Gay Robins and Charles C. D. Shute, "Mathematical Bases of Ancient Egyptian Architecture and Graphic Art," *Historia Mathematica* 12 (1985), 107–22.

[23] For further references regarding the fables and their lack of foundation, see Gillings, *Mathematics*, pp. 237–9. A recent, very careful treatment is Roger Herz-Fischler, *The Shape of the Great Pyramid* (Waterloo, Ontario: Wilfrid Laurier University Press, 2000).

[24] See the translation of the Palermo Stone annals in Clagett, *Ancient Egyptian Science*, vol. 1, 67–95.

a particular sign × for $^1/_4$) were in use – cf. the traditional way the batter of pyramids is expressed in RMP, in contrast to that of a structure with no links to the Old Kingdom. The first place where *ro*-writings of $\overline{4}$, $\overline{5}$, and $\overline{6}$ turn up is the twenty-fourth-century Abū Sir papyri – but since $\overline{5}$ appears in the sum $\overline{5}\ \overline{5}$ (for $^2/_5$) it is clear that the later canon for using the aliquot parts did not yet exist.[25] In Middle Kingdom administrative papyri, on the other hand, even aliquot parts too small to be noticed in practice are used routinely ($\overline{180}$ of a jug of beer! – in contrast, both 14÷10 and 16÷10 may quietly be equalled to 1 2' in economical papyri[26]). Since the mathematics of MMP and RMP is shaped as a coherent structure precisely by the use of aliquot parts and by the techniques for dealing with them, *Ancient Egyptian mathematics* seems to be a creation of the early Middle Kingdom, and to have been made immediately the fundament for the mathematical training of scribes.

Conversely, a new organization of scribal training may have been the motive force that transformed a bundle of mathematical techniques into a unified whole.[27] Old Kingdom scribes had been taught in an apprenticeship system, and not in a school; the first reference to a school postdates the collapse of the Old Kingdom. Early Middle Kingdom scribes, on the other hand, had gone to school. For purposes of practical computation, metrological sub-units are much more convenient than aliquot parts and their sums (just as decimal fractions are more convenient than ordinary fractions). The advantages of the aliquot parts will only stand out as such within a school context that has gained some autonomy from immediate practice:

- they permit exactness (and thus the teacher's decision whether "you have found it correctly," as written in MMP);
- they allow theoretical coordination (clearly appreciated by the author of RMP, where everything until number 34 is abstract and expressed in pure numbers and aliquot parts);
- their use permits (and asks for) the display of virtuosity.

If we compare Middle Kingdom mathematics with Old Babylonian mathematics (see chapter 3 in the present volume, "Mesopotamian Mathematics"), we shall find no systematic, openly supra-utilitarian pursuits similar to Babylonian second-degree algebra. Analogues of the "humanism" of the Old Babylonian scribe school are also absent from the Egyptian school texts that served to inculcate professional norms and pride in future scribes. As we

[25] Further references in J. Høyrup, "On Parts of Parts and Ascending Continued Fractions," *Centaurus* 33 (1990), 293–324, 310. On $\overline{5}\,\overline{5}$ in particular, see David P. Silberman, "Fractions in the Abu Sir Papyri," *Journal of Egyptian Archaeology* 61 (1975), 248–9.

[26] See Paul J. Frandsen, [Review of J. J. Janssen, *Commodity Prices from the Ramessid Period* (Leiden: Brill, 1975)], *Acta Orientalia* 40 (1979), 279–302, here p. 283.

[27] See Helmuth Brunner, *Altägyptische Erziehung* (Wiesbaden: Otto Harrassowitz, 1957), pp. 11–15; and John Wilson in Carl Kraeling and Robert McC. Adams (eds.), *City Invincible* (Chicago, IL: University of Chicago Press, 1960), p. 103.

see, however, the difference is not absolute, and even in Egypt the scribe school transmuted the knowledge and skills it had to impart. In one respect the impact of schooling was even stronger in Egypt than in Babylonia: the fundamental practical techniques created during Ur III (admittedly within the school) were only affected superficially by the new climate of the Old Babylonian scribe school; instead, "humanism" expressed itself in the grafting of an additional, supra-utilitarian discipline on the curriculum. In Egypt, the systematic use of aliquot parts (more supra-utilitarian than normally recognized) transformed even ordinary mathematical practice.

LINKS?

These similarities are evidently to be explained as parallel developments due to similar conditions, not as borrowings. On the general level, second-millennium Egyptian and Babylonian mathematics are wholly independent from each other. On the level of particulars, the occasional multiplication with a reciprocal in RMP has sometimes been seen as a borrowing, but the context where it occurs speaks against that assumption (cf. above). Only a single problem in RMP (*viz.* no 37) is certainly related to a Babylonian text:

> Go down I [a jug of unknown capacity] times 3 into the *ḥeḳat*-measure, 3′ of me is added to me, 3′ of 3′ of me is added to me, $\overline{9}$ of me is added to me; return I, filled am I.

This can be compared with a problem from Old Babylonian Ešnunna:[28]

> To $^2/_3$ of my $^2/_3$ I have joined 100 sìla and my $^2/_3$, 1 gur was completed. The *tallum*-vessel of my grain corresponds to what?

The Egyptian solution is quite regular, fully based on aliquot parts and grain metrology; the Babylonian solution is no solution at all but a trick which presupposes the solution to be already known. The problem is obviously one of those riddles which the early Akkadian scribe school borrowed around 1800 BCE (see chapter 3 of the present volume). On the other hand, the idiom of "ascending continued fractions" ("*a*, and *b* of *a*," where *a* and *b* are simple fractions) is typically Semitic, and alien to the Egyptian context.[29] A Babylonian borrowing from Egypt as well as an Egyptian adoption of a Babylonian school problem are thus excluded; both must build on

[28] IM 53 957, edited in Taha Baqir, "Some More Mathematical Texts from Tell Harmal," *Sumer* 7 (1951), 28–45, 37, corrections and interpretation in W. von Soden, "Zu den mathematischen Aufgabentexten vom Tell Harmal," *Sumer* 8 (1952), 49–56, 52.

[29] See in general Høyrup, "On Parts of Parts and Ascending Continued Fractions"; since I had not noticed the Ešhnunna parallel at the time, this publication contains some speculations about a possible common Hamito-Semitic language structure; they may now be happily dismissed.

a common source, probably a traders' environment in contact with both regions. In contrast to what happened in Babylonia, however, such borrowings are not likely to have had any deeper influence on Egyptian Middle Kingdom mathematics, which instead developed material and ideas already present in third-millennium scribal computation.

DEMOTIC CREATIVITY AND BORROWINGS

Autochthonous (but after the Middle Kingdom maturation very slow) development remains a characteristic of the Egyptian mathematical tradition into the demotic period – and even into the early Byzantine epoch, as revealed by the Akhmīm-papyrus (written in Greek).[30] Development *does* take place: one undated demotic papyrus[31] tabulates $p\cdot\overline{q}$, $1\leq p\leq 10$, $q = 90$ and $q = 150$ (similar tables are found in the Akhmīm papyrus; a modest beginning in RMP lists $p\cdot\overline{q}$, $1\leq p\leq 10$, $q = 10$, but the context here suggests that other q-values would not be considered); the demotic papyri also transform the old technique of the reference quantity so as to express occasionally proper fractions, treating (e.g.) $5\ \underline{11}$ ("5 seen as part of 11") as a legitimate final result and not as a problem whose solution is $3'\ 1\overline{1}\ 3\overline{3}$; what happens can be characterized as a process of "creative dissolution" of the old canon which does not bring about any new coherence – a close parallel to the changes in the character of the visual arts in Hellenistic Egypt.

In this phase, however, influence from Western Asia is strong – no wonder, given that Egypt had been regularly controlled by Assyrian, Achaemenid, and Greek armies and tax-officials since the seventh century. These contacts (and perhaps trading connections) are likely to explain the use of a variety of formulae which have no Egyptian antecedents but coincide precisely or almost with formulae that were in constant use in Mesopotamia since the third or even fourth millennium: the determination of the circular area as $^3/_4$ of the squared diameter; computation of the volume of a truncated cone as the height times the mid-cross-section;[32] and the use of the "surveyors formula" (average length times average width) to calculate the areas of approximately rectangular quadrangles. Contact is certainly the reason that 8 of 40 problems in P. Cairo *J. E. 89127B30, 89137B43* belong to a characteristic Babylonian type (a reed first standing vertically against

[30] Edition with translation and commentary in J. Baillet, *Le Papyrus mathématique d'Akhmîm* (Mission Archéologique Française au Caire, Mémoires 9, 1; Paris: Leroux, 1892).

[31] P. British Museum 10794, published in Parker (ed.), *Demotic Mathematical Papyri*, pp. 72 f.

[32] The details are of some interest: the surface of the mid-cross-section is not found as in the corresponding Old Babylonian text, but as 1/4 of the product of diameter and arc, the arc being 3 diameters – see Clagett, *Ancient Egyptian Science.* Volume II*: Calendars, Clocks and Astronomy* (Memoirs of the American Philosophical Society, 214; Philadelphia, PA: American Philosophical Society, 1989), p. 75. The latter formula belongs to the lay tradition, is found in one Old Babylonian school tablet (dealing with a semicircle), and recurs in the pseudo-Heronic material.

a wall and then moved to a slanted position), which always involves the Pythagorean theorem, often in a sophisticated way (asking, e.g., for the legs of a right triangle when one leg and the difference between the hypotenuse and the other leg are given).

INFLUENCE ON GREEK GEOMETRY?

From Herodotus onward, common Greek lore asserted that geometry was invented by the Egyptians (either, in agreement with etymology, for surveying and taxation purposes, or by the priests who had sufficient leisure for such concerns). Since Egyptian and Babylonian mathematics became known directly, historians of mathematics have been puzzled by this claim. There is no doubt that the Greeks took their way to deal with fractions from Egypt; the canonical system for pictorial representation certainly influenced sixth-century Greek sculptors, and similar architectural rules may be reflected in Vitruvius; but none of these have anything to do with Greek (theoretical) *geometry*. At least one strain in Greek geometry (the "metric geometry" of *Elements* II etc.), on the other hand, has striking structural similarities with Babylonian algebra (see chapter 3 of the preent volume "Mesopotamian Mathematics"). Before we dismiss the Greek account as pure legend we should take note that the Greeks would only encounter Egyptian geometric practice well after the arrival of Assyrian and Achaemenid surveyors, and that the borrowings into demotic mathematics concern geometry, in particular metrical geometry. It is not to be excluded that early Greek geometry was inspired by what the Greeks encountered in Egypt; if so, the Greeks will have had little chance to know that what they encountered was a fairly recent import there.

Part III

GREEK AND GRECO-ROMAN

9

PHYSICAL AND COSMOLOGICAL THOUGHT BEFORE ARISTOTLE

Daniel W. Graham

Early Greek philosophical and scientific thought began with cosmological speculation. Aristotle, followed by most ancient commentators, recognized Thales of Miletus as the transitional figure who moved from a mythological to a scientific conception of the world. Before him events in the world were understood as brought about by agents, human or supernatural, and often both acting together. After him some thinkers understood natural events as the product of purely natural forces without divine intervention. These thinkers, who later came to be known as philosophers, looked for natural regularities in the world and produced a scientific philosophy.

Before Thales, the most theoretical of mythographers was Hesiod, whose epic poems *Theogony* and *Works and Days* provided a quasi-systematic, if mythical, account of the world. The *Theogony* offered a genealogy of the gods, including in the first generations personified cosmic beings. First Chaos (a cosmic gap or opening) was born, then Earth, then Tartarus (the underworld), then (omitting some less cosmic figures) Heaven, then Hills, then (the inner) Sea, then (the outer) Ocean (116–33). Thus the birth of the gods, the theogony, was also the birth of the world and its familiar features. Among the gods there were struggles for power, which led to a succession of divine monarchs, first Heaven, then Cronus, finally Zeus. In a sort of anthropology, Hesiod tells of five generations of mortals fashioned by the gods, ending with the present race of iron, who are more wretched than their predecessors (*Works and Days* 106–201). Hesiod claims that the Muses appeared to him as he was pasturing sheep on Mount Helicon, and "taught [him] beautiful song," inspiring him with special knowledge (22–3). In his cosmography, which seems to agree with that of Homer, the earth is a flat disk consisting of land with an inner Sea and an outer Ocean circling its rim. Above it is a firmament of Heaven and below is the dark windy abyss Tartarus, perhaps roughly symmetrical with Heaven. Hesiod was considered the authority on theogony, and his cosmography expresses the traditional view of the earth and its environs for the early Greeks.

THE EARLY IONIANS

Thales was born in Miletus, a Greek city of Ionia (on the central Aegean coast of Asia Minor) with trading ties all over the Mediterranean, in about 624 BCE, and died around 545. He apparently left no writings, so we are limited to hearsay accounts of his accomplishments, though there is no doubt that he was renowned in his own day for his scientific and technological abilities. The fifth-century sophist Hippias of Elis records Thales' opinions in his *Collection* of opinions (perhaps from oral sources), and in the next century Aristotle makes a tentative reconstruction of his theories. Thales is said to have traveled to Egypt, and since Miletus had a colony on the Nile, he might easily have made the journey. He held that all is water in some sense. He viewed the earth as like a raft floating on a cosmic sea; ripples on its surface would be felt as earthquakes on land. He is said to have brought back geometry from Egypt. Since the Egyptians were experts in surveying ("geometry" means "land measurement") he may have brought back that skill rather than any mathematical theory. He learned from Phoenician sailors the value of Ursa Minor for navigation, since it lies close to the north celestial pole. He is said also to have helped King Croesus of Lydia cross a river to attack the Persian army by digging a trench behind the Lydian camp to divert half the river's water, thus lowering the water level in the main channel so the river could be forded. He calculated the height of the pyramids by measuring their shadows when the shadow of a man was equal to his height.

His most famous reported feat was his prediction of an eclipse in 585 BCE (modern calculations show that one took place on May 28 of that year). The story that he accomplished this was attested in early reports, and widely known in the Greek world; presumably there is some fact behind the reports. On the basis of his success later sources claimed that he understood that the moon is illuminated by the sun and that eclipses are caused by the blocking of the sun's light. The Babylonians had made progress in predicting eclipses on the basis of repetitions of a cycle of events, and they had begun to track the paths of sun and moon. But their observations, which they had kept for centuries, were valid only for Babylonia, and it is highly unlikely that the Greeks had any comparable data. Thus it is not clear how Thales could have reliably predicted a solar eclipse, even with Mesopotamian data. That he understood the nature of lunar light and eclipses is rendered unlikely by the fact that his followers in Miletus showed no awareness of these relationships in their own astronomical theories. Thus, although the era of scientific explanations can be said to begin with the eclipse of 585, it appears that the prediction was more a matter of luck than of science.

Anaximander of Miletus was born about fifteen years after Thales and died about the same time. He wrote a book, perhaps briefly summarizing his theories, that became one of the first prose treatises in Greek and a model for subsequent scientific writing. His book, like virtually all the writings of the Presocratic philosophers, was lost. But reports of philosophers' views in ancient sources ("testimonies") and quotations from philosophers' original writings ("fragments") were preserved in other writers such as Aristotle.[1] Aristotle's student Theophrastus collected his predecessors' opinions systematically. Theophrastus' own collection was lost, but gave rise to a genre of studies called by modern scholars "doxography," in which lists of opinions of thinkers are collected by topic (and usually without context) for the purpose of comparison and contrast. Doxographies together with fragments and testimonies provide the evidence for reconstructing early Greek philosophy and science.[2] None of these sources is infallible, but scholars can, by comparing and criticizing them, make plausible reconstructions of early theories.

Anaximander described the coming-to-be of the world out of the boundless, *to apeiron*. What exactly it was is unclear, and perhaps that is the point of calling it by a non-descriptive name, which seems to indicate its spatial extent rather than its temporal extent or its indeterminate composition. According to Anaximander,

> that part of the everlasting which is generative of hot and cold separated off at the coming to be of the world-order and from this a sort of sphere of flame grew around the air about the earth like bark around a tree. This subsequently broke off and was closed into individual circles to form the sun, the moon and the stars. (ps.-Plutarch *Miscellanies* 2 = A10)

In some sort of quasi-biological analogy, the concentric masses formed, which eventually produced a disk-shaped earth surrounded by rings of fire enclosed in air (obscuring the fire from our gaze). The rings of fire constitute the heavenly bodies, with the sun farthest out, the moon below it, and the stars (in many rings?) closest to earth. We see the fire in each ring at a hole in the covering of air. A gradual closing and opening of the moon's hole causes the phases of the moon, while sudden closings cause solar and lunar eclipses.

The earth is the shape of a column drum, with a diameter three times its height. The earth stays in place by a kind of equilibrium, since it is

[1] The standard edition of the Presocratics is Hermann Diels, *Die Fragmente der Vorsokratiker*, ed. Walther Kranz, 3 vols. (6th edn; Berlin: Weidmann, 1951) ("DK" for short). Testimonies are given A-numbers, e.g. A10, A11; fragments are given B-numbers. Most other collections of Presocratic materials reproduce the DK numbers. Here testimonies will repeat the A-numbers, while fragments will be marked by "fr." For Greek and Latin texts with English translations, see Daniel W. Graham, *The Texts of Early Greek Philosophy*, 2 vols. (Cambridge: Cambridge University Press, 2010) (translations in the present chapter come from this work) and André Laks and Glenn W. Most, *Early Greek Philosophy*, 9 vols. (Cambridge, MA: Harvard University Press, 2016).

[2] The original collection of ancient doxography is Hermann Diels, *Doxographi graeci* (Berlin, 1879; reprint Berlin: de Gruyter, 1965). The material is now being re-examined; see Jaap Mansfeld and David T. Runia, *Aëtiana: The Method and Intellectual Context of a Doxographer*, 2 vols. to date (Leiden: Brill, 1997–).

equidistant from the rings surrounding it. Anaximander went on to draw a map with the three known continents in the middle of the surface, surrounded by the encircling ocean that was part of the Hesiodic cosmography.

The surface of the earth was once wet, but it is gradually drying up. Life began in the seas and spread to land when animal embryos enclosed in a bark-like shell burst on the shore. Humans too began life in this way, and were nurtured for a long time in their shells so that they were mature enough to fend for themselves when they emerged.

Anaximander explained meteorological phenomena in a naturalistic way. Thunder and lightning result from winds enclosed within clouds. When they burst out, they create a tearing noise and the rent in the dark clouds produces a flash. Winds also cause the turnings or solstices of the sun and moon.

We have one important fragment from Anaximander that tells of the maintenance of the world:

> From what things existing objects come to be, into them too does their destruction take place, "according to what must be: for they give recompense and pay restitution to each other for their injustice according to the ordering of time," expressing it in these rather poetic terms. (fr. 1, from Simplicius *Physics* 24.18–21)

Ancient and early modern sources took this fragment to describe the repeated emergence and destruction of the world from the boundless, as a kind of cycle of original sin and retribution. But the phrase "to each other" indicates that the components of the world pay mutual recompense. The existing objects in question are probably the contraries that comprise the world, such as the hot and cold, the wet and dry, the light and dark. When the hot and dry and light become dominant in summer, they are eventually destroyed and replaced by their contraries in the winter. Despite transgressions of one power over another, the cosmic balance is maintained through time. The excesses of the world are regulated as by a tribunal, and the world can be understood to conform to a law-like order.

Anaximander seems to have invented the cosmogony, a story of how the cosmos emerged. He had a forerunner in Hesiod's theogony, but Hesiod provided a genealogy of persons rather than a succession of natural cosmic events. Anaximander explains how one event led to another until the present balanced world resulted. Whereas Hesiod and Homer explained (or at least understood) thunder and lightning to be caused by Zeus throwing thunderbolts, Anaximander explains them as natural phenomena resulting from natural processes. What his precise view of the gods of mythology was we do not know, but he attributes divine powers to the boundless, which seems to serve as a kind of remote intelligent power.

Anaximander's successor was Anaximenes of Miletus, who flourished around 545 BCE. Whereas Anaximander's cosmogony began from an unknown substance, Anaximenes identifies the boundless stuff of the universe as air. Further, he describes air as changing its character when it becomes more rare or dense. When it becomes rare, it turns into fire; when it is compressed, it turns into wind, and when successively more compressed it turns into cloud, then water, then earth, then stones (A5). Thus there is a series of basic substances ordered by relative density from fire to stones. Anaximenes seems to have used "felting" as a model for condensation. Just as one can apply pressure, heat, and moisture to wool to compress it into a thick, heavy fabric, namely felt, so pressure can make changes to physical substances (A6, A7).

The earth is like a flat disk floating on air. The heavenly bodies are like leaves blown around above the earth, or like nails fixed to a felt cap that rotates around the earth. The heavenly bodies do not move under the earth but around it. High mountains to the north hide the sun at night and produce darkness (A7, A14). As in Anaximander, the present world is the result of some sort of cosmogony. The earth formed first, then the heavenly bodies (A6).

Anaximenes is generally compared unfavorably with Anaximander by modern scholars. Against his teacher's perceptive account of the earth's remaining at rest because of equilibrium, Anaximenes has the earth float on a cushion of air. Against his teacher's cautious reference to the boundless, Anaximenes makes air the ultimate reality. Yet in at least one way he makes a significant advance: whereas Anaximander had left the changes between contraries mysterious and undefined, Anaximenes provides a simple mechanism for change: rarefaction and condensation – two sides of the same process that can change everyday objects as felting does. What was mysterious and metaphorical in Anaximander gets a straightforward theoretical treatment in Anaximenes and becomes comprehensible. Anaximenes even provides an everyday "proof": if one blows with an open mouth the air is hot, if one blows with pursed lips the air is cold, showing that heat and rarity go together, cold and compression (fr. 1). Although Anaximenes' treatment reverses the actual physical relationship between heat and compression, he does try to provide empirical evidence for his view.

Xenophanes of Colophon, a city to the north of Miletus, was exiled from his homeland around 546 and spent most of his long life as an itinerant poet in southern Italy and Sicily. His views seem to have been expounded piecemeal in his poems rather than in a systematic treatise. They included a description of the water cycle, by which water is evaporated, turns into clouds and then falls back to earth as precipitation. He claimed that the earth underwent periods of drying and flooding, as seen by the fact that the fossils of sea creatures are found inland in various places – the first time fossils had been noticed and used as evidence for theories of nature (A33).

Xenophanes held that the earth was an infinite plane dividing earth, which reached downward endlessly, from air, which reached upwards, probably endlessly (fr. 28). Different suns and moons were seen in different portions of this infinite world, apparently created by similar meteorological conditions. Each day a new sun passes over us from east to west and continues on endlessly to the west (A41, A41a). The heavenly bodies are luminous cloud-like entities, as are rainbows, St Elmo's fire, and other phenomena (A38, A39, fr. 32).

Heraclitus lived in Ephesus, between Colophon and Miletus, in the late sixth century. His riddling prose treatise criticized common knowledge. Though often taken since ancient times as a radical flux-theorist who denied the law of non-contradiction, he seems rather to have held a more subtle and attractive theory. Far from saying that you can't step twice into the same river (a view reported by Plato, *Cratylus* 402a), his actual words seem to say that on those stepping into *the same rivers, different waters* flow (fr. 12). In other words, changing waters constitute the same rivers. If the waters ceased to flow, the rivers would cease to be. Thus high-level order seems to emerge from low-level change and metabolism (much as in Aristotle's biological theory).

Heraclitus seems to have criticized his Ionian predecessors for misunderstanding the implications of their own theories. For instance, in Anaximenes, air is the ultimate reality, which changes into a range of stuffs from fire to stones. But if that is the case, Heraclitus suggests, why is air more important than any other stuff in the series? What is important is not any stuff, which forms only a temporary stage of a process, but rather the process itself, or the principle that governs changes in the series. Heraclitus himself envisages a series of three stuffs: earth, water, and fire. Each one changes into its neighbor, and does so according to a fixed proportion: so much earth becomes so much water, which in turn becomes so much fire, and vice versa (fr. 31). He likens the process to a monetary exchange: "All things are an exchange for fire and fire for all things, as goods for gold and gold for goods" (fr. 90). Thus there is a kind of law of interchange between the stuffs. Fire plays a key role, but arguably more as a standard of value and symbol than as a fundamental reality. For according to Heraclitus, fire does not remain, but each stuff perishes as another is born out of it (fr. 36, fr. 76). Fire in fact exemplifies the impermanence of stuffs and hence their subordination to the larger pattern of transformation.

This raises a question about the theory or theories of matter that were current before Heraclitus. According to Aristotle, the early Ionians were material monists, who held that there was only one reality or "principle" (*archê*) of which all other stuffs were appearances. For instance, Thales said that all was water, while Anaximenes said all was air and Heraclitus said all was fire (*Metaphysics* 1.3, 983b6 ff.). If this is right, then for Anaximenes water is just condensed air, that is, air in a dense, liquid state, while earth is air in a still more dense, solid state. Aristotle's interpretation has been tremendously influential and remains dominant to this day. Yet there is a competing interpretation according to which, for Anaximenes, when air becomes more condensed, it is transformed into another kind of matter, e.g. water and then earth, which is not air in a certain state but another replacement substance. This is explicitly how Heraclitus describes the successor substances of fire, so that if Aristotle takes Heraclitus as a material monist, he has misread him; and he may have misread Heraclitus' predecessors as well. In any case, Plato takes Anaximenes' theory as presupposing the generation of new substances out of old (*Timaeus* 49b–c), which should at least make us cautious about accepting Aristotle's reading.

PARMENIDES AND ELEATIC THEORIES

Parmenides of Elea in southern Italy wrote a remarkable philosophical poem around 490 BCE. The poem ostensibly records the narrator's visit to a mythical place at the edge of the world where he meets a goddess who teaches him about reality (fr. 1). Despite the religious trappings of the poem, the goddess tells the narrator to judge the argument by its own merits (fr. 7). She rejects the possibility of coming-to-be and perishing (fr. 2). There are verbal parallels with Heraclitus' work that suggest Heraclitus is an intended target of criticism, though this point is highly controversial. In any case, the poem provides an extended rejection of coming-to-be in favor of changeless being. The only thing that exists is what-is, since what-is-not is unintelligible and impossible. Coming-to-be presupposes what-is-not as a starting point, so it is ruled out. What-is (or, being) is also indivisible, motionless, and complete.

Although there are numerous problems in interpreting the argument in its poetic expression, clearly there is an extended argument, and it seems, on the dominant interpretation, to rule out not only coming-to-be but any kind of change and even any plurality. Thus Parmenides seems to defend an austere monism which precludes science as we know it. After the argument about what-is (the section known in the tradition as "Truth"), however, the poem surprisingly provides a detailed cosmology, one that is said to be "deceptive," but which the goddess claims is superior to other cosmologies (in the section known as "Opinion"). In this cosmology the phenomena of the world are derived from two "forms" or elements, light and night, which are said to be equal to each other but contrary in qualities (fr. 8.50–61; fr. 9). These forms seem to have the properties of being which the goddess had defended, in so far as they can have properties. Other stuffs are composed of mixtures of the two basic forms.

Parmenides' poem raises significant problems for interpretation. The Truth section seems to present an ontology that precludes cosmology, while the Opinion section advocates a particular cosmology as being preferable to others. Yet the Truth seems to obviate the need for the Opinion. Perhaps the cosmology is meant as a kind of anti-cosmology, as if to say: this is better than any competing cosmology, but it is false, so every cosmology is false. On this view, the ontology disqualifies the cosmology, which is presented for purely dialectical purposes. On the other hand, the cosmology could be taken as an example of how one *should* construct an ontology while abiding by the constraints of the ontology. For instance, the ontology rules out coming-to-be and perishing. Accordingly, the cosmology presents a cosmology in which the ultimate constituents of the world do not come-to-be or perish, although various phenomena *appear* to come into being and perish.

Modern interpreters have almost universally taken the first approach, reading the cosmology in light of the ontology, and so dismissing it as a mere exercise in refutation. They typically take the "pluralist" philosophers who developed cosmological theories after Parmenides as rejecting his theory, while retaining his ban on coming-to-be and perishing. The pluralists go on to posit a plurality of elemental realities that combine or separate to produce the appearance of coming-to-be and perishing. But this overlooks the extent to which the pluralists learned from Parmenides' cosmology: they might well have pursued the second approach, constructing their cosmologies using Parmenides' cosmology as a model, thus remaining within the constraints of his ontology. In fact the latter approach seems to capture their method better than the first: there is no clear indication that any of the pluralists actually criticized Parmenides, while there is ample evidence that they accepted his principles. Note that this could be true even if modern interpreters are right in making Parmenides an anti-cosmologist: the pluralists may have misunderstood his overall program, impressed by the brilliance of his success in both ontology and cosmology.

And Parmenides' cosmology was impressive. Although we have fewer lines of the Opinion than of the Truth section, we know that he made important discoveries, or at least advanced hypotheses that were later vindicated. In the first place, he identified the morning star with the evening star (A40a). While the identity of the two manifestations of Venus was known to Babylonian astrologers a thousand years before Parmenides, there is no sign that any Greek cosmologist had made the connection before him. Secondly, Parmenides recognized that the moon gets its light from the sun (fr. 14, fr. 15). This no one had recognized, to our knowledge, in any culture. Finally, Parmenides claimed that the earth was spherical (Diogenes Laertius 9.21, 8.48 = A44). All his Greek predecessors had assumed a flat earth, whether a raft, an infinite plane, or a disk floating in mid-air. We do not know what evidence or theory led him to this claim, but it would receive scientific support by the mid-fourth century. Given that the theory of the moon's light can be justified by careful observations of the moon's phases in relation to its angular distance from the sun during a lunar month, we can say that Parmenides made the first Greek astronomical discoveries with this and his recognition of Venus. For the first time Greek science made concrete advances. Thus Parmenides might well be thought to be a serious cosmologist and not merely a refuter of scientific inquiry.

Parmenides' followers, Zeno and Melissus of the so-called Eleatic school, proposed further arguments against change and plurality. Parmenides argues against change, but, while his successors understood him to be arguing against plurality, we find no explicit argument to that effect in his fragments. Later thinkers tended to see the Eleatics as sharing the same position, although there are some notable differences among them, and reacted to them as a group.

POST-PARMENIDEAN COSMOLOGY

Anaxagoras of Clazomenae (ca. 500–428 BCE), another city of Ionia, spent thirty years in Athens in the circle of Pericles before leaving to go to the northern Aegean. Accepting Parmenides' rejection of coming-to-be and perishing, he posited the existence of an indefinite number of stuffs and qualities which retain their natures but mix with each other. (One interpretation reduces the stuffs to mixtures of contraries, but the evidence for this is problematic; Anaxagoras mentions both stuffs such as earth and qualities such as hot and cold as components of things, but he does not expressly reduce one kind of being to another; see fr. 4a.) In the beginning these things were all mixed together, but Mind or intelligence (*nous*) caused a circular motion that turned into a vortex that caused a mechanical separation of ingredients (fr. 1, fr. 12). The heavy elements gathered in the center and eventually formed a flat earth, composed of earth; around it were water (as seas), air, and fire (fr. 15, fr. 16).

According to Anaxagoras, everything is mixed with everything (the "universal mixture" axiom), except for mind, which is found only in some things (fr. 12). This was obviously true in the initial state of the world, but it continues to be true everywhere, for matter is infinitely divisible (another axiom), and not all matter of one type (e.g. water) can be extracted from any stretch of matter. On this model, anything can emerge from anything, because there is always some quantity of matter X in any body of matter Y; on the other hand nothing, that is, no element or basic reality, ever comes to be or perishes ("no becoming" axiom). And whatever element is present in a mixture in the largest quantity gives its character to the whole ("predominance" axiom).[3] Thus when some stuff seems to appear, as in water

[3] For the axioms of Anaxagoras, see George B. Kerferd, "Anaxagoras and the Concept of Mattter Before Aristotle," *Bulletin of the John Rylands Library*, 52 (1969), 129–43.

condensing on a cold glass, what happens is just a change in concentration whereby water that was present in the air gathers together on the side of the glass. In fact, Anaxagoras' theory of matter follows the principles of elemental mixture exemplified in Parmenides' cosmology.

Anaxagoras' cosmology builds on the vortex motion of the primordial condition. The vortex precipitated differentiations in distinct concentric regions, as in a centrifuge. The vortex continues to expand over a wider area and to take in more undifferentiated matter at its periphery. The heavenly bodies are heavy earthy or stony spheres that are held aloft by the force of the vortex. Some vortex-like principle goes back to Anaximander, but whether it was conceived as a powerful cosmic force before Anaxagoras is unclear; in all earlier cosmologies the heavenly bodies (other than the earth) were composed of relatively light materials that did not need a strong centrifugal force to maintain them aloft. Anaxagoras, however, said that were it not for the powerful whirlwind above, heavenly bodies would fall to earth.

Anaxagoras went on to make some remarkable claims about heavenly bodies. As we have seen, he claimed they were massy, solid bodies. Their orbits (or some of them) passed below the earth. There were dark (invisible) bodies orbiting below the moon. The moon gets its light from the sun. The orbit of the moon lies below that of the sun. And solar and lunar eclipses are caused by a blocking of the sun's light, by the moon in the former case, and by the earth or a dark body in the latter (Hippolytus *Refutation of All Heresies* 1.8.6–10 = A42).

The claim about the sun's being the source of the moon's light comes from Parmenides. The other points of the theory seem to arise as a result of careful deductions from this point. If the moon's phases are caused by the light of the sun falling on it, the moon must be spherical to realize the given pattern of shadows; it must be opaque; and it must lie below the sun (or it would never become completely invisible at the time of conjunction). But if the moon is spherical, all heavenly bodies may be so as well. If the moon is spherical and opaque, it must be a massy, solid object – Anaxagoras says it is earthy and has mountains and valleys on its surface. This means that it is a physical object that continues in existence even when it is invisible. The sun must remain in existence during the night, since in the middle of the lunar month the moon shines all night, reflecting the light of the invisible sun, which now lies below the earth.

Some of these points seem trivial, but they were not in a time when some theorists, including Xenophanes and Heraclitus, said the sun was new every day; when some, including Anaximenes, Xenophanes, and presumably Thales, said heavenly bodies did not pass under the earth; when no one before Parmenides said the heavenly bodies were spherical, but rather ring-shaped, flat, cloud-like, or bowl-shaped; when everyone before Parmenides said the moon shone by its own light; and when no one gave a correct account of eclipses. After Anaxagoras most of the pre-Parmenidean theories of astronomy quietly disappear, as if later theorists were convinced by Anaxagoras' arguments and evidence. Aristotle offered Anaxagoras' theory of eclipses (he does not name its author) as a paradigm of scientific explanation.

One important event occurred serendipitously during Anaxagoras' career. A large meteoroid fell to earth about 467 BCE near Aegospotami on the Hellespont, in a fiery ball during daylight hours. The remaining meteorite was about the size of a wagon, presumably resting at the bottom of a large crater, and became a tourist attraction for five hundred years. It was said that Anaxagoras had predicted its fall. More likely his theory predicted the possibility of such an event, and when it happened, it became the stuff of scientific legend rather than of mythology (as would have happened a century earlier). Meteoroids were not known before this time (shooting stars could be taken as meteorological phenomena) but became a staple of scientific explanation thereafter, showing again that a rich interplay of empirical observation and theoretical speculation was already taking place. Significantly, all later Presocratic astronomical theories posited the existence of earthy or stony bodies.[4]

Empedocles of Acragas in Sicily (ca. 495–435) developed another pluralistic theory. Like Parmenides he wrote in verse in one or two poems (it is controversial whether the titles *On Nature* and *Purifications* refer to two poems or are alternative titles for the same poem). Deeply religious, Empedocles believed in reincarnation and integrated a psychology of a fall and redemption with his natural philosophy. His language is especially ornate and his expressions complex, making interpretation of his work more than usually difficult, despite the large number of lines we have.

Empedocles held that there are four "roots" (*rhizōmata*) or elements – earth, water, air, and fire – which are eternal and self-identical, but which mix with each other to produce other stuffs as their compounds, for instance flesh, blood, and wood (fr. 6). In language similar to that of Parmenides he denies that anything comes into being and perishes, though he admits he too uses these words to describe natural changes. There are also two forces that act on these elements, personified as Love and Strife, roughly a force of attraction and a force of separation. The former brings together unlike elements, while the latter repels them, making them return to their own kind (fr. 17.1–20).

The compounding and separating of elements take place against the background of a cyclical cosmogony. The cosmic cycle is controversial; it may consist of four stages (on the more traditional view) or two. The four stages are: (1) increasing Love, producing compounds; (2) a complete unification of elements in the "Sphere" where no differentiation is possible; (3)

[4] On astronomical advances, see Daniel W. Graham, *Science Before Socrates: Parmenides, Anaxagoras, and the New Astronomy* (New York: Oxford University Press, 2013).

the breaking up of the Sphere in a time of increasing Strife; and (4) a complete separation of elements, perhaps into concentric spheres with no compounds present. On this view there are two zoogonies or creations of living things, in stages (1) and (3). On the two-stage view there is (1) a period of conflicting Love and Strife in which compounds are present alternating with (2) the Sphere. Recently it has been suggested that there may be lesser cycles between recurrences of the Sphere. In any case, for Empedocles the cosmos we know alternates with one or more stages in which a cosmos is not present.

In astronomy Empedocles generally follows Anaxagoras. His sun is lentoid rather than spherical and may serve as a lens to concentrate light. But the sun's light illuminates an opaque moon, and night is caused by the sun's sinking below the surface of the earth. Solar eclipses are caused by the moon's blocking the sun's light to earth.

The post-Parmenidean theory that was destined to be the most influential was atomism. Begun by Leucippus, about whom we know almost nothing, it was developed further by Democritus, who was born between 470 and 460 and was thus a contemporary of Socrates. The basic realities are atoms and the void, which they called also "thing" and "not-thing" (Democritus A37, fr. 156). The notion of an indivisible (*atomos*) particle of matter may have arisen from trying to avoid paradoxes of divisibility raised by Zeno. It is also possible that Leucippus saw atomism as the natural development of Parmenides' cosmology: as night and light were respectively dense and rare, one had only to suppose that there was an absolutely dense and an absolutely rare state of affairs to account for all phenomena. Visible matter could not be absolutely dense because it can be broken up; but if there were tiny particles invisible to the eye, these could have the Eleatic properties of being. That is, each one could be absolutely dense and changeless – except for changing in place. The relatively empty spaces between physical objects cannot be completely empty because they are full of air (as earlier philosophers had argued). But the atmosphere would have fewer particles and more empty spaces between them at a microscopic level. Thus there would be no coming-to-be at the atomic level, in accordance with Parmenides' principle, but only rearrangements of atoms, while at the macroscopic level there would be an appearance of coming-to-be and perishing.

According to Leucippus' cosmogony atoms travel at random in an infinite space. Sometimes conditions are right for streams of atoms to interact in such a way as to produce a vortex. This action grouped like atoms with like, producing stratified masses. Some atoms clung together and formed a membrane-like container for the vortex. The earth formed from heavy atoms at the center of the vortex, while moist atoms at the circumference stuck together but eventually dried out and heated up, forming the heavenly bodies (Leucippus A1). According to Democritus, heat from the sun helped to dry out the earth, which was muddy at first. While it was drying out animals emerged from the earth by a process of spontaneous generation, which ceased when the earth became drier. Similar conditions produce similar results throughout the cosmos, so there must be other worlds than ours, although they do not necessarily have the same astronomical features as ours (Democritus fr. 5.1).

Atomism was a reductionistic theory that allowed simple bodies of different shapes and simple laws of motion to account for complex interactions at the macro level. Fire atoms were round and hence volatile; earth atoms were presumably larger and jagged so that they tended to clump together. Single atoms moved through the void, occasionally colliding with other atoms and, under certain conditions, combining with them. Aggregates of atoms could be dissolved by collisions with other aggregates.

In cosmology, Democritus ordered the heavenly bodies as follows: outermost the fixed stars, below them the planets (minus Venus), then the sun, then the morning star (Venus), then the moon. According to him the force of the vortex motion diminished with lower altitude, so that the more quickly the other bodies fell behind the fixed stars in their orbits, the lower and closer to the earth they were (A88). The earth itself is disk-shaped (drum-shaped according to Leucippus) and concave on its upper surface (Leucippus A26; Democritus A94). The vortex motion in the heavens somehow holds it in place (Philoponus *Physics* 262.8–13). The southern regions of the earth are heavier because of the greater mass of vegetation that grows there, causing the earth to tilt towards the south (A96). (Leucippus attributes the tilt to the fact that the air under the southern region is rarer than that under the northern because of its warmth, and hence does not support it as well; Leucippus A27.) According to Democritus, motions of wind and water in hollows of the earth cause earthquakes (A97, A98).

Democritus seems to have argued for this unusual ontology for the first time, saying that what-is is no more than what-is-not (fr. 156). What the premises were for this "indifference argument" is not clear. In any case, he seems to have tried to answer arguments offered by Melissus against a void. One of his arguments may have started from the fact that there is multiplicity in the world, and hence division and void.

The atomists seem to have operated for the first time with a purely mechanical system. There was no mind, no Love and Strife, no cosmic steersman, no *Logos*, but only atoms – particles of matter in motion – and void. Chance encounters of atoms accounted for all causal interactions and all visible phenomena. Higher-level regularities could arise out of microscopic actions under certain conditions, allowing for scientific accounts of meteorology, astronomy, and even biology and anthropology.

In the mid-fifth century Diogenes of Apollonia presented a monistic cosmology. Everything had to be composed of one thing, or objects could not interact with each other. There was only one reality, namely air. His theory was either a pedestrian revival of Anaximenes' theory (the majority view) or the introduction of material monism as a response to Eleatic concerns. Diogenes saw in the order of the world an indication that everything was arranged for the best, which could only happen with intelligence maintaining the world. Air was the basis of intelligence (since animals are intelligent so long as they breathe), so air controlled the world. Diogenes' astronomy posited rocky heavenly bodies that ancient testimonies tell us were inspired by Anaxagoras' meteorite.

Philolaus of Croton (in southern Italy), a Pythagorean, developed an unusual cosmology. Things were composed of limiters and unlimiteds, including the world and its contents. Unfortunately, we do not get any examples of these components. But they seem to come to something like Aristotle's notions of form and matter. Harmony was necessary to combine disparate elements such as the limiter and the unlimited. The first thing that came to be in the middle of the world was "the hearth," a fiery center point. Around this the heavenly bodies circled, including the sun, moon, planets, and stars. The earth was one of the circling bodies, which by traveling on its orbit (not by rotating on its axis) created day and night as it passed the sun. For the first time an astronomer treated the earth as a moving body, in a theory that had some explanatory power. Philolaus recognized harmony as a power necessary to hold unlike things in a unity; this notion is reminiscent of Empedocles' Love, which the latter also identifies with harmony.

The second half of the fifth century BCE saw the rise of the sophists, a group of itinerant educators who gave courses, usually on practical topics including public speaking and statecraft, for a fee to audiences. The need for such teaching arose from democratic governments which offered opportunities to citizens, but presupposed abilities that few possessed – such as skill in public speaking and knowledge of public administration. While the education that was most in demand related to political science and economics, some sophists offered teaching in areas including science and mathematics, notably Hippias of Elis. The writings of Antiphon the Sophist included cosmology, though he seems to have offered few original theories. Gorgias of Leontini (in Sicily) took up Eleatic themes in arguing against the Eleatics the paradoxical thesis that nothing exists.

THE SOCRATIC TRADITION AND PLATO

During the time of the sophists Socrates of Athens emerged as an idiosyncratic and charismatic figure who addressed some of the problems of the sophists – especially whether virtue or excellence was teachable – but who claimed no special expertise and charged no fees to those sharing in his inquiries. Although he may have studied with the natural philosopher Archelaus (Diogenes Laertius 2.23), he seems to have renounced any interest in cosmology in so far as he regarded questions about human goodness as being both more urgent and more accessible to human inquiry than cosmological questions (Plato *Apology* 19b–d; Xenophon *Memorabilia* 1.1.11–16). The sophists sometimes claimed that they possessed an art or craft (*technē*) of virtue or excellence, implying that they understood its principles and could pass them on to others. Socrates picked up on this claim and fashioned the craft analogy into a double-edged sword: on the one hand one who truly possesses a craft should be able to prove his worth by pointing to successful outcomes, as a doctor with his patients and a housebuilder with his buildings; but it is doubtful that any self-proclaimed expert (the meaning of "sophist") made his students more virtuous. On the other hand, there is no more important ideal than to achieve the "political art" of making people better, so that this should be the constant goal of the philosopher. After his death Socrates' devoted followers promulgated his views, or their own versions of them, so successfully that they produced a major shift of emphasis in philosophy toward the ethical and political.

Eventually Plato emerged as the most influential Socratic writer and educator. His early or Socratic dialogues stay within the range of issues ancient sources unanimously attribute to Socrates. But his middle dialogue *Phaedo* raises the possibility of a Socratic approach to science (96b–100b). The character Socrates recounts his early interest in science and the confusions his studies brought to him. At this point he heard someone reading the words of Anaxagoras, saying that Mind had organized all things. He recognized that if this were true, all things would be directed toward some good purpose, and this purpose or good could provide a starting point for scientific explanation. He subsequently got hold of Anaxagoras' treatise, only to find that the author did not invoke any such explanation, but only used mechanical and materialistic explanations of events. Socrates goes on to point out the limitations of such a theory in his present circumstances: if one were to explain why Socrates was in prison on the basis of the disposition of his limbs, one would fail to explain the most important causes of his incarceration, namely that he had made certain choices, the citizens of Athens had made other choices, and all of these with some good in view. Socrates goes on to say that he has failed to find any method to follow up on the insight of Anaxagoras, but he recommends a method of hypothesis by which one lays down a theoretical postulate and sees what will follow from that, regarding those implications as true. This sounds like a promising account of hypothesis in scientific theory construction, except that Plato says little or nothing about the crucial step of testing the hypothesis against facts or appearances. Socrates goes on to propose the theory of Forms, explaining that things are beautiful by Beauty itself, and so on. Furthermore, some characteristics "bring along" or entail others, for instance fire entails heat, so that there is some kind of implication among qualities.

Plato does not develop any explicit scientific theory in the *Phaedo*, but in a concluding myth (110b–115a) he presents a cosmography and an eschatology which strikingly portray a large, spherical earth whose interior is perforated with caverns where rivers of water and lava flow and wicked souls are punished. There is a hint that the cosmography may be colored by the principle gleaned from Anaxagoras: the world must be arranged for the best. Hence if one can divine what the best arrangement would be, one may postulate that as the correct shape and disposition of the world. Plato agrees with Parmenides in positing a spherical earth (though for what reasons, and with what scientific evidence we do not know). He agrees with Empedocles (from volcanic Sicily) in locating reservoirs of lava beneath the surface of the earth. (Biographers say that Plato had visited Mount Etna on his first voyage to Sicily. The *Phaedo* seems to have been written soon after this trip.)

In the late dialogue *Timaeus*, Plato returns to the subject of science for a detailed but also very tentative inquiry. In this dialogue a visitor from Italy, Timaeus, makes a long speech developing a cosmological theory, covering much the same ground as Presocratic theories. But he makes some very different assumptions from those of the Presocratics. Drawing on the Platonic distinction between what-is and what comes to be, corresponding to the cognitive states of knowledge and opinion, respectively, Timaeus asserts that the world, as a sensible being, falls into the latter class and is grasped only by opinion (*Timaeus* 29b–d). He postulates a cosmic demiurge or craftsman as the organizer of the world, without argument, and goes on to argue that this being must have modeled the world after a divine archetype, because it is beautiful (27d–29a). Whereas the archetype is changeless, the world is changeable. Because the world is an object of opinion, the best explanation that can be managed is a likely story about it, not a scientific demonstration. Timaeus goes on to argue that since the demiurge is good, and not jealous (tacitly criticizing the gods of Greek mythology), he wishes everything to be as good and like himself as possible (29d–30c). Accordingly he put reason in soul, and soul in the body of the world, making it a single living creature like its divine model (30b–31a). He made the world (including the heavens) spherical and pervaded by soul (33b–34c).

Despite the fact that the topics covered in Plato's cosmology approximate those probably found in most Presocratic treatises *On Nature*, his starting point is radically different from theirs. He begins with a mysterious creator god, unknown to mythology and philosophy alike, who has an eternal archetype to follow and the benevolence to exemplify it in a changing world. The framework Plato invents allows him to pursue the insight of Anaxagoras which was so imperfectly realized in Anaxagoras' own work: to show how the universe is arranged *for the best*, specifically as a piece of cosmic engineering. Hence Plato can use a speculation on how the best possible world would look to deduce its nature and even the order of its creation. This is, as has recently been emphasized, the beginning of systematic creationist accounts of the world in the Greek tradition.[5] Behind this approach is the faith that values and ideals are prior to facts; the world did not arise by chance and then progress to embody order and beauty, but rather it was designed from the outset so as to embody value. Thus rather than art imitating nature, nature imitates art, or rather is the expression of divine technology. The craft analogy introduced by Socrates comes to dominate the field of cosmology.

The demiurge first made soul, from which he created the world's soul, including the circle of the same and the circle of the different, the orbits that control the daily rotation of the heavens and the annual rotation of the sun, respectively. Time arose with the revolutions of the heavenly bodies to provide regular measures. The demiurge created the gods and commanded them to create lesser creatures. The gods designed the human body to be serviceable to its soul: first they made the spherical head, then the body and limbs to support it, distinguishing front from back and fashioning the organs (44d ff.; 69c ff.). Plato presents creations such as the human body as engineering problems to be solved sometimes by a tradeoff among demands. For instance, the human head could be made with a thicker skull and heavier flesh to protect it and make life longer, but then it would be dull and stupid; instead the gods decided to make it relatively light and sensitive, so that humans could live a better life through a higher level of intelligence (75b–c). In general, the ends aimed at by the demiurge and the gods who further his projects can only be carried out to the extent that matter and physical laws permit, so that this best of all possible worlds falls far short of the ideal.

The sensible powers such as hot and cold, wet and dry, were rendered orderly by the imposition of triangular structures that characterize the four elements. Three of these can change into one another by making different solid figures from the plane triangles, whereas earth is composed of an incommensurable triangle (an isosceles right triangle) so that it cannot turn into other elements, though it can form solid figures of different sizes.

One member of Plato's Academy who may have built on his cosmology was Eudoxus. He developed a mathematical model of the heavens that used the principles suggested in the *Timaeus*, namely geocentric uniform circular motion, to account for the complex motions of the planets. Eudoxus envisaged a structure of nested spheres, each moving on its own axis at its own speed in order to explain the apparently irregular motions of the planets as the result of regular motions.

Early Greek scientific thought appeared suddenly and developed steadily to the time of Aristotle. The scientific attitude itself marked a major departure from earlier modes of thought. Just the idea that the events of the

[5] David Sedley, *Creationism and Its Critics in Antiquity* (Berkeley, CA: University of California Press, 2007); Andrew Gregory, *Ancient Greek Cosmogony* (London: Duckworth, 2007), pp. 140–62.

heavens and earth were natural processes with natural causes invited inquiry into the nature of those processes. The habit of raising scientific questions in the public forum and submitting answers to criticism and testing have in the long run generated the scientific advances we now take for granted as the legacy of science. Speculative theory drove these early accounts, leaving only a limited role for testing conclusions against reality. Accordingly, critics, starting in the late fifth century BCE (including sophists, Socrates, and medical writers) have charged the Presocratics with being too theoretical. Two defenses can be offered. First, theories themselves are liable to a certain kind of criticism, which they received at the hands of other theorists. Thus Ionian accounts of material coming-to-be and perishing were rejected in favor of mixtures of elements – on purely theoretical grounds. Second, critics have consistently overlooked important empirical advances made by Presocratic thinkers, including meteorological and astronomical explanations that have proved correct: the water cycle; the nature of precipitation; lunar light; eclipses; the shape and composition of (some) heavenly bodies. Plato introduced mathematical considerations, including his triangular theory of matter and the notion of composite motions in the heavens. But even the Presocratics had adumbrated such ideas, for instance in Anaximander's ratios of orbits and Heraclitus' ratios of substances in their transformations.

In the end the early thinkers canvassed almost every possible theory of cosmogony and matter. Some held that matter was transformed from one kind to another; others held that there were changeless elements that mixed and separated; others held that one type of matter manifested itself in different ways under different conditions; some held that matter was continuous, others that it was discrete and atomic. When modern physics and chemistry revisited these questions, they had models of each possibility. The theory of elements turns out to be true for normal chemical interactions; the theory of transformation is true for nuclear changes; material monism is consistent with the phase-states of matter; matter is atomic in composition. Philosophers in the Socratic tradition demanded a further integration of facts and values, and assigned a greater role to mathematics in theory construction. As questions of the origins of the universe have come center stage in contemporary physics, the inquiries of the early philosophers seem more relevant than ever.

10

ARISTOTLE: AN OVERVIEW

Andrea Falcon

INTRODUCTION

Aristotle, the son of Nicomachus and Phaestis, was born in Stagira in the year 384. His father was a physician at the court of the Macedonian king Amyntas III. In 367, at the age of seventeen, Aristotle moved to Athens, where he joined the school founded by Plato, the Academy. He left the Academy and Athens shortly after Plato's death in 347. It has been suggested that his departure was the result of an intellectual crisis, or perhaps a response to the election of Speusippus as successor of Plato at the head of the school. We do not know. In all probability, his decision to leave the Academy and Athens was the result of a combination of several factors, including the political climate. In the summer of 348 the Macedonian king Philip II captured and destroyed the city of Olynthus, which was an ally of Athens. This generated an outbreak of anti-Macedonian sentiment in Athens. At that point, the close ties to the Macedonian elite that Aristotle cultivated throughout his life must have become a liability. It is telling that Aristotle was only able to return to Athens in 335. At that time, the city was *de facto* under Macedonian rule. Aristotle did not come back to the Academy. Instead, he established his own school, the Lyceum (also known as Peripatos due to the existence of a covered walk, *peripatos*). Under Macedonian protection, he could teach there until the unexpected death of Alexander the Great in 323. As the news of the death reached Athens, Aristotle had to leave the city. He retreated to Chalcis, where he died in the year 322.

The first period away from Athens has been considered especially important for the development of Aristotle's science. It has been observed that some of the data about animals collected in the *History of Animals* make reference to places on the coast of Asia Minor. This observation has suggested a connection between the research activities that constituted the basis of the impressive corpus of writings on animals (what we call, for a lack of better term, Aristotle's biology) and the years that Aristotle spent away from

Athens, especially the years of residence in the Troad (Assos) and on the island of Lesbos (Mytilene).[1] This connection is suggestive, but it is not possible to attribute Aristotle's interest in the study of animals, or more generally Aristotle's interest in the study of nature, to any particular period of his life. This interest is not only a prominent aspect of his thought but is also a central feature of his entire life. Arguably, it is also his most significant legacy.

A brief review of the reception of this legacy in antiquity will help us to appreciate how innovative Aristotle was as a thinker and as a scientist. In the *Lives and Opinions of Eminent Philosophers*, Diogenes Laertius tells us that Aristotle stood out for his ambition to find the causes not just of many things, but of everything:

> In physics, [Aristotle] explained more than anyone else, so that he gave the causes even of the smallest things. And for this reason, he wrote not a small number of books of notes on nature.[2]

This passage is evidence that, from very early on, many were impressed by the sheer scope of Aristotle's science. Yet, what was perceived by some as strength was regarded by others as weakness. Consider the ancient tradition reflected in the following passage from the tenth-century Byzantine lexicon *Suda*:

> Aristotle was a scribe of nature (*grammateus tês physeôs*), steeping his pen in the intellect; from whom perhaps it was not necessary to seek anything useful, even if it is fairly technical and exceptionally worked out.[3]

This tradition, which is certainly not a friendly one, has no difficulty recognizing that Aristotle has written more than anyone else on the topic of nature, and that his writings are exceptionally worked out. This tradition, however, has serious doubts about the usefulness, and presumably also the philosophical significance, of what is accomplished by Aristotle. Tellingly, Aristotle is compared to a scribe (*grammateus*). Given the polemical nature of the context in which this comparison is recalled, there is no doubt that *grammateus* is used as a term of contempt. It means something like clerk.[4]

1 This suggestion was first made by D'Arcy Thompson in the Prefatory Note to his 1910 translation of the *History of Animals*. It was further elaborated by Desmond Lee in "Place Names and the Date of Aristotle's Biological Works," *Classical Quarterly* 42 (1948), 61–7. *Contra* F. Solmsen, "The Fishes of Lesbos and their Alleged Significance for the Development of Aristotle," *Hermes* 106 (1978), 467–84.

2 Diogenes Laertius, 5.32, translated by R. W. Sharples, *Peripatetic Philosophy: 200 BC to AD 200* (Cambridge: Cambridge University Press, 2010).

3 *Suda*, 3930 Adler, *s.v.* "Aristotle," trans. C. Roth, *Suda On Line*, April 2012, www.stoa.org/sol-entries/alpha/3930.

4 The comparison of Aristotle to a scribe of nature is recalled by the Platonist Atticus (second century CE). See, Atticus fr. 7. 48–9 *Des Places* (E. des Places, *Atticus: Fragments* Paris: Les Belles Lettres,

Clearly, the ancients were ambivalent about knowledge won through a detailed study of nature. Aristotle addresses these worries toward the end of the first book of *Parts of Animals*. There, he introduces his study of animals with the following words:

> Since we have completed stating the way things appear to us about the divine things, it remains to speak about animal nature, omitting nothing in our power, whether of lesser or greater esteem. For even in the study of animals disagreeable to perception, the nature that crafted them likewise provides extraordinary pleasures to those who are able to know their causes and are by nature philosophers.[5]

The claim that it takes those who are by nature philosophers to enjoy the study of nature in all its forms and manifestations is best understood in light of a claim made right before our passage. There, Aristotle says that the study of the heavens is "philosophy concerned with divine things."[6] In antiquity, it was fairly common to identify philosophy with the study of the heavens. The ancient story that has Thales fall into a well while observing the stars is an obvious example of the association between philosophers and the study of the heavens. The following anecdotes associated with Anaxagoras, who was universally regarded as the champion of the study of nature before Socrates, are indicative of the same attitude:

> when he was asked for what purpose he was born, Anaxagoras replied: "for the study of the sun, the moon, and the heavens."[7]
>
> when someone asked him "have you no care for your country?" he [*sc.* Anaxagoras] replied, "I am very concerned about my country," and he pointed to the heavens.[8]

We do not need to multiply examples. In antiquity, it was relatively uncontroversial to consider the study of the heavens as contributing to the highest form of knowledge, namely philosophy. This was emphatically not the case for the study of animals and plants. While there was an interest in the study of animals and plants before Aristotle, this interest did not result in a systematic study of life. For such a study we have to wait for Aristotle.

The above passage from *Parts of Animals* suggests that Aristotle is self-consciously innovating when he advocates a study of the natural world in all

1977): "The scribe, *as they say*, of nature." From the way Atticus introduces the comparison, it is clear that he is relying on a well-established tradition.

[5] *Parts of Animals* 1.5, 645a4–10, translated by J. G. Lennox, *Aristotle: On the Parts of Animals I–IV* (Oxford: Oxford University Press, 2001).

[6] *Parts of Animals* 1.5, 654a4.

[7] Diogenes Laertius, 2.10, translated by R. D. Hicks, *Diogenes Laertius: Lives of Eminent Philosophers*, 2 vols (Loeb Classical Library; Cambridge, MA: Harvard University Press, 1925). Cf. Aristotle, *Eudemian Ethics* 1.5, 1216a12–14: "asked to what end one should choose to live, Anaxagoras replied: 'to study the heavens and the order of the whole cosmos.'"

[8] Diogenes Laertius, 2.7, translated by Hicks.

its parts, animals and plants included. His innovation is twofold. First, he is transforming the pre-existing interests in the study of animals and plants into a systematic study of life. Second, he is making this study contribute directly and immediately to philosophy. The philosophy Aristotle has in mind is natural philosophy (also known as physics). It is to be distinguished from what Aristotle calls first philosophy (we call it metaphysics). By his lights, metaphysics, physics, and mathematics are the three highest forms of knowledge. He refers to them as three kinds of theoretical knowledge. We seek them for their own sake or, to put it in a slightly different way, for the sake of *theoria*.

The rest of this chapter will introduce the reader to the second of the three highest forms of knowledge, namely physics.

ARISTOTLE'S PHYSICS: A FIRST IMPRESSION

When we speak of Aristotle's physics, we may refer to any one of the following three different groups of writings:[9]

1. The eight relatively independent and self-contained essays that make up the *Physics* (*Physics* 1–8).
2. *Physics* 1–8, *On the Heaven* 1–4, *On Generation and Corruption* 1–2, and *Meteorology* 1–4 (eighteen relatively independent and self-contained essays).
3. These eighteen essays together with the biological corpus (*History of Animals* 1–10, *Parts of Animals* 1–4, *Generation of Animals* 1–5, *Progression of Animals, Movement of Animals*, and the group of short essays known as *Parva Naturalia*[10]).

It is important to stress that the ambiguity highlighted above is not created, let alone encouraged, by Aristotle. He uses expressions such as "in the books on physics" (*en tois physikois* or *en tois peri physeôs*) to refer to ideas or doctrines found in any one of the four great physical works, namely *Physics, On the Heaven, On Generation and Corruption*, and *Meteorology.* Furthermore, he does not use these expressions in such a way that they refer to these four works to the exclusion of the so-called biological corpus. Quite the contrary. Aristotle speaks of what we call biology as physical science (*physikê epistêmê*). We prefer to speak of "natural science." But it should not be overlooked that this preference has to do with our conception of the

[9] On the ambiguity of the expression "Aristotle's physics," see J. Brunschwig, "Qu'est-ce que la 'physique' d'Aristote," in F. De Gand and P. Souffrin (eds.), *La Physique d'Aristote et les conditions d'une science de la nature* (Paris: Vrin, 1991), pp. 11–40.

[10] The title *Parva Naturalia* collectively refers to the following studies: *On Sense-perception; On Memory; On Sleep; On Dreams; On Divination in Sleep; On Length and Shortness of Life*; and *On Youth and Old Age, Life and Death, and Respiration*. The treatise *On the Soul* provides a theoretical foundation for these studies.

physical. It is because we do not think that life is just a physical phenomenon that we use a word based on the Latin *natura* to translate the Greek *physis*. The distinction between the natural and the physical is not available to Aristotle, who employs a rich concept of *physis.* This concept is so rich that he can use the word *physis* in connection with the study of life. In the first book of the *Parts of Animals*, which is also the official introduction to the study of animals, the reader is constantly reminded that the study of animals is part of a larger explanatory project. Aristotle describes this project as a theoretical science concerned with *physis* (*Parts of Animals* 1.1, 640a2 and 641b11). He also speaks of the study of animals as an investigation concerned with *physis* (*methodos peri physeôs; Parts of Animals* 1.4, 644b16). Of course, Aristotle could have coined a special word for the study of animals. He certainly has the conceptual resources to do so. At the very end of his short essay *On the Length and Shortness of Life*, Aristotle speaks of the investigation of animals in such a way that leaves no doubt that this investigation is regarded by him as a discrete and relatively self-contained investigation:

> It remains to investigate youth and old age, life and old age. Once this is done, the investigation of animals [*methodos peri tôn zôôn*] is concluded."[11]

And yet, in spite of the unity and distinctiveness that he finds in the study of animals, in the first book of the *Parts of Animals*, Aristotle takes pain *not* to give the impression that the study of animals is a science (*epistêmê*) in its own terms. This approach to the study of animals is far from being unique to this book. Consider how Aristotle introduces his investigation of animal locomotion in the *On the Progression of the Animals*:

> That the facts are these is clear from our natural history [*physikê historia*] and we have now to examine their causes. We must begin our inquiry by assuming the starting point that we are frequently accustomed to employ in the natural investigation [*pros tên methodon tên physikên*].[12]

This passage is interesting for at least two reasons. First, the opening sentence refers to a fact-establishing investigation (*historia*). We really do not know whether this is a reference to the *History of Animals* or to an equivalent investigation. What matters is that Aristotle refers to it as *physikê historia*.[13] Second, Aristotle speaks of what he is doing, which I take to be the

[11] *On the Length and Shortness of Life* 467b5–9 (D. W. Ross (ed.), *Aristotle: Parva naturalia* (Oxford: Clarendon Press, 1955). I owe this reference to J. G. Lennox, "The Place of Zoology in Aristotle's Natural Philosophy," in R. W. Sharples (ed.), *Philosophy and the Sciences in Antiquity* (Keeling Series in Ancient Philosophy; Aldershot: Ashgate Publishing 2005), pp. 55–71.

[12] *On the Progression of Animals* 2, 704a11–3, 704a12 (W. Jaeger (ed.), *Aristotelis De animalium motione et De animalium incessu. Ps-Aristotelis De spiritu libellus* (Leipzig: Teubner, 1913); my translation.

[13] This is not a unique phenomenon. For a reference to a fact-establishing investigation with the expression *physikê historia*, see *Parts of Animals* 2.3, 650a31. "Physical history" would have been an equally acceptable translation, since Aristotle is innocent with respect to the distinction between natural and physical.

combination of establishing the facts and their explanations, as a single investigation (a *physikê methodos*). Aristotle talks about what he is doing in the singular.[14] As a matter of fact, he never speaks, either in *On the Progression of Animals* or elsewhere, of what he is doing in the plural (*physikai epistêmai* or *physikai methodoi*). What is at stake is the unity and integrity of Aristotle's explanatory project.

What looks like a merely terminological dispute on how best to translate the noun *physis* and the adjective *physikê* has important consequences for how we conceive of Aristotle's explanatory project. It is time to look at this project in some detail.

GENERAL AND SPECIAL PHYSICS

The eight investigations that the tradition has transmitted under the title *Physics* play a prominent role in the project that Aristotle calls *physikê epistêmê* or *physikê philosophia*. We can think of these investigations as "general physics" because they provide an account of nature (*physis*) and motion (*kinêsis*) that applies throughout the physical world. The contents of the *Physics* supply the conceptual framework, which is to say the explanatory starting points (*archai*), for the subsequent investigations. We can refer to these other investigations as "special physics." Let us return to the study of animals. Since the study of animals involves applying the conceptual resources developed in the *Physics* to a narrow range of cases, we can think of it as special physics. Admittedly, general physics does not provide all the starting points for all the special investigations. A few additional starting points are needed to engage in an optimal study of animal life.[15] It is nevertheless clear that this study is shaped, and indeed controlled, by some of the results achieved in the study of general physics. Here suffice it to recall the foundational role that the second book of the *Physics* plays for the study of life. It is in this book that Aristotle introduces the concepts of form and matter, his theory of the four causes, and his general defense of teleology. It is simply impossible to study life, and in particular animal life, without having mastered *Physics* 2.

It is important not to mistake this picture of Aristotle's physics, which emphasizes the integration of general and special investigations and at the same time stresses the foundational role for the entire project played by the general investigations collected in the *Physics*, with an alternative one. This

[14] A point first made in J. G. Lennox, "Aristotle's Natural Science: The Many and the One," *From Inquiry to Demonstrative Knowledge: New Essays on Aristotle's Posterior Analytics: Special Issue of Apeiron* 43 (2010), ed. J. H. Lesher, pp. 1–23.

[15] For more on this point, I refer the reader to Lennox, "Aristotle's Natural Science: The Many and the One." The overriding goal of the first book of the *Parts of Animals* is to introduce standards of investigation that are specific to the study of animals.

other picture of Aristotle's physics can be extracted from an immensely influential paper by G. E. L. Owen.[16] The point of departure of this article is an account of Aristotle's use of the Greek term "*phainomena*." Owen famously argued that the *phainomena* invoked in the *Physics* are reputable opinions, whereas the *phainomena* that are the starting points in the special investigations are empirical data. An immediate consequence of this interpretation is the distinction between those areas of Aristotle's physics that are dialectical and as such are the province of "philosophy" and those that are empirical and as such are the province of "science."[17] The division of Aristotle's physics into general and special investigations does not depend on accepting the distinction between dialectical and empirical investigations. Even when Aristotle is engaged in a general investigation, he is engaged in an *empirical* investigation. Put differently, Aristotle's physics is empirical through and through.

It is important to try to understand why Aristotle thinks that *general* empirical investigations are needed in addition to *special* empirical investigations. To make progress on this front, it may be helpful to introduce a model for how to think about general physics. The model is given to us in the *Posterior Analytics* 1.4–5. There, Aristotle tells us that we have proper knowledge of the fact that the sum of a triangle's internal angles is equal to two right angles only if we know that this property belongs to every triangle *as* a triangle. Since this geometrical property belongs to all triangles, it belongs to equilateral, isosceles, and scalene triangles. But it does not belong to these triangles in virtue of the fact that they are equilateral, isosceles, or scalene. Rather, it belongs to them because they are triangles. Aristotle employs this example to show that there is a common explanatory level beyond that of equilateral, isosceles, and scalene triangles, and that we reach this common explanatory level by ignoring those facts that are specific to equilateral, isosceles, and scalene triangles. More precisely, we ignore those facts that make these triangles either equilateral, or isosceles, or scalene by treating them *as* triangles. Perhaps we can rephrase this last point by saying that Aristotle envisions a general treatment of triangles next to a special study of equilateral, isosceles, and scalene triangles. What is important, and must not be overlooked, is that the general treatment of triangles is not about triangles other than scalene and equilateral. On the contrary, those triangles *are* the object of this general treatment, except

[16] G. E. L. Owen, "*Tithenai ta Phainomena*," in S. Mansion (ed.), *Aristote et les problèmes de méthode* (Louvain and Paris: Publications Universitaires de Louvain and Éditions Béatrice Nauwelaerts, 1961), pp. 83–103 (the article is reprinted in G. E. L. Owen, *Logic, Science, and Dialectic: Collected Papers in Greek Philosophy* (Ithaca, NY: Cornell University Press, 1986), pp. 239–51).

[17] For an attempt to apply the results reached in this paper to Aristotle's philosophy as a whole, see T. Irwin, *Aristotle's First Principles* (Oxford: Oxford University Press, 1990). For an attempt to take issue with Owen's reading of Aristotle's physics, see R. Bolton, "Aristotle's Method in Natural Science: Physics I," in L. Judson (ed.), *Aristotle's Physics: A Collection of Essays* (Oxford: Oxford University Press, 1991), pp. 1–29.

that they are considered *in abstraction* from those properties that make them scalene and equilateral triangles.

We can use this geometrical example to shed some light on the nature of what we have called general physics. When we are doing general physics, we are studying concrete physical objects, except for the fact that we are studying them in abstraction from their specific properties. In other words, general physics is at the same time a *general* and an *empirical* investigation. It remains causally connected with the concrete physical objects. Furthermore, the geometrical example helps us see that general physics is also an *autonomous* investigation. It is autonomous in the sense that the results achieved in the context of general physics do not depend for their confirmation on anything that may be achieved at the level of special physics. Consider how the explanation of the property of having internal angles equal to the sum of two right angles is secured by Aristotle. It is secured without exploiting anything specific or peculiar to either equilateral, isosceles, or scalene triangles. To put it differently, their being equilateral, isosceles, or scalene does not contribute to our understanding of why these triangles have the internal angles equal to the sum of two right angles.

The scientific autonomy of general physics allows for a selective reading of Aristotle's physics. It is possible to read the *Physics* to the exclusion of the rest of Aristotle's physical writings. A reading of Aristotle's *Physics* does not do justice to the integrity of Aristotle's project. Such a reading is nevertheless permitted by how Aristotle conceived of his project of investigation. In connection with this last point, it should be noted that a study of Aristotle's physics that concentrates on general physics, or on a combination of general and special physics, has often prevailed in the subsequent tradition. Consider, for instance, how Aristotle's physics was studied in late antiquity. While this physics was an important part of the philosophical curriculum, its study was limited to *Physics*, *On the Heaven*, *On Generation and Corruption*, and *Meteorology*. It is very telling that no philosophical commentary was ever written on the biological writings. These writings remained outside the philosophical curriculum, as they were regarded as merely technical works. But it would be a mistake to think that this selective approach to Aristotle's physical writings was a late development motivated by a set of concerns extraneous to Aristotle's physics. With a wholly different set of concerns, a similar attitude toward Aristotle's biological corpus is found in Alexander of Aphrodisias (second and third century CE). He too did not write commentaries on Aristotle's study of animals. For an approach that insists on the integrity of Aristotle's physics we have to look beyond antiquity. Averroes (twelfth century) is the most obvious example of an interpreter who insists on the empirical nature of Aristotle's physics and at the same time on its integrity.[18] In the Latin world, Zabarella (sixteenth century) is the champion of the same approach.[19]

CELESTIAL AND SUBLUNARY PHYSICS

Special physics consists of a set of empirical investigations dealing with the different parts or aspects of the physical world on the basis of the conceptual framework developed in the context of general physics. In this overview, it is not possible to offer a comprehensive account of all the empirical investigations either promoted or developed by Aristotle. It is sufficient to concentrate on what is arguably the most controversial revision of the previous tradition of investigation, namely the division of the physical world into a celestial and a sublunary part. This division has important implications for how Aristotle conceived of his whole project. They are especially clear in the case of the special empirical investigation that we call meteorology.

In the opening lines of his *Meteorology*, Aristotle gives us an outline of his whole research program with an emphasis on how the empirical investigation he is about to launch fits within his larger project. Here is how he introduces this investigation:

> There remains for consideration a part of this inquiry which all earlier thinkers called meteorology [*meteôrologia*]. It is concerned with events that are natural, though their order is less perfect than that of the first element of bodies.[20]

In the subsequent lines, Aristotle lists a few phenomena to be explained, including comets, earthquakes, and winds. To the list we can add rain, snow, rivers, sea, metals, and minerals. At least at first sight, this looks like a random list of phenomena. Among other things, it is not obvious why these phenomena are to be studied by a single investigation. In the *Meteorology*, Aristotle is able to secure explanatory, which is to say causal, unity by showing that all these phenomena can be explained in terms of a material cause (a dry and a moist exhalation) together with an efficient

[18] His commentaries on Aristotle's physical writings can be found in Latin translation in volumes 4–6 of the *Juntine Edition of Aristotle and Averroes* (Venice: Iunctas, 1562; reprinted in Frankfurt am Main: Minerva, 1962).

[19] His essay on the constitution and parts of Aristotle's physics (*De constitutione scientiae naturalis liber*) is a manifesto for a reading of Aristotle's physics that insists on the integrity and unity of the explanatory project pursued by Aristotle. This essay is the opening one in Jacopo Zabarella, *De rebus naturalibus libri xxx* (Frankfurt am Main: L. Zetzner, 1607; reprinted in Frankfurt am Main: Minerva, 1966).

[20] *Meteor*. 338a25–b21, translated by H. D .P. Lee, *Aristotle: Meteorologica* (Loeb Classical Library; Cambrdige, MA: Harvard University Press, 1962), with modifications.

cause (the motion of the celestial bodies).[21] Aristotle is not content to show that there is a single discipline that is concerned with all these phenomena. By distinguishing these phenomena from the celestial ones, Aristotle is also proposing a radical revision of the previous investigation of nature. To say the least, Aristotle is not forthcoming about this revision in the opening lines of the *Meteorology.* In the passage quoted above, Aristotle is content to say that he is about to engage in what his predecessors called meteorology (*meteôrologia*). In fact, what they called *meteôrologia* was emphatically not Aristotle's meteorology. Physics before Aristotle was impervious to the distinction between meteorology and celestial physics. It is telling that the Presocratic study of *ta meteôra* is the study of all the things in the sky, including the celestial ones. Moreover, the epithet *meteôrologos* could be used to refer to someone who has an interest in the study of the whole physical word. For instance, Hippodemus of Miletus is remembered as a *meteôrologos*.[22] References to the study of *ta meteôra* are fairly common in the medical tradition. For instance, the author of *Ancient Medicine* speaks of a study of celestial and subterranean things with the expression *ta meteôra ê tôn upo gên*, and the author of *Airs, Waters, and Places* refers to the study of the seasons rising and settings of the stars as accounts of *ta meteôra* (or *meteôrologa*). Many more examples of the use of expressions such as *ta meteôra, meteôrologia*, and *meteôrologos* could be offered.[23] To make a long story short, we can say that the term *meteôrologos* could be employed to refer to those whom Aristotle would later call *physikoi* or *physiologoi*. In other words, meteorology as a branch of sublunary physics is an Aristotelian innovation. Moreover, it is an innovation that crucially depends on Aristotle's division of the natural world into a celestial and a sublunary part.

When we say that for Aristotle the natural world consists of a celestial and a sublunary part we do not do full justice to his position. In antiquity, it was common to think of the celestial world as a somehow special region. Stability and incorruptibility were often postulated as the differentiating features of the celestial world. Aristotle's position is, however, stronger than a generic commitment to the incorruptibility and stability of the celestial world. His position is that there is a material discontinuity between the celestial and the sublunary regions of the natural world. Aristotle is famously

[21] For an excellent study of the explanatory strategy employed by Aristotle to secure the unity of project offered in *Meteorology* 1–3, I refer the reader to M. Wilson, *Structure and Method in Aristotle's Meteorologica. A More Disorderly Nature* (Cambridge: Cambridge University Press, 2013).

[22] H. Diels and W. Kranz (eds.), *Die Fragmente der Vorsokratiker*, 3 vols (Berlin: Weidmann, 1966), vol. 3, 39.

[23] In some cases, these expressions are used with a negative connotation. In Plato's *Phaedrus*, for example, we are told that Pericles learnt from Anaxagoras "high speculations about [what is high in] nature (269c–272b, *meteorologia physeôs peri*). In other words, Pericles learnt from Anaxagoras speculations about what is high in nature, namely speculations about *ta meteôra*. But these speculations about *ta meteôra* were high-flown speculations of little or no use in life. For a study of the use of the expressions *ta meteôra, meteôrologia*, and *meteôrologos* before and after Aristotle, see W. Capelle, "*Meteôros – Meteôrologia*," *Philologus* 71 (1912), 414–48.

committed to the existence of a special simple body that naturally moves in a circle. This body is not just different from earth, water, air, and fire; at least for Aristotle, it cannot be reduced to earth, water, air, or fire. In fact, the expressions "fifth body," "fifth element," or "fifth substance," which are never used by Aristotle, are themselves an indication of how the celestial simple body was received in antiquity. This body was regarded as an additional substance alongside earth, water, air, and fire whose theoretical necessity was very dubious at best.[24]

But why does Aristotle posit the existence of a special element in addition to earth, water, air, and fire? This is a difficult question. It is best answered by reflecting on a picture of the physical world that is exactly like the Aristotelian one except for omitting the claim the heavens are made of the special simple body. In other words, let us suppose that the world is divided into a celestial and a sublunary part, but that the heavens are not made of a special simple body; instead, they are made of fire. In this case, the mobility of the heavens would be explained by invoking a physical property of fire. For instance, it is an empirical fact that fire moves up. Why not think that this fire keeps moving once it has reached the extremity of the world, but now in a circle rather than a straight line? If we take this suggestion seriously, we may conclude that we do not need to posit the existence of a body that moves in a circle and is not reducible to earth, water, air, and fire. Of course, Aristotle would have rejected this conclusion, but why? We should recall that for Aristotle earth, water, air, and fire are perishable in that they can change into one another. There is no reason to think that the fire that has moved up has lost its capacity to change into the other elements. In other words, even if it is successfully removed from the cycle of generation and corruption, this fire still retains a capacity for changing into the other elements. But it is clear that Aristotle would have considered this capacity for change (even if it is never fulfilled) a potential threat to the eternity of the heavens. In all probability, at least for Aristotle, the only way to secure a sufficiently robust version of the thesis that the heavens are eternal involves positing the existence of a celestial simple body lacking the capacity to change into the sublunary simple bodies.

We have reached a major commitment in Aristotle's physics. For Aristotle, the world is eternal in the sense that it is not subject to generation or destruction. In *On the Heaven*, he is quite emphatic that this view sets his physics apart from all the previous investigations of nature. See, in

[24] In addition to "fifth body," "fifth element," and "fifth substance," it was not unusual to refer to Aristotle's celestial simple body as *aithêr*. But Aristotle is very reluctant to use this word. In the *On the Heaven*, for example, he employs it only as the traditional name for the heavens. It is unfortunate that Aristotle's reticence to use this word is not fully appreciated. The fact that Aristotle avoids using it is often obscured, if not even denied, by routinely referring to Aristotle's celestial simple body as *aithêr*. For more on the language employed by Aristotle in connection with the celestial simple body, I refer to A. Falcon, *Aristotle and the Science of Nature: Unity without Uniformity* (Cambridge: Cambridge University Press, 2005), pp. 113–21.

particular, how he introduces the question in *On the Heaven*: "now, all (*pantes*) claim that the universe is generated."[25]

ARISTOTLE'S SYSTEMATIC STUDY OF LIFE

A study of the reasons that may have led Aristotle to introduce a special simple body in addition to earth, water, air, and fire helps us to better appreciate how unusual Aristotle's conception of the physical world was. The strong division of the physical world into a celestial and a sublunary region envisioned by Aristotle was not only exceptional but also anomalous even by ancient standards. Although Aristotle argues against all his predecessors that the celestial world is radically different from the sublunary world, he is not envisioning two disconnected or only loosely connected worlds, namely the heavens and the sublunary world. His view is that there is one and only one world, the physical world; however, at least for Aristotle, this world exhibits unity without uniformity. This view has enormous consequences for his whole research program. For one thing, Aristotle is extremely reluctant to engage in an investigation of the celestial world when and where the lack of empirical data at his disposal cannot be overcome by an appeal to the similarities between celestial and sublunary natures.[26]

Let us return to the exhortation to the study of the physical world in all its parts that can be read at the end of the first book of the *Parts of Animals*. We have already quoted the passage from this exhortation that marks the transition from the study of the heavens to the study of animals.[27] In this passage, Aristotle takes it for granted that the physical world is constituted by a celestial and a sublunary part, and argues that the study of each of these two parts has its own appeal. In particular, the study of animals gives us opportunities for knowledge that are not available in the study of the celestial world. But it would be a mistake to think that Aristotle was interested in the study of animals to the exclusion of plants. The research program outlined in the opening lines of the *Meteorology* makes it crystal clear that the project of investigation envisioned by Aristotle is not complete until a study of plants is also completed:

> After we have dealt with all these subjects let us then see if we can give some account, on the lines we have laid down, of animals and plants, both in general and separately; for when we have done this we may

[25] *On the Heaven* 1.10, 279b12–13.
[26] The epistemological consequences of this position are explored in Falcon, *Aristotle and the Science of Nature*, pp. 85–112.
[27] See the Introduction to this chapter.

perhaps claim that the whole investigation which we set before ourselves at the outset has been completed.[28]

Although Aristotle was committed to a study of animals and plants, he has left us only a study of animals. His references to plants are frequent, but they are mostly references to a study to come.[29] How is this phenomenon to be explained? One possibility is that Aristotle felt that plants and animals did not form a single investigative domain. On this scenario, Aristotle was not able to find explanatory starting points (*archai*) which could justify a common treatment of animal and plant life. But there is another possibility. Aristotle may have consciously postponed the study of plants until the conceptual apparatus for the study of animals was firmly in place. This apparatus would have been eventually adopted, and indeed adapted, for the study of plants.[30] This order of study – *first* animals and *then* plants – is reflected in the passage from the opening lines of the *Meteorology* quoted above: once the study of meteorological phenomena is in place, what remains to do is to study "animals and plants, both in general and separately."

The question of how Aristotle would have pursued his project of a systematic study of life, "animals and plants, both in general and separately," is a fascinating one.[31] Here, it suffices to say that such a study would not be exhausted by separate studies of animals and plants. In addition, Aristotle is expected to offer a general study of those aspects of life that are common to both animals and plants. The best-known example of this sort of study is given in the second book of *On the Soul* in connection with Aristotle's treatment of the nutritive capacity of the soul. Since nutrition is common to both animals and plants, an account which maps onto animals as well as plants is possible. One may ask why Aristotle should be expected to offer a general study of certain aspects of life in addition to separate studies of animals and plants. This question can be answered by invoking the methodological principle that explanation must be given at the proper level of

[28] *Meteor*. 339a5–8, translated by Lee with modifications.

[29] Complete list of references in F. Wimmer, *Phytologiae Aristotelicae Fragmenta* (Breslau: Grass, Barth, and Comp., 1838). The tradition knows of a treatise in two books on plants (the *On Plants*). But this treatise is not by Aristotle. It goes back to a compendium written by Nicolaus of Damascus (first century BCE).

[30] This is what Theophrastus seems to have done in his study of plants (*History of Plants* and *On the Causes of Plants*). For a study of how the conceptual apparatus employed in the first book of the *History of Animals* is adapted to the study of plants in the first book of the *History of Plants*, see A. Gotthelf, "*Historia* I: *plantarum* and *animalium*," in W. W. Fortenbaugh and R. W. Sharples (eds.), *Theophrastean Studies* (New Brunswick (NJ): Transaction Publishers, 1998), vol. 3, 100–35 (reprinted in A. Gotthelf, *Teleology, First Principles, and Scientific Method in Aristotle's Biology* (Oxford: Oxford University Press, 2012), pp. 307–42). A complete study of the explanatory strategies employed by Theophrastus in his study of plants is a desideratum of the literature.

[31] I explored it in A. Falcon, "Aristotle and the Study of Animals and Plants," in K.-D. Fisher and B. Holmes (eds.), *Frontiers of Ancient Science. Essays in Honor of Heinrich von Staden* (Berlin and New York: De Gruyter 2014), pp. 75–91.

generality. Recall the geometrical example given in *Posterior Analytics* 1.4–5. For Aristotle, to explain why the property of having internal angles equal to the sum of two right angles belongs to *isosceles* triangles is not an adequate explanation. Why? Because the property in question belongs not only to isosceles triangles but also to scalene and equilateral triangles. Hence, the only way to arrive at an adequate explanation of this geometrical property is to engage in a common study of triangles. We have already seen that this methodological insight motivates Aristotle to engage in the project we have called general physics. The same insight motivates him to develop a general study of life next to a separate study of animals and plants.

CONCLUSION

When Aristotle was engaged in a search for the causes of physical phenomena, he was working within a well-established tradition. Hence, it is not really surprising to see that Aristotle makes a self-conscious effort to make contact with this tradition (most notably, in *Physics* 1). But this effort does not mean that he is simply doing, although in greater detail and in a more systematic way, what was done by his predecessors. In his physical writings, he presents us with a new style of physics. What is truly remarkable is not simply the scope or ambition of his research program but also the nature of his explanatory project. First, the division of Aristotle's physics into a general and a special component is an Aristotelian innovation. It depends on the constraints that Aristotle builds on his theory of explanation as presented in the *Posterior Analytics*. Second, the prominent role that the systematic study of life plays in Aristotle's physics should not be underestimated. Such a study of life is another Aristotelian innovation. Third, the strong division that Aristotle introduces between the celestial and sublunary world is also unprecedented. This division, which is a result of a consequence of his view that the heavens are made of a special simple body, is needed to secure a strong version of the view that the physical world is eternal in the sense that it has no beginning and no end in time. Here too Aristotle is departing from the previous tradition of investigation

Even if Aristotle does not always insist on the novelty of his physics, there is no doubt that there was nothing like that before Aristotle. But what happened after Aristotle? This is a large question which cannot be fully addressed in the context of this overview, though it can be approached with the help of the following two quotes:

> Aristotle's physics is one of the most astonishing systems human reason has ever built; it gave answers to all the questions the ancients had about the heavens and their motions, the elements and their transformations, the most precise and complete answers offered up

> until then, and all these answers were logically organized in a theory compared to which all prior doctrines seemed to be mere beginnings. That such a system exercised on minds the powerful seduction that most of the Arabic or Christian philosophers experienced in the Middle Ages, is easy to understand. In contrast, it is surprising to learn that the immediate successors of Aristotle proved themselves to be, in general, rebellious to this influence.[32]

> In antiquity Aristotle's physics enjoyed little influence outside the Aristotelian school and was hardly known except in the Lyceum. It came to be enormously influential during the Middle Ages.[33]

The main message that both quotes convey is that when it comes to the reception of Aristotle's physics, there are two stories to be told. The first is a story of great success and enormous influence. It is the story of the reception of Aristotle's physics in the Middle Ages (and beyond). The second is a story of mixed success at best. It is the story of the reception of Aristotle's physics in antiquity. It is only in the Middle Ages that Aristotle achieved the status that made Dante refer to him as the teacher of those who know ("*il maestro di color che sanno*"). In antiquity his physics was often resisted if not openly criticized. Even when it enjoyed a revival of interest in late antiquity, the interest was limited to certain parts of Aristotle's project. We have seen, in particular, that the biological corpus remained outside the curriculum of study of late antiquity. This is no small thing if we consider that about half of the extant physical corpus is concerned with topics related to the study of life.[34]

[32] P. Duhem, *Le Système du Monde* (Paris: Hermann, 1916), vol. 1, 242, my translation.

[33] J. H. Randall, *Aristotle* (New York: Columbia University Press, 1960), pp. 165–6.

[34] The rediscovery and re-appropriation of Aristotle's biology was a gradual phenomenon that began in the Arabic world and continued in the Latin world. For an attempt to reflect on the transmission of Aristotle's biology in the Greek and Arabic traditions, see Cristina Cerami and Andrea Falcon, "Continuity and Discontinuity in the Greek and Arabic Reception of Aristotle's Study of Animals," *Philosophia antiquorum* 8 (2014): 35–56. For an introduction to the complexity of the reception of Aristotle's biology in the Latin world, see A. Falcon. "The Reception of Aristotle's Study of Animal Motion in the Latin World," *Documenti e studi sulla tradizione filosofica medievale*," 23 (2012), 521–39.

11

ARISTOTLE'S PHYSICAL THEORY

Eric Lewis

There are at least two unavoidable difficulties with any attempt to come to terms with Aristotelian physical thought. The first is the degree to which the physical thought is inseparable from the metaphysical thought. No discussion of Aristotelian physical thought can avoid discussion of form, matter, actuality, potentiality, substance, and the like.[1] The second is that Aristotle characteristically begins his scientific inquiries with an examination of his predecessors, culling from their views what he thinks is correct, and correcting, chastising, and modifying their positions.[2] One cannot therefore understand the motivations behind, and often the content of, much of Aristotelian physical theory without understanding, for example, what Aristotle takes to be the substance of, and problems with, Presocratic atomic theory, or Eleatic logic. These problems make it difficult to know where to begin an examination of Aristotle's physical thought. What follows is intended to give the

[1] On Aristotle's philosophy of science and scientific method see: R. J. Hankinson, "Philosophy of Science," in Jonathan Barnes (ed.), *The Cambridge Companion to Aristotle* (Cambridge: Cambridge University Press, 1995), pp. 109–39, and the extensive bibliography found at the end of this work.

[2] For a classic study of Aristotle's reception of his predecessors see: Harold Cherniss, *Aristotle's Criticism of Presocratic Philosophy* (Baltimore, MD: Johns Hopkins University Press, 1935); Harold Cherniss, *Aristotle's Criticism of Plato and the Academy* (Baltimore, MD: Johns Hopkins University Press, 1944). For a more recent consideration see: Robert Bolton, "The Origins of Aristotle's Natural Teleology in Physics II," in Mariska Leunissen (ed.), *Aristotle's Physics: A Critical Guide* (Cambridge: Cambridge University Press, 2015), pp. 121–43.

For an overview of Aristotle's *Physics* see: Auguste Mansion, *Introduction à la Physique Aristotélicienne* (2nd edn; Louvain: Institut Supérieur de Philosophie, 1946); W. D. Ross, *Aristotle's Physics. A Revised Text with Introduction and Commentary* (Oxford: Clarendon Press, 1936); Wolfgang Wieland, *Die Aristotelische Physik* (Göttingen: Vandenhoeck & Ruprecht, 1962). For more recent studies in Aristotle's and Aristotelian physics see: Richard Sorabji, *Necessity, Cause and Blame. Perspective on Aristotle's Theory* (London: Duckworth, 1980); Richard Sorabji, *Time, Creation and the Continuum. Theories in Antiquity and the Early Middle Ages* (London: Duckworth, 1983); Richard Sorabji, *Matter, Space and Motion. Theories in Antiquity and their Sequel* (London: Duckworth, 1988); Sarah Waterlow, *Nature, Change, and Agency in Aristotle's Physics. A Philosophical Study* (Oxford: Clarendon Press, 1982). See also the essays in Lindsay Judson (ed.), *Aristotle's Physics: A Collection of Essays* (Oxford: Clarendon Press, 1995); David Bostock (ed.), *Space, Time, Matter, and Form. Essays on Aristotle's Physics*. (Oxford: Clarendon Press, 2006); and Leunissen (ed.), *Aristotle's Physics.*

reader a sufficient understanding of certain essential aspects of Aristotelian physical science to enable them to pursue further examinations of specific issues in the Aristotelian scientific tradition.[3] Accordingly, I will concentrate on the core notions of matter, alteration and generation, space, time, and motion that collectively form the backbone of Aristotelian physical science. A related goal is to demonstrate just how embedded the notion of "potentiality" is in much of Aristotle's physical thought.

For Aristotle, the motions that objects manifest, and, more generally, the changes objects can undergo, are highly dependent on what it is that an object is composed of. Accordingly, I will begin with an examination of Aristotle's theory of matter and the elements. Once we understand his theory of matter, and the nature of his compositional hierarchy, we will turn to Aristotle's (under-examined) theory of weight, since it is the weight of objects which determines, to a great extent, how, and when, they move. This will lead naturally to a discussion of Aristotle's dual theories of natural and "unnatural" (forced, projectile) motion. Since much of Aristotle's theory of motion is developed as a critique of the theory of others, in particular that of the atomists (who endorse the existence of, and motion through, the void), we will next examine his arguments against the possibility of motion through the void, and, more generally, arguments against the possible existence of void.

Aristotle's arguments against the existence of void are bound up with his conception of place, to which we will then turn our attention. Having at this point laid the groundwork for Aristotle's theory of "matter in motion," we will examine two additional concepts: time and the infinite. Time is a necessary prerequisite for the possibility of any sort of change, while problems involving the infinite arise, famously, once one allows for continuous motion. In this sense what follows, although highly selective, attempts to give a fairly unified account of Aristotelian physical science.

A legacy of Presocratic thought, and an essential component of "phusikoi" writing in general, is to account for what the world is composed of. Prominent accounts of this sort that developed in the Presocratic period hypothesized that: all is composed of small changeless bits of stuff, called atoms (the atomists); and all is composed of earth, air, fire, and water (Empedocles), or one of these four (Heraclitus, perhaps fire; Anaximenes, air).[4] Any such theory which has it that there is a privileged set of entities which everything is composed of must face the question as to how one gets from the compositional basics to those things composed out of them. If you

[3] Still perhaps the best general survey of Aristotle's science is: Friedrich Solmsen, *Aristotle's System of the Physical World: A Comparison with His Predecessors* (Ithaca, NY: Cornell University Press, 1960).

[4] It is far from clear what any given author means by claiming that "X is composed of Y." The usual way this is put in Greek is to claim that "X is *ek* Y." This may mean that X is now a form of Y (as ice is a form of water), or simply that Y temporally precedes X, and is (suitably) related to Y's coming to be (as a child is from its mother). Other meanings are possible.

believe that everything is composed of air, how do we get paper? Is this piece of paper now a kind of air? The most obvious alternative to this view (that taken, perhaps controversially, by the atomists and Empedocles) is to deny that anything really exists other than those things which one conceives of as compositionally basic. One might claim that only earth, air, fire, and water exist, or only atoms exist. Any such view of the question of how we get paper, or what paper is, does not arise; there is no paper. Such a position will be called compositional reductionism, since it claims that the only things which really exist are the material components of those things which appear to exist. Aristotle is the first thinker to face these questions openly, and to offer to them a far more sophisticated answer than his predecessors.

The primary motivation behind compositional reductionism is to offer a response to a problem first articulated by Parmenides. Parmenides argued that nothing can come to be from what is not, nor from what is, the former being impossible (no creation ex nihilo), while the latter is inconsistent with coming to be (for the entity at hand already exists).[5] Compositional reductionism is a response to this dilemma if one adds the claim that the compositional simples are themselves uncreated (as are the atomists' atoms, or Empedocles' elements). The cost of such a move is an inability to account for the existence of those things which do seem to exist, that is, the ordinary substance of the world, animals, plants, artifacts, and the like. Solving this problem is the primary motivation behind Aristotle's physics, and it is the focus of the first two books of the *Physics*, where Aristotle presents for the first time his notions of form, matter, privation, substance, and existence in potentiality versus existence in actuality.

Aristotle presents a single solution to this puzzle, one which accounts for both alteration and substantial change. For Aristotle substantial change involves the coming into existence of a new substance. The paradigmatic example of such changes are biological comings-to-be, the birth of an animal, the maturation of a plant from seed. Yet the coming-to-be of new instances of kinds of things also count as substantial changes, for example the creation of gold from the four elements, or the smelting of iron. Alterations are, crudely put, changes in the features of a persisting substance, a person becoming tan from having been pale, some iron ore becoming hot from having been cold.

Aristotle's solution is to develop a retention–replacement model of change. Aristotle agrees with his predecessors that something must remain, or underlie, through a given change, yet there will also be things that are replaced. For non-substantial change, i.e. alteration, this model is simple.

[5] For a recent study of Parmenides' philosophy, with an extensive bibliography, see: John Palmer, *Parmenides and Presocratic Philosophy* (Oxford: Oxford University Press, 2009). For the fragments and English translations see: David Gallop, *Parmenides of Elea. Fragments* (Toronto: University of Toronto Press, 1984).

What remains through the alteration of a given substance just is the substance, while what is gained is a new quality or form, while what is lost is a quality, or privation. To employ the above example, if I become tan from having been pale, I remain through the change, and so play the role of the substrate or *hupokeimenon* of the change, while I lose the privation pale, and gain the form tan. To employ the language of the original dilemma, a tan person can come to be from what is not a tan person, but not from what is not a person.

It is with substantial change that the theory becomes more contentious, and more important for an overall account of Aristotelian science. What is it that remains through the coming-to-be of a substance? Most, but not all, agree that it is that which Aristotle calls matter. When a given substance comes to be, some quantity of matter persists through the change, and takes on a substantial form.

There are at least two difficulties with this account. If matter persists through substantial changes, why is such a change not just an alteration of the persisting matter? In effect, why is matter not substance? To solve this problem Aristotle denies that the material constituents of substances exist "in actuality" in what they compose. They exist, but "in potentiality."

A second problem, the solution to which is hotly contested, involves what persists through changes of what are often taken to be the materially most simple entities for Aristotle, the four elements, earth, air, fire, and water. Is there that which is their matter which can persist through their coming-to-be? This is not simply an academic problem, but is the most fundamental issue that Aristotelian chemistry must face.[6] At the heart of Aristotle's chemistry is the claim, and associated theory, that the four elements can both transform into each other, and collectively blend so as to form more complex materials. The level above the elements in the Aristotelian compositional hierarchy is that of the "homeomers," things parts of which are like the whole. Examples are gold, wood, blood, and flesh, for portions of these are all the same as the whole. Homeomers themselves can be structured so as to form what Aristotle calls anhomeomers, a good example being a nose, or a petal, for parts of noses are not noses, and parts of petals are not petals. Anhomeomers, when suitably unified, form substances. Aristotle devotes book two of *On Generation and Corruption* to a discussion of elemental transformations, the fourth book of the *Meteorology* to the creation of homeomers, the *Parts of Animals* to the coming to be of organic anhomeomers, and the *Generation of Animals* to the coming to be of biological substances. If the theory of elemental transformation and the creation of homeomers is suspect, then the whole of Aristotelian chemistry and biology tumbles.

[6] Eric Lewis, *Alexander of Aphrodisias. On Aristotle Meteorology 4* (London: Duckworth, 1996), pp. 15–59.

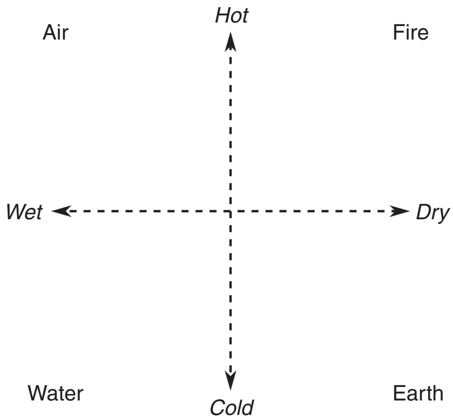

Figure 11.1. Schematic showing the relationship between elements and qualities.

The standard response to this problem, which goes back to the Neoplatonists, is that Aristotle postulated the existence of something called "Prime matter."[7] This prime matter is itself without qualities, and is the ultimate substratum of all material things. It persists as substrate through elemental transformations. This view, codified in the scholastic period, has always had to rely on very scant evidence. Indeed the strongest argument in its favor is that it does offer a solution to the problem of what persists through elemental transformations. Others have claimed that no material substrate persists through substantial changes.[8] A third alternative is available, which I will now sketch.

The four elements are each characterized by a pair of qualities, which are sometimes also called powers or contraries.[9] Consider the above Figure 11.1.

Here the elements are placed between the contraries characteristic of them. Apart from these contraries, the only other properties the elements seem to have are weight, resistance, and a propensity to move unaided to a particular part of the finite and bounded universe. Earth, if unimpeded, will move towards the center of the spherical-shaped cosmos, while fire will move

[7] For the history of prime matter in Neoplatonism see: Sorabji, *Matter, Space and Motion*, chapters 1–3. For modern accounts of Aristotle on prime matter see: Hugh R. King, "Aristotle Without Prima Materia," *Journal of the History of Ideas* 17 (1956), 370–89; William Charlton, *Aristotle. Physics Books I and II* (Oxford: Oxford University Press, 1970), pp. 129–45; C. J. F. Williams, *Aristotle's Generatione et Corruptione* (Oxford: Clarendon Press, 1982), pp. 211–19; Mary Louise Gill, *Aristotle on Substance. The Paradox of Unity* (Princeton, NJ: Princeton University Press, 1989), pp. 41–82, 243–52; Paul Studtmann, "Prime Matter and Extension in Aristotle," *Journal of Philosophical Research* 31 (2006), 171–84; Mary Krizan, "Prime Matter without Extension," *Journal of the History of Philosophy* 54 (2016), 523–46.

[8] Most prominently, Gill, *Aristotle on Substance*.

[9] See Aristotle *On Generation and Corruption*, 2.3.

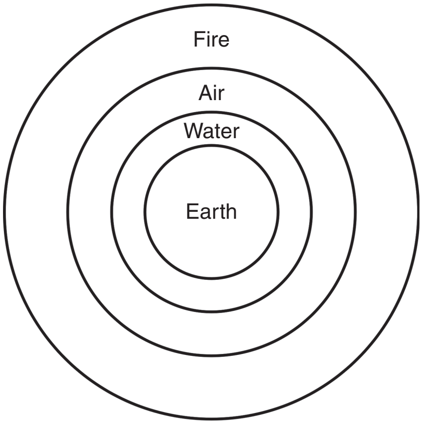

Figure 11.2. Diagram of the cosmos representing the natural positions of the four elements.

towards the perimeter. Water will move so as to be adjacent to earth, and air will move so as to be just below the fire. This would yield, were all the elements to so manifest their kinetic propensities, a cosmos which looked as in Figure 11.2.

Note that each element shares a contrary with that element adjacent to it, and with one other element. This suggests that the contraries play the role of the persistent substrate through such transformations. If these contraries play the role of the persistent substrate through these transformations they would be the matter of the elements. There is evidence that Aristotle endorses this position both in book two of *On Generation and Corruption*, and in the fourth book of the *Meteorology*.[10] The elements are composed out of the contraries, and so the contraries are the most basic or simplest matter in the Aristotelian universe. The propensity of elements to move to their natural place is another, often unrecognized, motivation behind the development of the notion of existence in potentiality and Aristotle's unique account of mixture as found in *On Generation and Corruption*, 1.10. Aristotle has a problem accounting for the coherence of mixtures of the elements. Again, this is a crucial problem for Aristotelian chemistry, since all things are composed out of mixtures of the elements. (In fact, Aristotle claims that all things are mixtures of all four elements; see *On Generation and Corruption*, 1.10.) If flesh is a mixture of all four elements, and these elements have a propensity to move to distinct parts of the cosmos, what accounts for the relative permanence of flesh?

There is evidence that Aristotle was aware of this problem from early on, since he mentions it, without solution, in the early work, *On Philosophy*.[11]

[10] Lewis, *Alexander of Aphrodisias*, pp. 1–60.

[11] Philo, *On the Eternity of the World*, 6.28–7.34=Ross fragment 19b: W. D. Ross, *The Works of Aristotle. Volume XII. Select Fragments* (Oxford: Clarendon Press, 1952) pp. 90–1.

The solution again hinges upon claiming that in mixtures the constituent elements do not exist actually, but only potentially. If they existed actually they would actually move towards their natural places, yielding the dissolution of the compound. So they exist, but potentially. Here we have the chemical motivation behind one of the essential features of Aristotelian metaphysics, the notion of existence in potentiality.

We now can complete the Aristotelian hierarchy of composition. Substances are composed out of anhomeomerous parts: organs and the like. These are composed out of homeomerous parts: flesh and blood, for example. The homeomers are composed out of the four elements, while the four elements have as their matter the four contraries: the hot, cold, moist, and dry. Substances are composed out of anhomeomers by juxtaposing the anhomeomerous parts in a structure that yields a whole with particular capabilities. Bone is next to sinew, for example. Yet this is not the relationship which elements have to each other in homeomers. The four elements are not juxtaposed to each other in, say, some flesh. Neither are they completely coextensive, since Aristotle makes it axiomatic that there cannot be coextensive bodies.[12] They are mixed together, and mixture is for Aristotle a very important topic, which he devoted *Generation and Corruption* 1.10 to discussing. It seems that the possible relationships which constituents in a mixture can have to each other are two: they are juxtaposed, or they are coextensive. Aristotle rules out the first, for he believes that the juxtaposition of elements cannot bring into existence something truly new.[13] If you juxtapose the four elements you cannot create, say, some flesh, that is something with properties and powers new, and other, than those of the four elements. Coextensive constituents are ruled out as simply being impossible. Here again Aristotle relies on his notion of existence in potentiality. In a true mixture the constituents do not exist actually, but do exist potentially. They are neither coextensive nor juxtaposed, and they do not manifest their propensity to move towards their natural places.[14]

Two of the most characteristic aspects of Aristotelian physics are the theories of motion (both natural and projectile) and weight. They form the core of Aristotelian mechanics, that very theory whose overthrow was essential for the development of modern mechanics, and all that follows from it. One must understand the Aristotelian paradigm if one is to understand the subsequent development of physics.

The virtually universally accepted "mistake" that Aristotle is thought to have made with respect to weight is to have claimed that heavier objects fall

[12] See Aristotle *Phyics*, 4.1 (209a4–209a7); 4.6 (213b7); *On the Heavens*, 3.6 (305a19–305a20); *On Generation and Corruption*, 1.5 (321a5–321a10); *On the Soul*, 1.5 (409b3); 2.3 (415a13–415a15). See also Sorabji, *Matter, Space and Motion*, chapter 5.

[13] Aristotle *Meteorology*, 3.2 (327b1–327b2).

[14] Aristotle *On Generation and Corruption*, 1.10 (327b23–327b26).

faster than lighter ones.[15] Indeed it is the correction of this mistaken claim by Galileo that many see as the crucial step towards the development of modern kinematics. More precisely, the mistake is to believe that in free fall (through a vacuum) heavier objects will fall faster than lighter ones. Some have used the fact that Aristotle denies the existence of the void to argue that no such claim can be made on Aristotle's behalf, and that, in any case, free-fall velocities through media are, to some degree, weight dependent.[16] Yet such defenses will only take one so far. If the orthodoxy is correct, Aristotle believed that an object twice as heavy as another one will fall twice as fast. This claim is so blatantly incorrect, and easily observed to be incorrect, that even the most fervent believer in the relativity of observations from within distinct scientific paradigms cannot explain it away. How could Aristotle have thought such a thing?

How certain are we that Aristotle's claims about the relative velocities of objects of differing weight concern free-fall motion (whether through media or the void)? Claims concerning relative velocities constitute how differences in weight manifest themselves for Aristotle. What velocity is Aristotle (or any Greek for that matter) likely to be considering in a discussion of weight – that is, in a discussion of how differences in weight are detected? Surely not free-fall velocity; things were not weighed by flinging them from heights. The Greeks employed balance beams, and the physics of balance beams offers support for Aristotle's claims. If one places progressively heavier objects in a balance beam with a normalized weight in the other pan, the velocity with which the heavier pan will fall down will increase. If it is this velocity which Aristotle had in mind, one can not only make sense of his claims about relative velocity, but one can see why he claims that it is the measure of velocity that is appropriate for the calculation of the weight of objects.[17]

However, much is still mysterious about Aristotle's conception of weight. For us weight is an absolute notion (particularly when it is conflated with mass), while heavy and light are relatives. For Aristotle things are reversed. Heavy and light are absolutes. Anything that moves naturally towards the perimeter of the cosmos is light (as with fire), while anything that naturally falls down towards the center of the cosmos is heavy (as with earth). However, weight is a relative notion. For Aristotle a talent of wood (about 37 kilos) is heavier than a mina of lead (about 431 grams) in air, but not in

[15] Galileo Galilei, *Discorsi e Dimostrazioni Matematiche, Intorno à Due Nuove Scienze Attenenti alla Mecanica & i Movimenti Locali, Leiden* (Leiden: 1638), pp. 62–3. More recently: Denis O'Brien, "Aristotle's Theory of Movement," *Proceedings of the Boston Area Colloquium in Ancient Philosophy* 11 (1995), 47–86.

[16] E.g.: Eric M. Rogers, *Physics for the Inquiring Mind. The Methods, Nature, and Philosophy of Physical Science* (Princeton, NJ: Princeton University Press, 1966), chapter 1.

[17] Eric Lewis, "Commentary on O'Brien," *Proceedings of the Boston Area Colloquium of Ancient Philosophy* 11 (1995), 87–100.

water.[18] Making sense of the relative nature of weight for Aristotle can only be accomplished once one realizes both how measurements of weight are made, and why it is that things have the weight that they do.

The answer to the first question is easy: we weigh things! If one weighs (in a balance beam) a talent of wood and a mina of lead in the air, then the wood is heavier, since the pan containing it falls down. If one conducts this experiment under water, them the mina of lead is heavier, since the pan containing it falls down, as the wood (as long as it is not black ebony) floats to the surface.

It is this (true) phenomenon that Aristotle wishes to explain. Why do compounds have the relative weights that they do? It is not the ratio of elements which compose some body that determines its weight, but the actual amounts of each element present. If one ball of lead is composed of 1 part fire, 1 part air, 1 part water, and 10 parts earth, while another contains 10 parts of fire, air, and water, and 100 parts of earth, the latter ball is 10 times heavier than the first, while according to the ratio test they would both (*per impossibile*) weigh the same. That this is so is demonstrated by weighing the two samples.[19]

The motion of compounds, in so far as they are light or heavy, is a function of their elemental make-up, and so a product of the natural motions of the four elements. Yet what of projectile motion, and what further conditions, if any, does Aristotle place on the possibility of motion?

Projectile motion is a problem for Aristotle since there is nothing about, or internal to, the object being propelled that can account for its motion contrary to nature. When I throw a rock upwards, why does it not immediately fall down (the direction the rock would fall naturally if unimpeded) as soon as it leaves my hand? As Aristotle puts it, things in motion are moved by something in contact with them (if they do not move themselves).[20] What can move a projectile that is in contact with it? Rather mysteriously, Aristotle believes that my hand can impart to the surrounding medium the power to be like my hand, which is to be a mover. The medium retains this power even after my hand stops moving. In other words, the air surrounding the rock can continue pushing the rock after my hand no longer does.[21]

This rather ad hoc explanation of projectile motion comes at the very end of the *Physics*. It was to be rejected, and ridiculed, by many subsequent physicists.[22] Interestingly, Philoponus' alternative theory, which has it that a force can be directly imparted to the moving object by the mover, is clearly a forerunner to modern impetus theory. Indeed the very passage where he

[18] Aristotle *On the Heavens*, 4.4.

[19] On difference test: Aristotle *On the Heavens*, 4.4 (311a30–311a32); on ratio test: Aristotle *On the Heavens*, 4.2 (309b8–309b16).

[20] Aristotle *Physics*, 4.4 (211a24–211b4, 212a29–212a30); 4.5 (212b19).

[21] Aristotle *Physics*, 8.10 (267a20).

[22] Most prominently in antiquity: Philoponus, *On Aristotle Physics* (641,13–642,20).

makes this innovation seems to have influenced (perhaps even directly) Galileo.[23]

A further aspect of Aristotelian dynamics is his denial of the possibility of motion through void.[24] This is part and parcel of his arguments against the existence of void. Melissus had argued that void was a necessary precondition for motion, yet that it did not exist, and so motion was impossible.[25] The Presocratic atomists accepted the need for void if there is to be motion, and so made void, along with atoms, the only truly existing things.[26] Aristotle, as part of his more global attack on atomism, claims that the void would make motion impossible (and, indeed, rest too).

Some of the arguments are famously obscure, yet Aristotle's overall stratagem is clear. He begins by claiming that a body placed in void would have no reason to move one way as opposed to another, since "the void, in so far as it is void, admits no difference," thereby concluding that a body placed in void would rest, rather than move.[27] This argument, in and of itself, does not answer the question as to whether a body, already moving, might continue to move through void. It therefore is perhaps best interpreted as part of the objection to the void being a cause of motion, as the atomists seem to have thought (at least according to Aristotle).[28]

Aristotle goes on to claim that the natural motion of the elements would be impossible if such motion were indexed to particular locations within a void.[29] This objection has two parts. First, if one assumes (with the atomists) that the void is infinite, there could not be preferred directions of motion, for "there will be no up or down or middle." Second, in so far as one is considering motion in a void, there could be no motion towards particular locations within a void, for "there is no difference in what is nothing." Since for Aristotle natural motion is prior to forced motion, if the existence of the void cannot support the existence of natural motion, it cannot support the existence of any motion.

Aristotle goes on to demonstrate that motion through the void is inconsistent with his theory of projectile motion (which requires a medium to become a mover, something which the void, being "a nothing" cannot do).[30] He also claims that nothing moving through a void would ever stop moving "for why should it stop here rather than there?"[31]

[23] Galileo *de Motu* and *il Saggiatore*. There are persistent rumors that a copy of Philoponus' *On Aristotle Physics* exists with marginalia in Galileo's own hand.
[24] Sorabji, *Matter, Space and Motion*, pp. 142–59.
[25] H. Diels and W. Kranz (eds), *Fragmente der Vorsokratiker*, 3 vols. (6th edn; Berlin: Weidmann, 1952) (henceforth "Diels–Kranz"), 30 B7=Simplicius, *On Aristotle Physics* (111,18–112,15).
[26] Diels–Kranz 67 A7=Aristotle *On Generation and Corruption*, 1.8 (324b35–325a26)
[27] Aristotle *Physics*, 4.8
[28] Ibid., 4.8 (215a1–215a14) and (214b12 ff.).
[29] Ibid., 4.8 (215a7 ff.).
[30] Ibid., 4.8 (215a14 ff.).
[31] Ibid., 4.8 (215a20).

In addition, Aristotle raises two objections crucial to the subsequent development of dynamics. First he claims, in effect, that motion with a finite velocity through void is impossible (and so is all motion through the void, since motion with infinite velocity is not, for Aristotle, either possible, nor would it be motion at all), since it would entail that the density of the void stands in a finite ratio to "thicker" media.[32] The argument is powered by the assumption that as a medium becomes progressively thinner, the velocity of a given body through it increases according to the same progression. In other words, half the "thickness" of a medium, and a body will move twice as quickly through it. Now assume that a body moves through a void interval of a given length in time X. There will exist some non-void medium through which the same body will move through the same length in time 2X. Now consider a medium half as thick as this medium. The body should move through it in time X. "But this is impossible."

Aristotle draws another conclusion from these considerations. He seems to think that the manner in which the weight of an object manifests itself is by cleaving a medium – heavier objects cleave a medium more quickly then light objects, and so move more quickly through media. If, therefore, there were to be no medium (that is, a void medium) all objects would move with the same velocity.[33] In other words, Aristotle takes the modern truth that all objects have the same free-fall velocity through the void as a *reductio* of the existence of the void. Why? The argument is not, I think, simply based on thinking that the regularity of such motion is counterintuitive. For on Aristotle's theory the similitude of motion yields the similitude of weight, and Aristotle thinks, more plausibly, that it is absurd to think that all bodies would have the same weight in a void.

Both these arguments assume a faulty relationship between the velocity of a body through a medium, and the density of the medium. It was not until Philoponus that it was recognized that the removal of a resistant medium would not increase velocity to infinity, but merely remove the extra time it takes to traverse an interval. Philoponus, followed by many others, claims that velocity through a void would be finite, and a thickening of the medium merely adds time to such journeys. These arguments were eventually to influence Galileo, who refers directly to Philoponus.[34]

These arguments against the possibility of motion through a void are only part of a more global attack on the existence of the void. Here I wish to consider the arguments against the existence of an extra-cosmic void.[35] Recall that for Aristotle the universe is finite in extent and spherical-

[32] Ibid., 4.8 (215b22–216a4).
[33] Ibid., 4.8 (216a11–216a21).
[34] Philoponus *On Aristotle Physics* (690,34–691,5, 681,17–681,30).
[35] Sorabji, *Matter, Space and Motion*, chapter 8; Edward Grant, *Much Ado about Nothing. Theories of Space and Vacuum from the Middle Ages to the Scientific Revolution* (Cambridge: Cambridge University Press, 1981).

shaped.[36] These features allow his theory of natural motion to be employed. A spherical-shaped cosmos has a center and a perimeter (and is suggested by celestial movements), while a finite cosmos allows the elements actually to reach their natural places. However the postulation of a finite bounded cosmos suggests that the cosmos is itself embedded in, or surrounded by, void. This Aristotle denies. He takes literally the description favored by the atomists of the void being "nothing" and claims that the cosmos, in so far as it is surrounded by nothing, does not have beyond it something void.

Why not? Aristotle first establishes that the cosmos must itself be finite, that, in effect, body must be finite.[37] Only finite body could: rotate; move at all; have an up and a down; be composed out of simple bodies (here the four elements); or act and be acted upon, and so be perceptible. Aristotle goes on to make a crucially distinct argument. Not only must body, or matter taken as a whole, be finite, but a finite piece of matter could not "escape" beyond the limits of the cosmos, and so come to be outside it.[38] Consider some body outside the cosmos. It is there naturally or unnaturally. It cannot be there naturally, for if so, where it is would be its natural place. Yet it has already been established both that there are only four elements, and what their natural places are (this argument crucially relies on the fact that the existence of only four elements, with their four natural places, has already been established). Since none of the four elements have their natural place located beyond the perimeter of the cosmos, this body must be beyond the cosmos unnaturally. If it is there unnaturally, then the place where it is is unnatural for it. So, this place must be the natural place for something else (given that all places are natural for something or other). But there is not anything for it to be the natural place of (since the natural place of all things has already been accounted for), so, the object also cannot be there unnaturally. If it cannot be there naturally or unnaturally, it cannot be there at all.

Aristotle goes on to claim, in an argument found wanting by many, that since it is impossible for body to be received outside the cosmos, and since both void and place are defined as that in which there can come to be body, there is neither void nor place outside the cosmos.[39] Many think that this argument is faulty, since it seems to move from the incapacity of matter to become extra-cosmic to there not being anything extra-cosmic that could so receive it.

Some resolve this difficulty by assuming that Aristotle is here relying on the modal principle that there cannot be any truly eternally unactualized

[36] See especially: Aristotle *On the Heavens*, 1.5.
[37] Ibid., 1.2–6.
[38] Ibid., 1.9 (278b22–279a7).
[39] Ibid., 1.9 (279a11–279a18). In antiquity, Cleomedes was especially dissatisfied with Aristotle's argument: Cleomedes, *The Heavens*, 1.1.81 ff.

possibilia.[40] On this account if place or void is what can receive body, and no body will ever so be received, then there is no place or void in fact present. Alternatively, some think that Aristotle has confused lack of opportunity with the lack of capacity. There could be extra-cosmic void, so they claim, which has the capacity to receive body, but will never have the opportunity to do so.[41]

I believe that both camps have failed to realize the true nature of Aristotle's argument here against extra-cosmic place. He argues (as I just sketched above), in effect, that there is no extra-cosmic place for this object to be, since there can be neither any extra-cosmic natural place, nor unnatural place. There is, therefore, no extra-cosmic place at all.[42] This argument assumes, as many do in these chapters of *On the Heavens*, Aristotle's theory of the four elements and the associated theory of natural (and so unnatural) place. As such the argument is a good one.

In the passages just examined Aristotle considers place and void in the same breath. But he famously had a theory of place itself thought to be both odd and problematic.[43] Aristotle devotes the fourth book of the *Physics* to a discussion of place. Here he rejects assorted definitions of place, concluding that place is the first unmoved surface surrounding a body.[44] This problematic definition becomes standard in the Peripatetic and scholastic tradition. The rejected candidates include conceiving of place as form, matter, or the extension between the limits of a body. It is the rejection of this third option that is most curious, for two reasons. First, the rejected candidate seems to fit best what we might want to take the place of an object to be, namely that bit of space that an object occupies and is coextensive with. Second, this conception of place as a three-dimensional extension which a body occupies seems to be the very definition of place which Aristotle does accept in the *Categories*.[45] Why reject it in the *Physics*, and what accounts for the seeming change of heart?

There are two camps concerning the relationship of the *Categories*' definition of place to that of the *Physics*. The first camp claims that the differences are merely seeming, that the two theories are in fact the same. So in antiquity Ammonius, Simplicius, and Philoponus all read the *Categories*' definition as being compatible with the *Physics*' definition, or indeed as endorsing it.[46] More recently Taylor and King have argued that the *Physics*'

40 E.g.: David E. Hahm, *The Origins of Stoic Cosmology* (Columbus, OH: Ohio State University Press, 1977), p. 106; F. H. Sandbach, *Aristotle and the Stoics* (Cambridge: Cambridge Philological Society, 1985), p. 42.

41 Sorabji, *Matter, Space and Motion*, p. 133.

42 Aristotle *On the Heavens*, 1.9 (278b22–279a11).

43 Sorabji, *Matter, Space and Motion*, parts 2–3.

44 See especially: Aristotle *Physics*, 4.4 (212a14–212a21).

45 Especially: Aristotle *Categories*, 6 (6a8–6a14).

46 Ammonius *On Aristotle Categories* (58,16–26); Philoponus, *On Aristotle Categories* (87,7 ff.); Simplicius, *On Aristotle Categories* (125,17 ff.).

definition is, in effect, the same as the *Categories*' definition, in so far as the *Physics* (contrary to first appearances) presents a conception of place as an extension.[47]

The second camp claims that Aristotle's view changes, or develops, from the writing of the *Categories* to the *Physics*. So Mendell argues that the *Categories* is basically the view of the Platonic Aristotle, while the *Physics* definition is freed of Aristotle's early Platonism (itself a contentious issue).[48] Algra argues that in the *Categories* Aristotle was not really theorizing about place, but only in the *Physics* turned his attention directly to the notion of place.[49]

I will now sketch a position, akin to that of Algra, which seems to me to be the most defensible. There are two senses in which place may be extension in the *Categories*. It may be an extension independent of the body so occupying it (akin to a region of absolute space; this is Algra's view of its status in the *Categories*) or the extension which a body bears as a property. It is difficult to decide which is the view Aristotle holds in the *Categories*, and there are philosophical problems with both. The former view suffers from having it that place comes out as a substance, and not something dependent on ("said of" or "present in" an object, in the technical vocabulary of the *Categories*) the body so in place (as it is clear place should be as a subdivision of the non-substantial category of quantity). The latter view suffers from two main problems. First, Aristotle systematically employs the verb *katekhein* to describe the relationship of bodies to places, and this is best rendered by "occupy"; and, second, such an account of place which has it as a property of an object makes it difficult to see how bodies can move or exchange places. In the *Categories* Aristotle is clearly guilty of one oversight or another.

However in the *Physics* he is certain of (at least) two things about the status of place. First it is still said of a subject.[50] This rules out the "place as independent extension" model. Second, he is clear that place can be separated from that which is in place.[51] Aristotle believes this, since he believes that bodies can come to be in the same place as each other. This, however, rules out the "place as extension of a body" model. Aristotle is stuck; he needs a conception of place that is separable from the extension of a body in place, yet still dependent on body, still "said of a subject." His brilliant solution to this dilemma is to divorce the dependence of place from the body which is in place (and so it is separable from that body), yet have it still said

[47] H. R. King, "Aristotle's Theory of ΤΟΠΟΣ," *Classical Quarterly* 44 (1950), 76–96; A. E. Taylor, *A Commentary on Plato's Timaeus* (Oxford: Clarendon Press, 1928), pp. 664–77.

[48] Henry Mendell, "Topoi on Topos: The Development of Aristotle's Concept of Place," *Phronesis* 32 (1987), 206–31.

[49] Keimpe Algra, *Concepts of Space in Greek Thought* (Leiden: Brill, 1995), chapter 4. See also: Benjamin Morison, *On Location. Aristotle's Concept of Place* (Oxford: Oxford University Press, 2002).

[50] As the opening line of *Physics* 4.2 makes clear.

[51] As Aristotle *Physics* 4.2 (209.22 ff.) makes clear.

of body, just not the body in place, but the surrounding bodies. And so we have the definition of body as the surface of the first surrounding (immobile) body.

However, this new definition of place as found in the *Physics* is problematic. Most of the problems arise from the condition that the surface which forms the place of an object must be immobile. These problems were to exercise subsequent thinkers from the time of Theophrastus (Aristotle's student and successor as head of the Peripatos) through the seventeenth century.[52] The history of these worries is a rich one. For now it will do merely to mention some of the problems. Consider a boat moored in a river. Here the surrounding water is not immobile, and so if one were to hold on to the definition, such a boat would be changing its place. (Note: this is only problematic if change of place is thought to be sufficient for a body moving, for what we object to in this case is having to conclude that the boat is moving.) Alternatively, a boat floating downstream with the flow of the river would not be changing its place (since the surrounding water is the same). Note, however, that this is only problematic if change of place is a necessary condition of motion.

So, it is standardly thought, Aristotle amends the definition of place to be the surface of the first surrounding immobile body, and so abandons the requirement that a body's place is in immediate contact with it. On the example just given the place of the boat is now determined by the banks of the river. Yet this revised view is also problematic. First, Aristotle seems to remain committed to the contact requirement of place, and a body will often not be in contact with its first immobile surrounding body. Second, two boats in the same river seem to have the same place, one boat moving across a river seems not to change its place, and a boat in a still lake has its place suddenly move from being the surrounding water to being the banks of the lake every time the wind blows the waters about. These problems, and others related to them, were authoritatively discussed by Theophrastus, and formed the basis for many subsequent discussions of place.

Aristotle's views on time proved to be equally influential to those on place.[53] Perhaps the main source of Aristotle's influence on subsequent inquiries concerning time is his creation of some puzzles or paradoxes about time, the solutions to which are far from clear, and indeed Aristotle's own response to them (if he had one) is hotly contested. These

[52] Theophrastus' arguments are preserved by Simplicius, *On Aristotle Physics* (604,5–604,11). For an account of these arguments and subsequent Greek responses see: Sorabji, *Matter, Space and Motion*, part 2.

[53] Sorabji, *Time, Creation and the Continuum*; G. E. L. Owen, "Aristotle on Time," in Peter Machamer and Robert Turnball (eds.), *Motion and Time, Space and Matter: Interrelations in the History of Philosophy and Science* (Columbus, OH: Ohio State University Press, 1976); Tony Roark, *Aristotle on Time. A Study of the Physics* (Cambridge: Cambridge University Press, 2011); Ursula Coope, *Time for Aristotle. Physics IV.10–14* (Oxford: Oxford University Press, 2005).

puzzles threaten to short-circuit any discussion of the nature of time by demonstrating that time does not in fact exist.

In the *Physics* 4.10 Aristotle presents the paradoxes intended to demonstrate the unreality of time. The problem is that time is composed of the past and the future. But the past has occurred, and so is not, while the future will be, and so is not yet. Therefore time as a whole is not. If one complains that the now, or the present, has been ignored, Aristotle adds that "now" is not a part of time, since a whole is composed of its parts, and time is not composed out of nows. So, time does not exist since no one of its parts exist.[54]

This first puzzle is followed by a second, whose purpose is partially to convince one that time is not in fact composed out of nows.[55] Is it always the same now, or a different now? Assume it is always a new, different now. If so, we can ask, when does a given now (say noon) cease to exist? It cannot cease to exist while it exists (a simple impossibility), nor at the next instant (for instants, by their very nature, do not have nexts – they are densely ordered). The only remaining option, that noon ceases to exist at some arbitrary instant later than noon (say noon and one second) is also absurd, for then the instantaneous now that is noon would last for a second.

The final puzzle disallows the remaining possibility, namely that the now is always the same. Were this to be the case, there would only be one instant, but any finite span of time must be bounded by two instants. More importantly, all events would be simultaneous if they all happen during the same now. These last two puzzles should, among other things, convince one that time is not composed out of nows.

Aristotle never explicitly addresses these paradoxes, although the paradoxes themselves were to greatly influence subsequent Greek discussions concerning the reality of time.[56] Aristotle's solutions are clearly intimately bound up with his views on infinity and divisible magnitudes. That he thinks the paradoxes can, and indeed must, be solved is made clear by the fact that he goes on to give an account of what time is. There have been many conflicting accounts of just what his definition of time amounts to. He defines time as "the number of motion with respect to the before and after, and continuous."[57] This definition is taken to be equivalent to another, namely that time is the measure of motion.[58] These definitions raise (at least) two fundamental questions: in what sense is time a measure or a number, and what motion, or sort of motion, is it the number or measure of? There is little consensus concerning the answers to either of these questions.

[54] Aristotle *Physics*, 4.10 (217b29–218a7).
[55] Ibid., 4.10 (218a8–218a20).
[56] Diodorus Cronus, Chrysippus, Epicurus, Apollodorus, Posidonius, Pseudo-Archytas, Plutarch, Sextus, Eusebius, Basil, Alexander, Augustine, Iamblichus, Damascius, and others address the paradoxes. For an account of these responses see: Sorabji, *Time, Creation and the Continuum*, part 1.
[57] Aristotle *Physics*, 4.11 (220a25).
[58] See ibid., 4.11 (219b3–219b5), but the equation is suspect.

Concerning the first question Aristotle thinks that time is that which is countable with respect to change.[59] Aristotle seems to talk of counting both instants (which, recall, are not parts of time per se), and instantaneous stages of changes. Both these candidates, and others, have been seen to yield a potentially circular definition. In a sense the confusion is between a conception of time as something with which we count versus something that we count.

As for the relationship between number and measure, opinion divides as to whether (1) number and measure are simply used equivalently, or (2) time is thought of as what measures change, or (3) time is itself a measurable aspect of change. In any case it is clear that for Aristotle, like many other Greek thinkers, time requires change. It is sometimes thought that Aristotle makes the regular circular motions of the heavens the measure of time, since their number is best known.[60] The issue is whether the very existence of time is contingent upon the existence of the regular circular motions of the heavens or, rather, whether such motions, both easily visible and well studied, are merely the easiest way to measure time. In other words, some think that the celestial motions are the motion which time is the number of, while others think such motions are merely a best example. Aristotle's thoughts on infinity have proven to be both greatly influential and subtle. Famously, he denies the existence of any actually infinite collections, while again bringing the modal metaphysics of potentiality versus actuality to the fore. Much that he has to say of interest concerning the infinite arises in two related contexts: responses to Zeno's paradoxes of motion, and refutations of atomism (itself developed partially in response to Zeno).[61]

The dichotomy paradox, or the paradox of the half-distances, claims that one must traverse an infinite number of half-distances in order to reach a given destination. This follows from the fact that in order to traverse any interval, one must first traverse half of it. This half-interval is itself an interval, and so to traverse it, first one must traverse half of it, and so on. If no intervals can be traversed, then motion is impossible. Responses to this paradox depend crucially on what one takes the impossibility to be. Aristotle takes Zeno to have thought that the infinite number of sub-intervals would require an infinite amount of time to traverse, since each would take some time to traverse.[62] Here Aristotle gives the proper response, that time is infinitely divisible just as is the interval, and since the interval is itself finite in length (but infinitely divisible) so too the time to traverse the interval is finite in duration (but infinitely divisible). Aristotle realizes that the problem, in its most general form, concerns the possibility of traversing any infinite

[59] Ibid., 4.14 (223a29–223b1).
[60] Ibid., 4.14 (223b18–223b20).
[61] Aristotle's central account of the infinite is found at ibid., 3.5–7. But see also: ibid., 6.2, 6.9, 8.8; Aristotle *On Generation and Corruption*, 1.2.
[62] Aristotle *Physics*, 6.2 (233a21 ff.).

collection of intervals, be they spatial or temporal. The problem is a particularly acute one for Aristotle, since he denies that any infinite collections can be traversed. This denial itself is secondary to Aristotle's denial that there can even be any infinite collections. Why? An absurdity Aristotle sees with infinite collections is that they would have to contain, *per impossibile*, sub-collections which would also be infinite.[63] This argument against the existence of actual infinite collections becomes standard in the western tradition and, indeed, is finally (and ironically) made definitory of infinite collections by Cantor. In addition, Aristotle believes that if an infinite collection were to be traversable, it would be countable, which he deems impossible.[64]

The impossibility of counting an infinite collection naturally suggests Aristotle's own conception of the infinite as an extendible finitude – something which one can always take more of, or add on to. Something infinitely divisible, for example, can always be divided again, while number is infinite in the sense that you can always count one higher.

Yet there is still a major problem concerning motion, which Aristotle is aware of. If it is impossible to traverse any infinite collections, how can any interval be traversed, since Zeno's paradox seems to demonstrate that all intervals do contain an infinite number of sub-intervals?[65] Here is where Aristotle relies again on the notion of existence in potentiality. He denies that any continuous interval actually contains, or is actually composed out of, an infinite number of sub-intervals. The act of dividing an interval brings into existence a sub-interval that did not previously actually exist. Sub-intervals are created, not discovered. This point is generalized to apply to positions. As an object moves you can pick out, and therefore bring into actual existence, as many positions as you care to. But these positions are not already actually there according to Aristotle.

This last point is part and parcel of Aristotle's recognition that continua are not composed out of punctiform entities. Time is not composed out of instants, and space is not composed out of points. There is a deep sense in which Aristotle is correct here, for sizeless entities cannot compose something with size. As Aristotle correctly puts its, point-like entities are not parts of continua.[66] This insight is directly brought to bear on his (correct) solution to another paradox of Zeno's, that of the arrow.

The arrow paradox attempts to prove that an arrow in flight is actually at rest. Consider the arrow at any instant you care to pick during its flight. At this instant it occupies a space equal to itself. Yet this seems to be characteristic of being at rest; during any temporal interval that some thing is at rest it occupies a space equal to itself, unlike something moving during this same

[63] Ibid., 3.5 (204a20–204a26).
[64] Ibid., 8.8 (263a7–263a11).
[65] Ibid., 8.8 (263a4–263b9).
[66] Ibid., 6.1 (231a24 ff.).

temporal interval. Yet, so the paradox goes on, since this is true at any instant you care to pick, it must be true of the whole temporal interval during which the arrow is said to be moving, and so the arrow is actually at rest.

Aristotle's reply is to deny that time is composed of nows, and so to deny that we can infer that the arrow is at rest during the whole interval because it is at rest at any (and every) instant.[67] The instants are not the proper parts of the interval. The interval can be divided into sub-intervals, but the arrow of course occupies a space larger than itself during any given sub-interval. Motion through, or during, an interval is not built up out of successive motions through, or during, point-like parts of the interval. These point-like entities do not actually exist, and when you divide an interval at a point, the point is but the boundary of two sub-intervals that have been brought into existence, and not even now in any sense a part.

The preceding discussion allows us now to make sense of Aristotle's objections to atomism, as found in *On Generation and Corruption*, 1.2.[68] The argument for atoms, that is for the existence of indivisible body, is a *reductio*.[69] Assume body is divisible everywhere. If so, it might be divided everywhere. This would yield problems, in particular the dissolution of the body into nothings, for the resultant punctiform entities could not compose an extended body. So body cannot be divided everywhere, so it is not divisible everywhere, so there must exist indivisible extended bodies, that is atoms. The worry seems to be that division is, in general, division into the parts that compose something. So, if a body is divisible everywhere, one is committed to claiming that bodies are composed out of punctiform entities, which is impossible. So, one must deny the infinite divisibility of body.

Aristotle agrees that there will always exist undivided bodies, what he denies is that there are therefore indivisible bodies. For Aristotle you can divide a given body anywhere you like, and as often as you like; you cannot, however, divide a body everywhere. Why not? Well, where will these divisions be made? Points are not next to points, as Aristotle says, and so the notion of simultaneous division everywhere is nonsense.[70] I can divide a body at any point I care to pick, but upon doing so, there is no "next" division to make, that is no division next to that first made, such that were I to complete this process all at once, I would complete the division of the body everywhere.

[67] Ibid., 6.9 (239b5 ff.)

[68] For Aristotle on the ancient atomists see: Sorabji, *Time, Creation and the Continuum*, chapter 21; David J. Furley, *Two Studies in the Greek Atomists* (Princeton, NJ: Princeton University Press, 1967); David J. Furley, *The Greek Cosmologists. Volume 1. The Formation of Atomic Theory and its Earliest Critics* (Cambridge: Cambridge University Press, 1987); Stephen Makin, *Indifference Arguments* (Oxford: Blackwell, 1993); Eric Lewis, "The Dogmas of Indivisibility: On the Origins of Ancient Atomism," *Proceedings of the Boston Area Ancient Philosophy Colloquium* 14 (1998), 1–21.

[69] Aristotle *On Generation and Corruption*, 1.2 (316a10–317a12).

[70] Ibid., 1.2 (317a1 ff.).

12

ARISTOTLE AND THE ORIGINS OF ZOOLOGY

James G. Lennox

PRECURSORS OF ZOOLOGY

If understood as a systematic investigation aimed at scientific knowledge of animals as such, zoology begins with Aristotle. Nevertheless, there are two different groups of texts that can be seen as precursors to his more systematic, theoretically motivated investigations: the various discussions of animals that one finds in the writings of many of the "Presocratic" natural philosophers, which are regularly discussed and criticized by Aristotle; and those works in the Hippocratic corpus that are specifically focused on, or appeal to, anatomy and dissection.

In the first category one must distinguish between an earlier and a later group. Thales, Anaximander, and Anaximenes lived in the Ionian port of Miletus in the early to middle sixth century BCE. No manuscripts of their written work survive, and the evidence for their views about "life," "soul," and the origins of animals is very slim – Aristotle speaks as if most of his information about them is second-hand. The same cannot be said for Anaxagoras, Empedocles, and Democritus: they hailed from widely diverse locations, and it appears that each of them included discussion of the origins, development, and attributes of animals as part of their writings on nature. Aristotle quotes directly from them and often discusses and criticizes their views on topics ranging from respiration and perception to embryological development. As far as we can tell from our sources, these thinkers did not view animals as a distinct subject of investigation, but as the locus of phenomena for which their cosmological first principles would need to account. Nevertheless, Aristotle takes their accounts seriously.

In the Hippocratic medical writings, some of which are certainly from the fifth century BCE, we have a significant collection of entire works rather than fragments. But medicine is typically portrayed in these treatises as a craft, focused on the states of health and disease in human beings; so these treatises do not count as zoology, as it has been defined above. It is likely, however, that any dissections performed by these authors employed animals, since Greek city-states had proscriptions against defiling corpses.[1] The rapid development of medical anatomy and physiology in Alexandria in the Hellenistic period is generally attributed to the fact that human dissection was practiced systematically there for the first time.[2]

The Hippocratic corpus is important, too, for treatises that illuminate the role of experience and theory in the study of living things. For example, the explanation for epileptic seizures in *On the Sacred Disease* is based on a humoral theory, a theory of respiration, and an account of the vascular system that appears to have little basis in dissection or experimentation. By contrast, works like *Fractures*, *Epidemics*, and *Regimen for Health* appear to rest, at least partially, on careful, first-hand observation, including anatomical observations and some dissection. Works such as *On Ancient Medicine*, *On the Art*, and *On the Nature of Man* take different sides on the question of the relative value of the patient accumulation of medical experience versus the theoretical search for causes. These documents may well prefigure a central debate in Hellenistic medicine over the role of experience and theory in the advance of medicine. The Hippocratic corpus is thus a rich source of information about the context in which zoology proper originated.

ARISTOTLE

The first printed Latin edition of Aristotle's major zoological treatises, the *Libri de animalibus* of Theodore Gaza (1476), came during a period when anatomical dissection was undergoing radical transformation, and when Galen's anatomical claims were being challenged. Aristotle's zoology presented a clear alternative: the systematic study of animals as a branch of natural philosophy, and comparative anatomy as a part of that study. His zoology served as a model for the founders of modern comparative anatomy, embryology, and physiology in the Renaissance, and there was nothing of similar scope and sophistication until the eighteenth century.[3] Even in the nineteenth century, the great anatomist Richard Owen declared that "Zoological Science sprang from his [Aristotle's] labours . . . in a state of

[1] G. E. R. Lloyd, "Alcmaeon and the Early History of Dissection," *Sudhoffs Archiv* 59 (1975), 171–92; reprinted in G. E. R. Lloyd, *Methods and Problems in Greek Science* (Cambridge: Cambridge University Press, 1991), pp. 164–93.

[2] Heinrich von Staden, *Herophilus: The Art of Medicine in Early Alexandria: Edition, Translation and Essays* (Cambridge: Cambridge University Press, 1989).

[3] See Stefano Perfetti, *Aristotle's Zoology and its Renaissance Commentators (1521–1601)* (Leuven: Leuven University Press, 2000).

noble and splendid maturity."[4] And Charles Darwin, after having read a significant portion of a new English translation of *De partibus animalium* by William Ogle, wrote:

> From quotations I had seen I had a high notion of Aristotle's merits, but I had not the most remote notion what a wonderful man he was. Linnaeus and Cuvier have been my two gods, though in very different ways, but they were mere school-boys to old Aristotle.[5]

ARISTOTLE'S PHILOSOPHY OF ZOOLOGY

Aristotle's remarkable achievements in zoology owe a great deal to his philosophical reflections on the nature of science, as set out in his *Posterior Analytics* (APo.). According to that work, the goal of inquiry – scientific knowledge (*epistêmê*) – is a system of concepts and propositions organized hierarchically, resting on a rational grasp of the essential natures of the objects of study and other first principles. The *form* of a scientific demonstration is modeled on the syllogistic proofs presented in the *Prior Analytics* (APr.), but, in addition to entailing the conclusion, the premises of a scientific demonstration must be true, primary, and immediate (not themselves in need of proof); and they must be better known than, be prior to, and give the cause of the conclusion (APo. I.2, 71b20–5).

The second book of the *Posterior Analytics* discusses *how* to achieve this goal. It distinguishes two paired goals of inquiry: establishing the fact (*to hoti*) and discovering the reason why (*to dioti*) it holds; determining whether something exists (*ei esti*) and determining what it is (*ti esti*) (APo. II. 1, 89b23–5). The first pair proceed from establishing *the facts* – about which attributes belong universally to which subjects – to establishing the causal explanations for those facts: for example, after finding that the moon regularly suffers eclipse, we go on to determine what causes it to be eclipsed. The second pair proceeds from establishing *that* some object of investigation actually exists (e.g. centaurs, eclipses) to seeking to determine their essential natures, as reflected in their definitions.

The second chapter, however, makes it clear that these two, two-stage inquiries are intimately intertwined – more like two perspectives on scientific inquiry than distinct inquiries. They are both said to be engaged in a search for *causes*, and many causal demonstrations can be reformulated as definitions of the phenomena being explained: for example, if thunder is due to fire being extinguished in the clouds, a proper scientific definition of

[4] Phillip R. Sloan (ed.), *Richard Owen: The Hunterian Lectures in Comparative Anatomy, May and June 1837* (Chicago, IL: University of Chicago Press, 1992).

[5] Cf. A. Gotthelf, "Darwin on Aristotle," *Journal of the History of Biology* 32 (1999), 3–30.

thunder would be "a noise in the clouds caused by the extinguishing of fire" (APo. II.10, 94a1–10).

The *Posterior Analytics* provides a highly abstract account of scientific knowledge and inquiry, applicable equally to many disciplines. The first book of *On the Parts of Animals* (PA I) provides the more specific norms for the task of an organized inquiry into the world of animals. Its recommendations, and their relationship to those in the *Posterior Analytics*, can be summarized in the following six points.[6]

[i] Scientific inquiries have a stage in which factual information is systematically collected and organized and another, later, stage during which causal explanations are sought; but in zoological inquiry the search for causal explanations is complicated by the fact that the order in which parts develop is determined by the final cause of development, that is the form of the living organism (PA I.1, 639b7–11, 640a10–19, 640b1–4).

[ii] Scientific inquiry aims at knowledge of causes, but in the zoological context *goal* causation (the basis of teleological explanation) takes precedence over motive (efficient) causation (PA I.1, 639b12–21, 642a2–13).

[iii] Our demonstrations must display *necessary* relations, but in zoology the necessity is often *conditional* – *given* a certain goal, certain materials and processes are necessary; but they do not necessitate the goal (PA I.1, 639b21–640a3, 642a3–13).

[iv] All sciences seek knowledge of essences, but in the case of animals this means understanding both the organic systems that make up living *bodies* (material natures) and, more fundamentally, the souls (living capacities) for the sake of which those living bodies are organized (formal natures). The work of our predecessors shows the limitations of relying solely on the material constituents of animals and their interactions in attempting to understand their natures, without inquiring into the activities that constitute their lives (PA I.1, 640b5–641b10).

[v] Analytical division of the attributes of general kinds into their differentiated sub-kinds can help explain the natures or essences of animals; but,

[6] In the twentieth century this view was first vigorously defended by Wolfgang Kullmann, *Wissenschaft und Methode: Interpretationen zur aristot'lischen Theorie der Naturwissenschaft* (Berlin: de Gruyter, 1974) and developed by Allan Gotthelf, "First Principles in Aristotle's Part of Animals," in Allan Gotthelf and James G. Lennox (eds.), *Philosophical Issues in Aristotle's Biology* (Cambridge: Cambridge University Press, 1987), pp. 167–98; James G. Lennox, "Divide and Explain: the Posterior Analytics in Practice," in Gotthelf and Lennox (eds.), *Philosophical Issues*, pp. 90–119; James G. Lennox, "Between Data and Demonstration: The Analytics and the Historia Animalium," in Alan C. Bowen (ed.), *Science and Philosophy in Classical Greece* (New York: Garland Publishing, 1991), pp. 261–94; and David Charles, *Aristotle on Meaning and Essence* (Oxford: Oxford University Press, 2000). Recent research on the introduction to Albertus Magnus' *De animalibus* has shown that he too was aware of the relationship between PA I and the *Analytics* (see Michael Tkacz, "Albert the Great and the Revival of Aristotle's Zoological Research Program," *Vivarium* 45.1 (2007), 30–68).

for beings of such complexity, dichotomous division as practiced in Plato's Academy will not do. A method of multi-axis division must be used, with every differentiation being into determinate forms of the one above: divisions must reflect all the differentiated forms of structures and their associated functions (PA I.2–3, 642b5–644a12).

[vi] We need to establish the existence of the kinds to be investigated; but zoological kinds are complex and must be identified at many levels of generality. Aristotle identifies certain "great kinds" (*megista genê*) that have many specific forms closely similar to each other (e.g. birds, fish, insects, crustaceans). The specific forms of these kinds share attributes that differ only in degree; but he also recognizes that there are analogical identities *across* these kinds as well as *specific* kinds that unite individual animals that are formally identical (PA I.4, 644a12–b16; PA I.5, 645b1–646a2).

Those norms for *zoological* inquiry are systematically discussed in the first four chapters of PA I. The fifth and last chapter opens with a stirring encomium to the study of animals as a worthy subject for the philosopher of nature. Here is a taste of it:

> it remains to speak about animal nature, omitting nothing in our power, whether of lesser or greater esteem. For even in the study of animals disagreeable to perception, the nature that crafted them provides extraordinary pleasures to those who are able to know their causes and are by nature philosophers. Surely it would be unreasonable, even absurd, for us to enjoy studying likenesses of animals – on the ground that we are at the same time studying the art, such as painting or sculpture, that made them – while not prizing even more the study of things constituted by nature, at least when we can behold their causes. For this reason we should not be childishly disgusted at the examination of the less valuable animals. For in all natural things there is something marvelous. (PA I.5, 645a5–16)

PA I concludes with a brilliant integration of its principal themes – applying ideas about using similarities among *parts* to identify kinds and their forms to the *activities* performed by animals, and then providing a sort of map of the network of possible functional relationships *between* parts and activities. With that theoretical framework in place, one can get down to the task of actual zoology.

ARISTOTLE'S ZOOLOGY IN OUTLINE

Aristotle's zoological writings make up roughly a quarter of the entire extant Aristotelian corpus. They include: *History of Animals* (HA, 10 books);

Dissections (an uncertain number, all now lost);[7] *On the Parts of Animals* (PA, 4 books); *On the Generation of Animals* (GA, 5 books); *On the Locomotion of Animals* (MA, 1 book); *On the Progression of Animals* (IA, 1 book); *Small Natural Studies* (PN) = *On the Senses and their Objects*; *On Memory and Reminiscence*; *On Sleep and Waking*; *On Dreams*; *On Divination from Dreams*; *On Length and Shortness of Life*; O*n Youth and Old Age*; *On Life and Death*; and *On Respiration*.[8] There are, in addition, two other works that are intimately tied to this animal project: *On the Soul* (An., 3 books), an investigation of fundamental capacities of living things (nutrition, reproduction, perception, locomotion, reason); and *Meteorology* IV (Mete. IV), a systematic study of the emergence of uniform materials with powers and qualities not possessed by any of the elements, among which are the "tissues" of living things such as flesh, bone, blood, or hair. Aristotle depends heavily on its conclusions in the discussion of uniform parts in PA II and their development in GA II and V.

There are many self-contained and potentially interdependent studies here. This may not seem unusual, but Aristotle was the first person in the history of science to see the study of nature as an articulated complex of interrelated, yet partially autonomous, investigations. The study of animals that Aristotle initiated is itself similarly articulated – while Aristotle is at great pains to stress that each work is part of a complex unity.

HISTORY OF ANIMALS

Aristotle concludes a long methodological introduction to the *History of Animals* by characterizing its theoretical task in relationship to the zoologist's ultimate goal of causal understanding of why different animals have their characteristic features.

> These things have now been stated in outline, in order to provide a taste of which things . . . we need to study *in order that we may first grasp in all cases the differences that are present and the attributes . . . After this we must attempt to discover their causes*. For this is the natural way (*kata physin*) to carry out the inquiry (*tên methodên*), when the history (*historia*) about each thing is present; for it is from these things

[7] These are mentioned in the ancient lists of Aristotle's writings (e.g. Diogenes Laertius, *Lives of the Philosophers*, V.25 (translated by R. D. Hicks, 2 vols. (Loeb Classical Library; Cambridge, MA: Harvard University Press, 1925)) lists eight books of *Dissections* and one of *Selections from the Dissections*), and Aristotle refers to them twenty-eight times in the other zoological works. Most of the references appear to concern a collection of illustrations or diagrams, although some simply explain what would be seen in a dissection, or right and wrong ways of tackling it.

[8] Aristotle appears to refer to two other works related to these studies: *On Nutrition* and *On Plants*. But the former may not have been a separate work, and Aristotle's references to the latter might well refer to writings of Theophrastus. The medieval work *On Plants* is spurious. A work that is typically considered spurious is *On Breath*, which is usually placed last in manuscripts of the *Parva naturalia*.

> that both the things *about which* there needs to be demonstration and the things *from which* there needs to be a demonstration become apparent. (HA I.6, 491a7–14)[9]

A "history" (*historia*) is the report of an inquiry into facts about the objects of investigation. Zoological investigations begin with this fact-gathering and organizing stage, and on completion of a "history" we are prepared to search for the causes that will reveal why animals have certain parts and behave and develop as they do. Aristotle regularly reminds us of this methodological maxim, for example at the beginning of his causal investigation of animal parts in PA II and his causal investigation of variations in animal locomotion in IA 2.

While we have no first-hand reports about how Aristotle actually gathered the vast amount of information found in his zoological works, there are helpful clues scattered throughout. The HA itself shows how he thought the results of such an investigation should be organized.

Book I opens with a theoretical discussion of different kinds of parts and the ways in which they are alike or different (in form, in kind, or by analogy). Uniform parts (tissues such as bone, blood, and flesh) are distinguished from non-uniform (organs such as heart, liver, hand, and eye). Animals the same in form are those with parts that are identical in form (so one horse's teeth, eyes, hair, hooves, etc., will be identical in form to those of another horse); animals the same in kind but different in form (for example different species of the kind Bird) will have parts that differ in degree or by "more and less" (all members of the kind Bird have beaks, but these differ in size, shape, texture, color, and so on); and finally, animals that *differ* in kind may still be *alike by analogy*, as the feathers of birds are analogous to the scales of fish, or the lung of a breathing animal is analogous to the gills of fish.

A passage characterizing the way different forms of animals within one kind are distinguished from each other gives a flavor of Aristotle's approach.

> Most of the parts in these kinds differ by way of oppositions of their affections, for example of color or shape, by way of the same part in some cases being affected more, in some cases less; and again by a greater or smaller number [of the part], and by [the part] being large and smaller, and generally by "excess and deficiency." For some animals will have soft flesh and some hard, some will have long and others a short beak, some will have many feathers and others few. Yet some parts in these kinds belong to some and others to others, i.e. some [birds] have spurs and others do not, some have crests and others do not. But roughly speaking, most of the parts from which the bulk of the body is constituted are either the same or differ by these

[9] Compare this passage with the more general account of preliminary scientific investigation in *Prior Analytics* I.30, 46a22–7.

> oppositions and by excess and defect – for "the more and less" can be regarded as a sort of "excess and defect." But some of the animals do not have parts that are the same in form or in virtue of excess and defect, but in virtue of analogy, for example bone in comparison to fish spine, nail to hoof, hand to claw, or feather to scale; for what is feather in a bird is scale in a fish. (HA I.1, 486a25–b22)

Shortly after this passage (at 487a10 ff.) he argues that this same analysis of likeness and difference can be applied to three other broad categories of animal differentiae: their way of life (*bios*), activities or actions (*praxeis*), and dispositions of character (*êthê*).[10] While we know that in gathering information about the similarities and differences among the *internal* parts Aristotle used dissection systematically and across a wide range of animals, the opening chapter of HA indicates that he did so with a sophisticated theory of the *kinds* of similarity to look for, and that besides parts there were also differences in *activities*, *ways of life*, and *character* to be investigated.

After this introduction, books I.7–IV deal with differences among parts, doing so in a highly organized fashion. External non-uniform parts of human beings are discussed first, on the grounds that they are the most familiar to us (I.7–15); but I.16 begins with a revealing statement of the need for a different approach in studying our internal parts:

> The parts visible externally, then, are ordered in this way; and as we have said, they have been given names and on account of their familiarity are well known; but the situation is just the opposite with the internal parts. For we are mostly ignorant about the internal parts of mankind, so that it is necessary to investigate them by referring to the parts of the other animals, which are similar in nature to ours. (494b18–24)

And, indeed, while the rest of book I is devoted to the internal, non-uniform parts in humans, it is rich in comparative information about these internal organs as they are found in blooded animals. Book II then returns to the external non-uniform parts, but organizes the discussion around the "great kinds" of blooded animals; it then goes on to the internal non-uniform parts of the blooded animals (II.15–III.1). The uniform parts of blooded animals are described from III.2–10; and, finally, Aristotle considers the internal and external parts of the bloodless animals (IV.1–9).

With respect to each part, the general strategy is first to identify it at the most common level. Then ways in which it is differentiated in different subclasses are noted, and the other parts, with which the part under discussion is related, are universally correlated at each level. Within each of these broad categories of parts, Aristotle uses a privileged group of what he refers to as

[10] To see this method applied to these other categories, see the introduction to HA VII.1, 588a16–b12.

Aristotle's Great Kinds	**Modern Classes and Orders**
Blooded	Vertebrates
Bird	Birds
Fish	Fish
Cetaceans	Cetaceans (aquatic mammals)
Four-legged live-bearers	Land mammals (except humans)
Four-legged egg-layers	Reptiles, Amphibians
Bloodless Animals	Invertebrates
Soft-bodied	Cephalopod Mollusks
Soft-shelled	Crustaceans
Hard-shelled	Testaceous Mollusks
Insects	Insects, Arachnids

"great kinds" (*megista genê*)[11] to order the data. These kinds are characterized by a significant number of correlated parts that only differ by continuous degree (or "by the more and less") from one form of the kind to the next. Identifying these kinds is discussed as a theoretical problem in PA I.4, and most of HA I.6 is devoted to deciding how many such groups there are. A list of blooded "great kinds" is reiterated in HA II.15; and in HA IV.1 he lists the bloodless kinds.

Of utmost importance for understanding the place of HA in Aristotle's zoological research, brought home convincingly by the research of David Balme and Pierre Pellegrin, is that these kinds are not the primary tools for organizing information in HA, nor are they intended as a means of ordering all animals into a systematic, hierarchically organized taxonomy of the sort Carl Linnaeus aimed to produce in the eighteenth century. The primary organization of HA is by means of the four categories of differentiae identified in its opening pages (parts, activities, ways of life, habits of character), and above all Aristotle searches for universal, and especially coextensive, correlations among and between these differentiae, at various levels of universality and specificity. A useful way to think of his "great kinds" is as identified loci of such correlations where the correlated differences all vary along continua from one group to another within the kind.

This point will be illustrated in the case study, below. But, for now, think about the ways in which beaks, shared by all birds, differ from one kind of bird to another, or the ways in which the external shells shared by all crustaceans differ from crabs to lobsters to shrimp.

The structure of the rest of HA can be more briefly described. Books V–VI deal with different activities related to reproduction (differences in

[11] I am purposefully avoiding rendering *megista* as a superlative. The same phrase, *megista genê*, was used for a much more abstract set of concepts in Plato's *Sophist* (where the great kinds were things like Being, Same, Other, Motion, and Rest), and Aristotle uses the same phrase to refer to four less extensive kinds of soft-shelled animals in PA IV.8, 683b25–30.

methods of copulating, breeding seasons, modes of reproduction, gestation times, and development). Some sections of these books are more highly organized and systematized than others, suggesting that this may have been a work still in progress when Aristotle died. Book VII deals with ways of life, but typically described in relation to activities and habits connected with nutrition and reproduction. Its opening chapters are theoretical in orientation, discussing questions of the application of categories of similarity and difference and the use of division in studying variations in ways of life, while some of the discussions later in the book are not well integrated into the wider discussion. Book VIII deals with the character traits of different animals, and again appears to be less systematically organized in later sections. Finally, Book IX deals with human reproduction. Some manuscripts of HA include a Book X, but there is now a consensus that it was not part of the HA project.[12]

Throughout, Aristotle uses a methodology of multi-axis "division and correlation" to establish relationships of similarity and difference at various levels of generality, using "great kinds" as an aid in this process. One primary goal of *historia*, for which these kinds are immensely valuable, is to find, among the *universal* correlations, *coextensive* correlations among differentiae. These will be the correlations that will allow Aristotle to identify causally fundamental properties. He is unconcerned that many animals (including human beings) do not fall into these "great kinds." The kinds are of value because they identify groups with many distinct forms that share systematically correlated attributes. Being able to track these systematic networks of correlations among parts, activities, and behaviors will aid in the discovery of the causal relationships among animal differentiae. Primary among causal relationships are those that explain a part's differential attributes by reference to the functional demands of an animal's specific way of life.

For example, "bird" is a scientifically useful kind since it identifies blooded, egg-laying animals that fly, have beaks, feathered wings, two

[12] Some (e.g. D. M. Balme, "Historia Animalium Book Ten," in J. Wiesner (ed.). *Aristoteles: Werk und wirkung. Paul Moraux gewidmet*, 2 vols. (Berlin: de Gruyter, 1985), pp. 191–206) believe it is an early work of Aristotle, others that it is not by him (for references see Aristotle, *History of Animals, Books I–III*, trans. A. L. Peck (Cambridge, MA: Harvard University Press, 1965), pp. lvi–lviii). How modern editions should order these books is at issue. All the manuscripts have the material ordered as I have described it above, but the Renaissance editor Theodore Gaza moved the discussion of human generation to adjoin the other books on generation – thus book IX became book VII and the books on ways of life, activities, and character traits became books VIII and IX. All modern printed editions followed Gaza until Harvard University Press asked David Balme to produce a text and translation of the last three books for the Loeb Classical Library. Balme made a strong case for following the manuscript ordering, and did so in the third volume of the Loeb HA in 1991. Cf. Aristotle, *History of Animals Books VII–X*, ed. and trans. D. M. Balme, prepared for publication by Allan Gotthelf (Cambridge, MA: Harvard University Press, 1991). Balme also left a nearly completed draft of a new text and commentary of the HA based on a new collation of the extant manuscripts, which also reproduces the original book order. This was published under the editorial guidance of Allan Gotthelf (D. M. Balme, *Aristotle: Historia Animalium Volume I: Books I–X: Text* (Cambridge Classical Texts and Commentaries 38; Cambridge: Cambridge University Press, 2002).

fleshless legs, and so on. Each of those common attributes performs an important biological *activity* for birds, and is differentiated by more-and-less differences to suit the different *ways of life* of different kinds of bird; carnivorous birds, for example, have correlations of hooked talons, hooked beak, keen eyesight, powerful wings, and soaring and diving flight. Kinds are articulated into a large number of sub-kinds down to "indivisible forms," and at each level Aristotle searches to identify more specific coextensive correlations among the differentiae at that level. How this aids Aristotle's causal/explanatory project we shall soon see.

ON THE PARTS OF ANIMALS

As the corpus of Aristotle's work has come down to us, the first book of the *On the Parts of Animals* is actually Aristotle's philosophical introduction to the study of animals. The investigation into the causes of the parts of animals begins in book II, with a statement that the *historia* of the parts has been accomplished – we know which animals have which parts, and now we need to investigate *the causes* that explain this. (Remember that the first four books of HA are devoted to the parts of animals (I.7–IV.7) and the next two to animal reproduction and development – the causes of which are the subject of PA and GA respectively.)

He begins that inquiry by proposing a constitutional hierarchy of animal composition: the four elements (earth, air, fire, and water) are composed of four elementary powers (hot, cold, moist, and dry); the uniform parts (such as flesh, bone, blood, sinew, and fats) are composed from the four elements. The non-uniform parts or organs (such as hand, eye, heart, and lung) are composed of the uniform parts and these are the parts out of which the whole organism is composed. The last item in the order of coming into being, the whole organism, is first in the order of being: which uniform parts are generated, and where and when they are generated, is a function of which non-uniform part is being constructed, and the organization of organic parts is a function of which activities the animal must perform to live its particular form of life.

Aristotle here displays the relationship between a material-causal perspective on the organism and a final-causal perspective, and displays the way in which he prioritizes the final cause. As he puts it in PA I, paraphrasing a remark of Socrates in Plato's *Philebus*, coming-to-be is for the sake of being: the way in which an animal is composed as it develops is determined by the requirements of the fully developed animal's life. He exemplifies the relationship here by noting that a hand, in virtue of the many different activities it needs to perform, must be constituted out of hard rigid parts, soft pliant parts, and parts that stretch and contract, and all these must be so arranged that the hand can squeeze, grasp, and release. The bones, sinews, flesh, blood vessels, and nails must be in exactly the right overall arrangement for the hand to work.

The entire organization of the three books follows this introduction, starting with blooded animals; the bloodless animals (our "invertebrates") are taken up in Book IV, chapter 5. PA II.1–2 categorize the different uniform parts; chapter 3 discusses hot and cold in relationship to the nature of blood; and chapter 4 is a discussion of blood as the most basic uniform part, the final stage of nutriment before it is assimilated by the other parts. Blood varies in its fibrousness and this affects animal character, since fibrous blood tends to heat up more easily and quickly. The forms of fat found in organisms vary for the same reason: it occurs as lard in the thin-blooded but as suet in the thick-blooded. Aristotle goes on to discuss the material character of the brain, bone (and similarly hard materials), and flesh and then, in chapter 10, turns to the non-uniform or organic parts. There are relatively few teleological explanations in these chapters: much of Aristotle's explanatory work regarding the uniform parts explains their differences by reference to their distinctive material make-up.

Moving on to the discussion of non-uniform parts or organs (655b27–656a13), he explains that we need to have a way of organizing the investigation so that no part is overlooked. He suggests taking human beings as our model, because our "up" and "down" agree with cosmological "up" and "down" (656a11–13). (As he explains in IA 705a30–2: "Up and down are determined by function, not by position . . . alone. For the region from which the distribution of nutrient and growth for each animal comes is up, while the region to which this finally extends is down.") He thus starts by looking at the structures around the brain including the sense organs, their locations, and their orientation in different kinds, including a fascinating explanation of the elephant's trunk that is also used as a hand – "extraordinary in both size and potency" (658b34–5).

Teleological explanations now come to the fore, often focused on explaining why parts identical at a general level of description are differentiated as they are, and sometimes explaining why an animal kind has a part that is missing in otherwise similar animals.[13] Why do some eyes have eyelids, eyelashes, or eyebrows while others do not? (PA II.13–15) Why do some mouths come equipped with teeth, or tongues, while others do not (PA II. 16–17)? Why do four legged live-bearing animals with horns have fewer teeth but more stomachs than others (PA III.2)?

[13] Mariska Leunissen, *Explanation and Teleology in Aristotle's Science of Nature* (Cambridge: Cambridge University Press, 2010) has stressed that some of these teleological explanations (which she calls primary) insist that the animal's life requires that a certain part with a specific character be present, while others (secondary) describe the animal's nature as using residual material, produced as a by-product of the operation of primary teleology, in order to improve or enhance the animal's life in a certain way.

In PA III.3, after explaining that the neck is present in order to protect the larynx, windpipe, and esophagus, Aristotle shifts to discussing the internal parts of blooded animals, a discussion that continues until PA IV.5 and moves progressively towards parts that are not explained teleologically but as simply the products of material necessity. Indeed, during a long discussion of the presence of bile in certain species but not others and its variable quantity, Aristotle makes a methodological remark that is as important today as it was when he wrote it:

> Now sometimes nature even makes use of residues for some benefit, yet it is not on this account necessary to seek what something is for in every case; on the contrary, when certain things are such as they are, many other such things happen of necessity . . . So it is apparent that the bile is not for the sake of anything, but is a by-product. (677a16–19, 27–8)

Chapters 5–9 of book IV turn to the bloodless animals and are organized as a comparative review of first the internal and then external parts of each kind: soft-bodied animals (cephalopods); soft-shelled animals (crustaceans); hard-shelled animals (testaceous molluscs); and insects. Aristotle then returns, in chapter 10, to discuss the remaining external parts of blooded animals: first human beings; then the four-legged land animals; the birds; the fish; and finally organisms that share features of two different kinds (cetaceans, seals, bats, and ostriches; apes were earlier (689b32–5) said to share features of bipedal and quadrupedal live-bearing animals). In all these chapters there is a focus on the way all of an animal's external parts are organized around their way of life or *bios*.[14]

ON THE GENERATION OF ANIMALS

On a number of occasions during the PA Aristotle explicitly puts off providing explanations for the parts related to generation and development, saying he will do so in his study of generation. Developmental biology or embryology has, since Aristotle, been treated as a distinctive branch of biology, with its own methods and its focus on the causes of the generative process and on the sequential emergence through time of a complex, living being from a simple, undifferentiated "seed." There are a number of treatises in the Hippocratic corpus that discuss the nature of the seed, the

[14] Gotthelf, "First Principles" provides more detail on this structure and argues that it reflects the project of the *Posterior Analytics*. Allan Gotthelf, "The Elephant's Nose: Further Reflections on the Axiomatic Structure of Biological Explanation in Aristotle," in Wolfgang Kullmann and Sabine Föllinger (eds.), *Aristotelische Biologie: Intentionen, Methoden, Erbebnisse* (Stuttgart: Franz Steiner Verlag, 1997), pp. 85–96 focuses on the complexity of the explanatory project of PA through a detailed unpacking of the discussion of the elephant's "nose."

development of the human embryo and what would today be referred to as obstetrics and gynecology, but Aristotle's *Generation of Animals* is the first, and one of the greatest, works in the history of comparative developmental biology. This was acknowledged by E. S. Russell, who devoted a chapter to Aristotle in his *The Interpretation of Development and Heredity: A Study in Biological Method* (1930).

Like all of Aristotle's animal studies, GA is carefully organized. After a brief introduction placing this work in the wider context of his study of animals, focused especially on the *motive* cause of the coming to be of a new living being, he distinguishes animals generated through the coupling of male and female from those that are not, and identifies the capacities of the male and the female as the starting points of animal generation (except in cases of spontaneous generation, which he discusses in GA III.11 (762a9–763b16)). The next fifteen chapters concern the different kinds of male and female generative organs, and then in chapter 17 he discusses the nature of the male and female seed or contribution to generation (characterizing this as moving from examination of the *non-uniform* parts of generation to the *uniform* parts: 721a27–30).

The discussion of male and female contributions begins with a critical examination of previous theories, especially those claiming that particles from every part of the male and female constitute their contributions and those claiming that the female contributes nothing but a receptacle for germinating the male's seed. At 724a16, Aristotle begins the task of determining what the male and female contributions are, eventually concluding that they are residues of excess blood that have been specially prepared for their generative roles: in animals in which the female menstruates he take the menstrual discharge to be the female contribution, and the semen the male contribution. He argues that the male semen's role is only to transfer a generative motion, in the form of heat, to the female material, so it contributes nothing *materially* to the developing organism. Its heat is the source of generation and of the functional organization that emerges gradually and continuously during development. He uses shorthand for this theory, saying the male contributes the form and motive cause while the female contributes the matter, but this is misleading: in GA II.5, noting that in most birds and fish the female can produce an egg without any fertilization from the male, he argues that the heat contributed by the male is required to bring the potential organism within the egg to full completion. The evidence of the production of eggs without male help leads to the problem around which the chapter is organized: why does generation require a male at all?

The first four chapters of GA II develop and refine this theory, beginning with a highly abstract explanation of why there is generation at all, and then why the male and female powers are in separate organisms. Generation, Aristotle argues, allows mortal beings to be eternal by maintaining their form in their offspring. Where possible, this is best accomplished by having the efficient cause separated from the material cause, for its potency is best maintained in that way. The process is goal-directed, as the nutritive power of the male parent is transferred by coitus to blood especially prepared to develop into a mature organism once acted on by that nutritive power. It may not be too anachronistic to think of it as a "program" transferred to the material, since Aristotle himself occasionally describes this special heat as having a *logos* (plan?) which determines the sequence of changes transforming a uniform material (prepared blood) into a complex, differentiated living being. After outlining the initial stages of this process in blooded live-bearing animals through the remainder of book II, he turns in book III to those that lay eggs, then to bloodless animals that are nonetheless sexually generated, and finally to animals that are generated "spontaneously," without the interaction of male and female. In III.11 (762a36–b18) he presents a theory for how such generation can occur, which has parallels to his theory of sexual generation: environmental heat is enclosed in a moist, earthen material, producing a simple animate structure which is not itself capable of reproducing.

In book IV, Aristotle explains features that differentiate members of the same species: why do some animals develop into females and others into males? Why do offspring resemble one parent or the other, and why do they sometimes resemble both in different respects? Why is there variation in the number of offspring produced in certain species but not in others? What causes the occasional production of deformed offspring?

Finally, in book V, he considers parts that vary in their features by "more and less" or change during the maturation of the organism, including many that occur simply due to material necessity (such as gray hair or baldness). The first chapter of this book includes a theoretical introduction discussing criteria for determining whether some attribute of an animal is present for the sake of an end or not. This and the passage discussed earlier in PA IV.2 are important reminders that Aristotle is not what would today be called (through Stephen Jay Gould referencing Voltaire) a "Panglossian adaptationist." Not only did he not always seek teleological explanations for every organic attribute, he reflected philosophically on the question of how to determine when such explanations were appropriate.

A CASE STUDY: ARISTOTLE ON RESPIRATION

DATA ORGANIZATION IN THE *Historia Animalium*

Most of the first four books of HA are devoted to presenting information about the *parts* of animals, most of book I being about the external, non-uniform parts of human beings. Aristotle introduces the discussion as follows:

> First let us take up the parts out of which animals are constituted. For it is most of all and primarily in virtue of these that whole animals also differ, either by having or not having a part, or by their position or order, or in virtue of the differences previously mentioned, in form, by degree, and by analogy and by oppositions of the affections. And first we must take up the parts of the human being; for just as every city reckons its currency in relation to the what is best known to itself, so too in other domains; and mankind is from necessity the animal best known to us. (HA I.6, 491a15–22)

He starts with parts in order to clarify the differences between animals; he starts with humankind because this is familiar to us; and he will use the analysis of likeness and difference introduced in the first chapter.

The parts relating to breathing are first discussed in I.16 (495a19–b24), with information from comparative anatomy used in what is allegedly a discussion restricted to human beings; they appear again in II.15, where Aristotle explicitly considers these organs across the *full range* of blooded animal kinds. He first notes the presence in the neck of both windpipe and esophagus, their relative positions in all animals that have them, and the coextensive relation between the windpipe and the lung. He describes the material composition of the windpipe, and notes the existence of the epiglottis; at this stage he does not discuss its function (though he does so in *On the Parts of Animals*, which is more concerned with causal explanations). He next notes that the windpipe extends to the lungs and then branches into two parts (regarding the lung as a single organ with two lobes: 495a33–b5), and that there are branches of the aorta and "great blood vessel" (*vena cava*) that also extend from the heart into the lung. Later (496a28–b7), he discusses the heart's relation to the lung in more detail, concluding that contact between the different kinds of passageways attached to the heart and lung enables breath to be received and conveyed to the heart.

Aristotle next turns to the lung. He claims that among live-bearing animals it is the part that is most suffused with blood, criticizing people who are misled by a lack of visible blood into claiming that the lung is empty (496b1–7). The nature of human internal organs can be inferred from findings in all the other live-bearing animals; the wide-ranging and systematic anatomical dissection of animals is implied.

> So then, as many as are four-legged and live-bearing, all have an esophagus and a windpipe, *positioned in the same manner as in human beings*; and [they are positioned] similarly among the four-legged that are egg-laying, and in the birds, though they differ in the

> forms of these parts. Speaking generally, as many as receive air all breathe in and out, and all have a lung, windpipe and esophagus; and the esophagus and windpipe are positioned similarly [in all of them], but are not alike [in character], while the lung is neither alike [in character] nor positioned similarly. (505b32–506a5)

Aristotle is the first thinker in history to identify these extremely wide-ranging coextensive relationships among biological processes and systems of organs, laying the methodological groundwork for all future zoologists.

He also points out that, while all animals with these respiratory parts have hearts and are blooded, not all blooded animals with hearts have these respiratory parts: "for example, fish do not, nor any others animals there may be that have gills" (506a11–12). This last qualification appears because Aristotle is aware of newts (known as *kordulos*, he says), with four legs and feathery appendages near their head, which he correctly identifies as gills.[15] He is aware that these animals do not have lungs and spend most of their time in water. Aristotle's recognition that these parts play some role crucial to the functioning of the heart will lead him to the fundamental discovery that gills must be analogues of lungs.

Some things are absent from the HA: all language associated with definition, explanation, and causality seems to have been deliberately excluded. In none of the passages I have discussed is any part, or are any of the correlations noted, said to be necessary; nor is any part said to be present for the sake of some end; nor is there any reference to nature acting for an end, nor any use of the distinction between form and matter. It is unlikely this is due to the HA representing an early stage of Aristotle's exploration of nature; David Balme has argued that the work itself (as opposed to the inquiry it reports on) was written *after* causal investigations like those reported in GA, PA, and PN had been completed.[16] It is far more likely that the data are presented and organized as they are because Aristotle saw HA as playing a particular role in the pursuit of scientific knowledge of animals, either in continuing investigation or as an aid to teaching, or both.

EXPLAINING RESPIRATORY PARTS: PARTS OF ANIMALS III.3–6

As well as in HA, Aristotle discusses the parts related to respiration in PA III.3 and 6. While he concludes that the gills of fish have an analogous

[15] Cf. HA VII.2, 589b26–7; *On Respiration* 10, 476a6–7; PA IV.13, 695b25. He may have been thinking of the larval stage of some sort of salamander, such as *Triturus macedonicus*, which is endemic to Macedonian Greece. A number of these species have members of their populations that are paidomorphs, that is, they stay in the larval, gilled form their entire life.

[16] See Aristotle, *History of Animals Books VII–X*, ed. and trans. Balme, pp. 21–5; his argument and evidence are reviewed in James G. Lennox, "Aristotle's Biological Development: The Balme Hypothesis," in W. Wians (ed.), *Aristotle's Philosophical Development: Problems and Prospects* (Savage, MD: Rowman and Littlefield, 1996), pp. 229–48.

function to the lung in the remainder of the blooded animals,[17] he restricts the term "respiration" to the function performed by taking in and expelling air. He identifies five primary structures related to it: the lung; the windpipe (sometimes divided into the bronchial tubes, larynx and trachea); the esophagus; and the epiglottis. Here, his goal is to explain why these parts are present in the animals that have them: lung and windpipe are present *for the sake of* respiration; the esophagus is *necessary* for transporting food, though not to aid in the act of digestion; and the epiglottis is required because of the necessity of having the windpipe in front of the esophagus.

In the following passage, explaining why animals have the parts related to respiration, we see again the complex way in which final, efficient, and material causality are intertwined (I have highlighted the causal/explanatory language):

> Not all animals have a neck, but only those with the parts *for the sake of which* it is naturally present; and these are the larynx and the part called the esophagus. The larynx is present *by nature for the sake of breathing*; for through it animals draw in and expel breath when they inhale and exhale. *This is why* those without a lung do not have a neck, e.g. the kind consisting of fish. The esophagus is that through which nourishment proceeds to the gut; so that all those without necks manifestly do not have an esophagus. But it is *not necessary* to have the esophagus *for the sake of nutrition*; for it prepares nothing for nutrition . . . And the esophagus is fleshy, with a sinuous elasticity – sinuous *so that* it may dilate when food is ingested, yet fleshy *so that* it is soft and yielding and is not damaged when scraped by the food going down . . . The windpipe, by being positioned . . . in front of the esophagus is interfered with by the food; but *for this, nature has constructed the epiglottis*. Not all the live-bearing animals have this part, but rather all those that have a lung and hairy skin, and that are naturally neither hard-scaled nor feathered, do. In those that are hard-scaled or feathered, in place of the epiglottis the larynx contracts and opens . . . These animals do not have an epiglottis *because their flesh is dry and their skin hard, so that* in them a part of this sort, constituted from flesh and skin of this sort, would not move easily. (Selections from PA III.3, 664a14–665a3)

Teleological explanations are here provided for the neck, lung, and windpipe; the esophagus is necessary because in animals that breathe there is a significant distance between mouth and stomach. All these animals must solve the "design flaw" created by having the windpipe in front of the esophagus, but their *material nature* determines how they do so: animals

[17] PA I.5, 645b5–10; III.6, 669a3–6; how he determined that can be inferred from the way his discussion of respiration develops in the concluding section of the *Parva Naturalia*.

with soft, moist flesh develop an epiglottis, while those with hard, dry flesh simply collapse the opening of the windpipe. While in that case material differences do the explaining, the sinuous and fleshy nature of the esophagus is explained by its purpose. Finally, since fish use gills and water for cooling rather than lungs and air, they have no need for windpipe, esophagus, or neck.

Why should the larynx/windpipe be placed in front of the esophagus in the first place? Aristotle says that is a design flaw, and then explains:

> Let it be assumed, then, that we have stated the following things: *the cause owing to which* some animals have an epiglottis and others do not, and *why nature has remedied* the inefficiency of the position of the windpipe by constructing the part called the epiglottis. The larynx lies in front of the esophagus *of necessity*. For the heart . . . lies in front and in the middle . . . and the lung lies where the heart is, i.e. surrounding it, and *respiration takes place both on account of this and on account of the source* [of nutrition, perception and locomotion] *being present in the heart*. But respiration comes about in animals through the windpipe; so, *since it is necessary* that the heart be placed first among things in the front, it is *also necessary* that the larynx and windpipe be placed in front of the esophagus. (Selections, PA III.3, 665a10–22)

All of these structures are present in animals because these animals must breathe to survive – but *why* do animals need to breathe? What is it for?

EXPLAINING THE ACTIVITY OF RESPIRATION: ON RESPIRATION

Aristotle discusses respiration as a physiological activity in the final chapters of what has come down to us as the *Parva Naturalia*, the *Small Natural Studies*, chapters sometimes identified independently as *On Respiration*.[18] He there argues that respiration is present for cooling. The heart generates a great deal of heat to carry out its various functions, which for Aristotle include generation, nutrition, perception, and locomotion. Indeed he thinks the heart generates sufficient heat that the area around it needs to have that heat moderated. As the lobes of the lung are full of blood vessels coming from the heart and tubes for air coming from outside, this allows the air to cool the blood and carry off its heat. In the absence of this exchange, the heat would soon consume the nutrients needed by the animal, and the animal would die. (*Resp*. 13, 15–17, 20)

Respiration is introduced in the next lines of PN, traditionally marked as the opening lines of *On Youth and Old Age*:

[18] The Prussian Academy edition (I. Bekker (ed.), *Aristotelis Opera* (Berlin: Georg Reimer, 1831), on which the standard pagination of Aristotle's works is based, begins to renumber the chapters of *De Juventute* (*On Youth and Old Age*) when the discussion of respiration begins, breaking up what is clearly intended as a continuous discussion of "youth, old age, life, and death."

> We must now discuss youth, old age, life and death; and presumably at the same time it is also necessary to discuss the causes of respiration; for in some animals living and not living are dependent on this. (PN, 467b10–13; cp. 472b24–9)

This much is an unquestionable datum of common sense, which gives one good reason to investigate it from a teleological perspective: it is clearly doing *something* important for life, but *what*, precisely? Because Aristotle's answer is that it promotes cooling in the region around the heart, investigating why breathing animals breathe is transformed gradually into an investigation of methods of *cooling* across the animal kingdom, of the need for a *variety* of methods of cooling, and of the importance of effective cooling for the preservation of life. That discussion will depend heavily on the central topic of *On Youth and Old Age*: the intimate relationship between the nutritive soul and internal, natural, or psychic heat – and the physical changes in old age that lead to natural heat no longer serving as an effective agent of nutritive soul.

On Respiration, then, is embedded within these wider concerns. It opens with a long discussion of earlier views, during which Aristotle repeatedly criticizes his predecessors for failing to focus on the function respiration serves.[19] His message is that you will not succeed in figuring breathing out at all, or even getting the material or mechanical story right, if your investigation is not guided by questions regarding function.

The hearts of all blooded land animals employ a natural heat in preparing blood for distribution to the rest of the body as its nourishment. The need for that heat underlies the need for respiration. Aristotle's argument begins by focusing almost exclusively on the bloodless animals, arguing that they can be adequately cooled by their surroundings and so do not need organs of respiration. But the chapter also looks forward, making *comparative* comments about animals that cool themselves by gills or lungs. Next it specifies as precisely as possible the *extension* of the relevant attributes, in a form reminiscent of HA.

Groups are carefully correlated based on differentiae relevant to how cooling is accomplished. The widest background group of interest is now the *blooded* animals, all of which have a heart. Among them, the group with a lung is coextensive with those that cool themselves by inhaling and exhaling air.

Aristotle goes on:

> But *as many as* have gills, *all* cool themselves by taking in water; both the kind consisting of the so-called selachians and the other footless

[19] Contra Democritus, 471b30–472a2; contra Plato, 472b24–9; contra Empedocles, 473a15–17.

> animals.[20] And all the fish are footless; and indeed what part they have [for locomotion] is named for its likeness to wings.[21] (PN 476a2–5)

He then mentions that, of those so far studied, only one animal with feet also has gills, "the so-called kordulos" (see footnote 15, above). But he notes that it has no lung, and in fact, "[n]o animal has yet been seen possessing both lungs and gills," because:

> the lung is for the sake of cooling by breath . . . while the gills are related to cooling by means of water; and there is one instrument for one use, and one is sufficient for cooling in all cases. So since we see nature doing nothing in vain . . . but one would be in vain, on this account some animals have gills, some lungs, but none have both. (PN, 476a7–15)

The kordulos suggests that gills perform in the water the function that lungs perform in the air.

Aristotle next announces that he must speak "about *cooling*, in what way it comes about both for those that breathe and for those having gills" (477a11–12); the shift to this broader question is a consequence of his having focused, from the start, on the *purpose* of respiration and thus being poised to understand the analogical identity between lungs and gills. Cooling is a central theme of the remainder of the treatise, intertwined with discussions of the topics of generation, life and death, and the movements of the heart and blood vessels.

The evidence of dissection, which reveals tubes of a sinu-vascular character passing from the top of the heart to the gills (478b7–15), and other related blood vessels, allows Aristotle to conclude that "on account of their nature being aquatic, [fish] produce cooling by the water passing through the gills." Finally, there appears to be a reference to the "fanning" of the gills, analogous to inhaling and exhaling.

Aristotle elaborates upon these analogies in chapter 21, once again noting the reason why respiration should be discussed as part of an investigation of youth, old age, life, and death:

> the air enters into many ducts in the lungs, like channels, and alongside each of [these ducts] blood vessels are extended, so that the whole lung seems to be filled with blood. And we call the inhaling of the air "respiration" and its exhaling "expiration." And this always takes place continuously, and the animal lives so long as this part moves continuously. And for this reason life depends on respiration and expiration. And in the same way motion arises in the gills of fish.

20 Footless creatures such as snakes or cetaceans are excluded, as it is only footless animals with gills that are under consideration.

21 The Greek for "fin" is *pterugion*, the plural for which is *pterugia*, and for "wing" *pteruks*.

> For when the heat in the blood throughout the parts increases, the gills also rise up and water flows through [them]. And when water is allowed to flow in relation to the heart through these channels, the gills are also cooled and collapse, and the water is expelled. Thus whenever the heat of the heart increases water is always taken in, and when it is cooled again the water flows out. For this reason too the final authority concerning living and not living for those with lungs rests with breathing, and for those with gills with the taking in of water. (PN, 480b6–20)

Earlier her notes a further analogy between gills and lungs:

> And breathing animals are suffocated in a small amount of air that remains the same; for each of them rapidly becomes hot since contact with the blood heats each animal. And being hot, the blood prevents cooling; and when, owing to sickness or old age, the lung in the case of breathing animals, or the gills in the case of water-dwelling animals, is unable to move, at that point death results. (PN, 478b15–21)

By pursuing comparative anatomy with a focus on the goal of respiration, Aristotle established that its function, cooling the region where the heart concocts the blood into its final, nutrient form, is a function shared by all blooded animals, including fish. And that led to the discovery that, in animals with gills, their structure, their vascular connections to the heart, their movements, and finally their age-related pathology, make them analogous to the lungs in all the other blooded animals.

CONCLUSION

Aristotle's zoology is organized in broad outlines in accordance with his general views about scientific inquiry and scientific knowledge. To understand some general truth about a group of animals, such as that all animals that breathe have lungs and a windpipe, one must identify the cause of that general truth; and a scientific explanation of such a general truth consists of a syllogistic proof of it in which the term shared by its premises identifies the cause. For example: having established in a "history" of animals that breathe that they all have lungs and a windpipe, one seeks an explanation for why this is so. In Aristotle's study of respiration he argues that breathing is for the sake of moderating the heat generated by the heart, and thus there is a deeper teleological explanation for breathing. And, since there are animals with hearts that do not have lungs, he arrived at a new question: if all hearts need cooling, how is this cooling achieved in fish, which have no lungs and do not breathe?

We have now seen Aristotle's characteristic approach to such issues. It is not surprising that some of the greatest biologists of the nineteenth century – Cuvier, Owen, and Darwin – were deeply impressed by his zoological treatises. They represent one of the greatest achievements in the history of biology, all the more so for being the very first attempt at a systematic science of animals.

13

BOTANY

Laurence Totelin

Studies of Greek and Roman botany have traditionally adopted a chronological approach, with a focus on important contributors to the field such as Theophrastus of Eresus (372/1–287/6 BCE), author of *Enquiry into Plants* and *Causes of Plants*; Dioscorides of Anazarbus (first century CE), compiler of *Materia Medica*; Pliny the Elder (23–79 CE) and his *Natural History*, whose books 12 to 25 focus on plants and medicaments derived from them; and pseudo-Aristotle with his *On Plants* – in fact the work of Nicolaus of Damascus, a first-century CE Aristotelian philosopher, who transmitted some Peripatetic theories on plants that may have been extracted from one of Aristotle's writings. Mention might also be made of the Roman agronomists M. Porcius Cato (234–149 BCE), M. Terentius Varro (116–27 BCE), and L. Iunius Moderatus Columella (first century CE), and of various botanical writers whose works have been lost, such as Crateuas the Root-Cutter (second/first century BCE), who collaborated with Mithradates VI, king of Pontus on botanical projects, and Sextius Niger (early first century CE), an important source for both Dioscorides and Pliny.[1]

[1] For a recent work on ancient botany, see Gavin Hardy and Laurence Totelin, *Ancient Botany* (London: Routledge, 2015). For traditional approaches to ancient botany, see e.g. Edward L. Greene and Frank N. Egerton, *Landmarks of Botanical History*, 2 vols. (Stanford, CA: Stanford University Press, 1983), vol. 1. The bibliography on Theophrastus' botany is extensive. Most important is Suzanne Amigues' work, e.g. *Études de botanique antique* (Mémoires de l'Académie des Inscriptions et Belles-Lettres, 25; Paris: De Boccard, 2002). On Dioscorides, see in particular Lily Y. Beck, *Pedanius Dioscorides of Anazarbus: De materia medica* (Altertumswissenschaftliche Texte und Studien, 38; Hildesheim: Olms-Weidmann, 2005); John M. Riddle, *Dioscorides on Pharmacy and Medicine* (History of Science Series, 3; Austin, TX: University of Texas Press, 1985). On Pliny's botany, see A. G. Morton, "Pliny on Plants: His Place in the History of Botany," in Roger French and Frank Greenaway (eds.), *Science in the Early Roman Empire: Pliny the Elder, His Sources and Influence* (London: Croom Helm, 1986), pp. 86–97. On Nicolaus of Damascus, see Hendrik J. Drossaart Lulofs and E. L. J. Poortman, *Nicolaus Damascenus De plantis: Five Translations* (Verhandelingen der Koninklijke Nederlandse Akademie van Wetenschappen. Afd. Letterkunde, Nieuwe reeks, 139; Amsterdam: North-Holland Publishing Company, 1989). All references to *De plantis* in this article are to the Arabic translation, which is the central text in the complex transmission of Nicolaus' botanical work. Information and bibliography on all other authors listed

It is undeniable that these figures helped systematize the study of plants, but such an approach tends to obfuscate the fact that botanical knowledge was not the preserve of highly literate writers in the ancient world – a fact of which these authors were fully aware. Thus, Pliny wrote in negative tones:

> But the reason why more herbs are not known is because those who experience them are illiterate folk people, who are alone in living among them. (*Naturalis historia* (henceforth "*HN*") 25.16)

In a more positive fashion, Theophrastus and Dioscorides acknowledged as sources various non-literate craftspeople (see below, p. 240).

Botanical knowledge was spread through Greek and Roman society, from the simplest country-women to the highest kings, and people at all levels of the social scale could contribute valuable information on useful plants. Indeed, utility emerges as the most important characteristic of plants in all ancient writings that deal with them, whether philosophical (e.g. Theophrastus and Nicolaus of Damascus), encyclopedic (e.g. Pliny the Elder), medical (e.g. Dioscorides and Galen), agronomical (e.g. Varro and Columella), or even poetic (e.g. Theocritus and Nicander). Drawing a distinction between "pure" botany (i.e., the study of plants for their own sake) and "applied" botany (i.e., the study of plants for practical purposes) is therefore almost meaningless for the ancient world. In fact, discerning a body of clearly "botanical" works in Greek and Latin is very difficult. The ideal approach would be to include all material that deals in one way or another with plants – a monumental task if there is one.[2]

In this chapter, in order to avoid some of the pitfalls of a chronological approach centered on key figures, I have adopted a thematic approach. I will discuss how the ancients gained knowledge of plants; how they classified, named, described, and depicted plants; and how they conceived of their life cycles. I will conclude with some thoughts on the utility of plants in the Greco-Roman world.

here can be found in Paul T. Keyser and Georgia L. Irby-Massie (eds.), *The Encyclopedia of Ancient Natural Scientists: The Greek Tradition and its Many Heirs* (London and New York: Routledge, 2008).

[2] On the contribution of kings to the study of botany, see Laurence M. V. Totelin, "Botanizing Rulers and Their Herbal Subjects: Plants and Political Power in Greek and Roman Literature," *Phoenix* (2012), 122–44. On Galen's pharmacology, see Sabine Vogt, "Drugs and Pharmacology," in R. J. Hankinson (ed.), *The Cambridge Companion to Galen* (Cambridge: Cambridge University Press, 2008), pp. 304–22. On medical botany more generally, see John Scarborough (ed.), *Pharmacy and Drug Lore in Antiquity: Greece, Rome, Byzantium* (Variorum Collected Studies Series, 904; Farnham: Ashgate, 2010). On Nicander, see the works of Jean-Marie Jacques, e.g. "Situation de Nicandre de Colophon," *Revue des études anciennes* 109 (2007), 99–121. On Theocritus, see Suzanne Amigues, "De la botanique à la poésie dans les 'Idylles' de Théocrite," *Revue des études grecques* 109 (1996), 467–88 A distinction between "pure" and "applied" botany is drawn by John Scarborough, "Botany," in Simon Hornblower and Anthony Spawforth (eds.), *Oxford Classical Dictionary* (Oxford: Oxford University Press, 1996), pp. 255–6.

ACQUIRING KNOWLEDGE OF PLANTS IN THE ANCIENT WORLD

In his work on agriculture, Varro (*De re rustica* (henceforth "*Rust.*") 1.1) explains how he acquired his knowledge from three sources: his personal experience; what he had read; and what he had heard. Most authors writing on plants in the Greco-Roman world could have claimed to have drawn on these three sources, albeit in varying proportions.

Personal experience was the most reliable means to acquire plant knowledge in the ancient world. Roman agronomists could claim to have first-hand experience of growing and caring for plants (in particular cereals, vines, and olive trees) on their farms; ancient pharmacological writers (such as Dioscorides, Galen, or the root-cutters mentioned by Theophrastus) had expertise in gathering plants for medical preparations; others traveled to observe for themselves plants in their original environment. Thus, Dioscorides reports that he traveled extensively thanks to his "military life" (preface 4), and Theophrastus too journeyed through the Greek world and observed plants in their original habitat (e.g. plants on Mount Ida, mentioned regularly by Theophrastus in the *Enquiry into Plants*). Ancient botanists, however, could not have observed themselves all the plants they describe – they sometimes had to rely on oral or written reports made by others.[3]

In their travels, ancient botanists collected oral reports from individuals belonging to named ethnic or professional communities. These stories, often introduced in writing with phrases such as "they say" or "it is said," are a rich source of information on the practices of, for instance, peasants (e.g. Theophrastus, *Historia plantarum* (henceforth "*Hist. pl.*") 2.6.2: on the cultivation of palm trees), makers of musical instruments (e.g. Theophrastus, *Hist. pl.* 4.11.4: on the use of reeds to make musical instruments), perfumers (e.g. Dioscorides 1.20: on the use of camel's-thorn as a fixative), wreath-makers (Dioscorides 3.75: on the use of rosemary), root-cutters, and drug-sellers (both Theophrastus, *Hist. pl.* 9.8.5), all people who were directly involved in the handling of plants. These oral accounts can also inform us on local practices, for instance the use of ground pine as an antidote by the people of Pontic Heraclea (Dioscorides 3.158.2) or the recourse of Thessalian women to *orchis* to increase sexual desire (Dioscorides 3.126.2).

[3] On olive, cereal, and vine cultivation in the ancient world, see Marie-Claire Amouretti and Jean-Pierre Brun (eds.), *La production du vin et de l'huile en Méditerranée* (Bulletin de correspondance Hellénique. Supplément, 26; Athens: École française d'Athènes; Paris: De Boccard, 1993); Peter D. A. Garnsey, *Food and Society in Classical Antiquity* (Key Themes in Ancient History; Cambridge: Cambridge University Press, 1999); Lin Foxhall, *Olive Cultivation in Ancient Greece: Seeking the Ancient Economy* (Oxford: Oxford University Press, 2007). On the preface of Dioscorides, see John Scarborough and Vivian Nutton, "The Preface of Dioscorides' *Materia Medica*: Introduction, Translation, Commentary," *Transactions and Studies of the College of Physicians of Philadelphia* 4 (1982), 187–227.

They tell, among other things, of rituals involved in the gathering of plants (for instance, avoiding the use of iron in collecting a plant (see Pliny, *HN* 24.103), or gathering during a particular phase of the moon (see Pliny, *HN* 19.113)). These tales, however, should be treated with caution, as much can be lost or added in the translation from the oral to written medium.[4]

The reliance on written material, unsurprisingly, increases with the passing of time: Varro, Columella, Pliny, Dioscorides, and Galen provide long lists of authors from whom they have borrowed; whereas the number of written authorities named by Theophrastus is much shorter (it includes the poets Homer, Hesiod, and Musaeus). Lists of written sources, however, should be treated with caution: these could themselves be borrowed and used as a means to increase the prestige of a particular branch of knowledge. Conversely, ancient botanical writers sometimes omitted to name their sources. For instance, Theophrastus does not name his sources for the information he transmits on Persian (e.g. *Hist. pl.* 4.7.5–6) and Indian (*Hist. pl.* 4.4.4–11) vegetation. From parallel passages in the historical works of Arrian (ca. 86–160 CE), *Anabasis* and *Indica*, we can infer that Theophrastus obtained this information from scholars, such as Aristobulus and Nearchus, who had accompanied Alexander the Great on his conquests.[5]

Indeed conquest, colonization, and expedition to foreign lands were significant means of acquiring botanical knowledge in the ancient world. Conquering territories that grew such precious plants as silphium or myrrh could mean significant economic returns, and there are – dubious – reports that Alexander intended to conquer Arabia for the sake of its famous spices (Arrian, *Anabasis* 7.20; Pliny, *HN* 12.62). Fabulous stories relating to exotic plants, found in the writings of numerous historians and geographers (see e.g. the stories relating to the gathering of Arabian aromatic herbs in Herodotus 3.107–12), were highly entertaining but also hinted at the real economic value of these products.[6]

[4] On the contribution of non-literate sources to ancient botany, see Geoffrey E. R. Lloyd, *Science, Folklore and Ideology: Studies in the Life Sciences in Ancient Greece* (Cambridge: Cambridge University Press, 1983), pp. 119–49; John Scarborough, "The Pharmacology of Sacred Plants, Herbs, and Roots," in C. A. Faraone and Dirk Obbink (eds.), *Magika Hiera: Ancient Greek Magic and Religion* (New York: Oxford University Press, 1991), pp. 138–74. On rituals involved in gathering plants, see Guy Ducourthial, *Flore magique et astrologique de l'antiquité* (Paris: Belin, 2003).

[5] Varro gives his list of sources at *Rust.* 1.1.8–10; Columella at *Rust.* 1.1.4–14; Pliny lists his sources in book 1 of his *Natural History*, which functions like a table of contents; Galen lists his sources in various places in his pharmacological books (*Composition of Medicines according to Types; Composition of Medicines according to Places; Simple Drugs; Antidotes*); Theophrastus' references to his sources are spread through his works. On Theophrastus' sources, see Théophraste, *Recherches sur les plantes, vol. 1: Livres 1–2*, ed. and trans. Suzanne Amigues (Paris: Les Belles Lettres, 1988), pp. xx–xxx. That lists of sources could be borrowed can be inferred from the fact that the same mistakes can be found in two separate lists.

[6] On expeditions and botanical knowledge, see Suzanne Amigues, "L'expédition d'Anaxicrate en Arabie occidentale," *Topoi Orient-Occident* 6 (1986), 671–7, reprinted in her *Études de botanique antique*, pp. 57–62. See also Hugo Bretzl, *Botanische Forschungen des Alexanderzuges* (Leipzig: Teubner, 1903).

CLASSIFYING, NAMING, DESCRIBING, AND DEPICTING PLANTS

There were in the ancient world as many ways of classifying plants as there were botanical writers. Theophrastus discussed his principles in the first book of *Enquiry into Plants*. Following Aristotle's method of biological classification, Theophrastus based his system of plant classification on differences (*diaphorai*): differences in cultivation status (domestic or wild); in habitat (earth or water); in fertility (barren or fruitful); whether the plant is evergreen or deciduous, flowering or not; but above all differences in parts (stems, roots, flower, etc.) (*Hist. pl.* 1.14.3). His attention to parts led him to divide the vegetable kingdom into trees, shrubs, under-shrubs, and herbs. The philosopher, however, eschewed all systems of classification that are too rigid, arguing that classification can only be "typical" (*Hist. pl.* 1.3.5).[7]

Other authors developed systems of classification that were meant to be useful, to guide their readers in finding information. Thus, Dioscorides developed a system of classification of *materia medica* (most of which are plants) based on their properties, that is, on the effect the ingredient had on a patient. Dioscorides was proud of his system, which he developed through experience, criticizing the principles of some of his predecessors, and in particular alphabetical classification (preface 3). Modern studies have shown how Dioscorides' system of classification reflects deep knowledge of the working of plants on the human body. To his contemporaries, however, the benefits of Dioscorides' system were not apparent. Indeed, Galen, who had a high opinion of Dioscorides' work and used it as a source for his pharmacological works, had recourse to an alphabetical system of organization in his *Simple Drugs*. At some point in late antiquity, probably in the fourth century CE, Dioscorides' *Materia Medica* itself was alphabetized, as the manuscript tradition testifies: for instance, the famous Vienna Dioscorides manuscript (see below, p. 244) presents pharmacological materials in alphabetical order.[8]

Pliny's system of classification too was most probably meant to be useful to his readers, allowing them quickly to find information they needed, for instance, on the vine (book 14), the olive tree (book 15), or cereals (book 18) – topics that were typically discussed by Roman agronomists. Pliny's system, however, also integrated some of Theophrastus' more complex principles whereby plants are defined by their parts. Indeed, the encyclopedist's discussion of plants follows Theophrastus' system of classification in that it first

[7] On Theophrastus' system of classification, see Suzanne Amigues, "Problèmes de composition et de classification dans l'*Historia plantarum* de Théophraste," in J. M. van Ophuijsen and M. van Raalte (eds.), *Theophrastus: Reappraising the Sources* (Rutgers University Studies in Classical Humanities, 8; New Brunswick, NJ and London: Transaction Publishers, 1998), pp. 191–202, reprinted in her *Études de botanique antique*, pp. 45–54; Jacques Desautels, "La classification des végétaux dans la *Recherche sur les plantes* de Théophraste d'Erésos," *Phoenix* 42 (1988), 219–43.

[8] On Dioscorides' method of classification, see Riddle, *Dioscorides on Pharmacy and Medicine*, chapter 1.

covers trees (books 12 to 17), followed by cereals, shrubs, herbs, and flowers and the medicines derived therefrom (books 18 to 22), and ends on a discussion of drugs obtained from trees (books 22 to 24) and a miscellaneous book on "the natures of self-grown plants" (book 25).

Beside the classificatory systems devised by literate authors, there existed a folk system of classifications of plants into "female" and "male" plants (Theophrastus, *Hist. pl.* 3.8.1). Within that system, male plants were considered harder, stronger, more knotty than female plants, which were softer, easier to work, and smoother.[9] Whatever the merits of these various classification systems, they never led to – nor were they underpinned by – an established plant nomenclature. Ancient plant nomenclature was extremely unstable. The same plant could have several different names (hence the existence of numerous lists of plant synonyms both within botanical works and as an independent genre); and the same name could be used for several different plants. Ancient botany writers did not name plants themselves; they reported appellations that they had learnt from their sources (country people, farmers, etc.). Some plants did not even have names: this was particularly the case of wild plants (see Theophrastus, *Hist. pl.* 1.14.4). When plants were named, they often took their name from a prominent morphological characteristic, including resemblance to an animal (e.g. the scorpion plant, Pliny, *HN* 25.122); after the person (real or mythological) who had allegedly discovered them (e.g. the gentian named after a certain king Gentius; Pliny, *HN* 25.71); after their habitat (e.g. water lilies, in Greek *nymphaia*, thus named because, like Nymphs, they lived in water); or after their efficacy (e.g. the *aristolochia*, literally "good birth," thus called because it was believed to help women in childbirth; Dioscorides 3.4.1). To the main name of the plant, an epithet (what modern botanists would call a "modifier") could be added: geographical epithets were particularly frequent, especially for fruits. Thus, the Greek and Latin names of our peach and apricot were respectively Persian and Armenian apple.[10]

A name (or names) was seen as an integral part of a plant's description, which would have also included a botanical description, information on the plant's habitat, its uses, and its methods of harvesting, preparation, and storage. It is difficult to identify plants from ancient botanical descriptions,

[9] See M. Negbi, "Male and Female in Theophrastus' Botanical Works," *Journal of the History of Biology* 28 (1985), 317–32.

[10] There are numerous studies of ancient nomenclature: for instance *Actes du colloque international "Les phytonymes grecs et latins" tenu à Nice les 14, 15, et 16 mai 1992 à la Faculté des Lettres, Arts et Sciences humaines, Université de Nice-Sophia Antipolis* (Nice: Centre de recherches comparatives sur les langues de la Méditerranée ancienne, 1993); Jacques André, *Les noms de plantes dans la Rome antique* (Paris: Belles Lettres, 1985); James L. Reveal, " What's in a Name: Identifying Plants in Prelinnaean Botanical Literature," in Bart K. Holland (ed.), *Prospecting for Drugs in Ancient and Medieval European Texts* (Amsterdam: Harwood Academic, 1996), pp. 57–90. On lists of synomyms, see e.g. Willem F. Daems, *Nomina simplicium medicinarum ex synonymariis Medii Aevi collecta: semantische Untersuchungen zum Fachwortschatz hoch- und spätmittelalterlicher Drogenkunde* (Series in Ancient Medicine, 6; Leiden: Brill, 1993).

as botanical writers tended to describe a plant by comparison with another plant instead of using a fixed set of description criteria. On the other hand, the ancients took care to describe habitat: they realized that some plants could only grow under certain environmental conditions, as numerous stories of failed transplantations show (e.g. the failed transplantation of laurel and myrtle to Panticapaea: Pliny, *HN* 16.137; Theophrastus, *Hist. pl.* 4.5.3). Clues as to the identity of an ancient plant can also be found in the account of its usages. Indeed, Greek and Roman botanical writers devoted much space to describing how and why to use a plant. Plants were thought to exist for the benefit of animals, of which man was the highest species, as expressed by Nicolaus of Damascus: "Plants are created for the sake of animals only, and animals are not created for the sake of plants."[11] Their utility for men extended to all aspects of human life, from the building of dwellings to the healing of dangerous diseases. Even dangerous, poisonous plants could be useful, either as medicaments if used in the right dose, or as poisons to destroy enemies.

Ancient written descriptions of plants were sometimes accompanied by illustrations. Pliny (*HN* 25.8) informs us that authorities such as Crateuas, Dionysius, and Metrodorus used painted likenesses of plants in their medical herbals, a practice that was not without pitfalls as painters could only represent the plant in one of its growth aspects, and copyists made mistakes when reproducing the paintings. Fragments of such illustrated herbals have survived on papyri. Thus, two plant illustrations (*phlomos* and *symphyton*) are preserved on the "Antinoopolis illustrated herbal" (P.Johnson + P.Ant. 3.214), a codex leaf from a herbal copied around 400 CE. The earliest complete manuscript of an illustrated herbal is the Vienna Dioscorides (Vienna, Österreichische Nationalbibliothek, Cod. med. gr. 1), a magnificent work produced around 512 CE for the benefit of the princess Julia Anicia.[12]

THE LIFE OF A PLANT

Greek philosophers debated as to what type of living beings (animals) plants were. Some Presocratic philosophers, such as Empedocles and Anaxagoras

[11] Nicolaus of Damascus, *De plantis: On Plants*, ed. and trans. Hendrik J. Drossaart Lulofs and E. L. J. Poortman (Amsterdam and Oxford: North-Holland, 1989), p. 142.

[12] On botanical illustrations, see Minta Collins, *Medieval Herbals: The Illustrative Traditions* (British Library Studies in Medieval Culture; London: British Library, 2000); Daniela Fausti, "Erbari illustrati su papiro e tradizione iconografica botanica," in Isabella Andorlini (ed.), *Testi medici su papiro. Atti del Seminario di Studio (Firenze, 3–4 giugno 2002)* (Florence: Istituto Papirologico 'G. Vitelli', 2004), pp. 131–50; David Leith, "The Antinoopolis Illustrated Herbal (PJohnson + PAnt. 3.124 = MP3 2095)," *Zeitschrift für Papyrologie und Epigraphik* 156 (2006), 141–56; Charles Singer, "The Herbal in Antiquity and its Transmission to Later Ages," *Journal of Hellenic Studies* 47 (1927), 1–52; Riddle, *Dioscorides on Pharmacy and Medicine*, chapter 5.

(whose views are preserved in Nicolaus of Damascus' *On Plants*),[13] considered plants to be endowed with sensations (pain and pleasure) and even feelings (joy and sadness). Plato, in the *Timaeus* (77a–c), argued that plants were living beings partaking of the third type of soul (the appetitive soul), "having no part in opinion or reason or mind, but only in feelings of pleasure and pain and the desires which accompany them." Aristotle (*De anima* 410b24 ff.) endowed plants with some faculties of the soul (nutrition, growth, and generation) but not motion or perception. Theophrastus too described these functions in plants, but insisted far less than Aristotle on the concept of soul. In many ways, Theophrastus' botanical project was the continuation of Aristotle's study of animals (see chapter 12 in this volume), but Theophrastus showed his independence in this lack of interest in vegetative soul and in teleology.[14]

Whatever the philosophical position taken by botanical writers in this debate, a great anthropomorphism is observable in all ancient works on plants, and in particular in the description of vegetable life-cycle. Thus, the four first books of Theophrastus' *Causes of Plants* follow the life of a plant from generation through growth to diseases and death, and they use concepts and vocabulary borrowed from human biology. The question of plant generation puzzled the ancients; they debated whether plants contained both female and male principles in them (Aristotle, *De generatione animalium* 730b33–731a4), or whether the earth acted as a female element (Theophrastus, *De causis plantarum* (henceforth *Caus. pl.*) 4.4.9–10). They discerned several modes of plant generation, which can all be seen as forms of "cloning" rather than sexual reproduction: "spontaneous, from the seed, from a root, from a shoot, from a branch, from a twig, from the trunk itself, or even from wood broken into small pieces" (Theophrastus, *Hist. pl.* 2.1.1). While reproduction by seed was considered the simplest means of plant reproduction, spontaneous generation exercised the mind of Greco-Roman botanists, with particular attention paid to the example of silphium, a plant which allegedly appeared in Cyrene (Libya) a few years before the arrival of Greek settlers after a pitch-like rain (Theophrastus, *Caus. pl.* 1.5.1; *Hist. pl.* 3.1.6; Pliny, *HN* 19.41). Plant propagation through various grafting methods was discussed in detail by Roman agronomists, who noted that new grafting techniques could be discovered at all times. Thus Columella had invented a way to combine trees of different natures such as figs and olives (*Rust.* 5.11.12–15; see Pliny, *HN* 17.137).[15]

[13] Nicolaus of Damascus, *De plantis: On Plants*, ed. and trans. Drossaart Lulofs and Poortman, p. 126.

[14] On Theophrastus and teleology, see Roger French, *Ancient Natural History. Histories of Nature* (London: Routledge, 1994), pp. 89–92.

[15] On plant reproduction, see D. M. Balme, "Development of Biology in Aristotle and Theophrastus: Theory of Spontaneous Generation," *Phronesis* 7 (1962), 91–104; Negbi, "Male and Female"; C. G. Torzten, "Male and Female in Peripatetic Botany," *Classica et Mediaevalia* 42 (1991), 81–110.

Once they had been generated, plants grew, some slower than others, until they became able to produce buds, flowers, and fruits (the plant's raison d'être) – processes which were compared to pregnancy and birth in animals (Theophrastus, *Caus. pl.* 1.6.3; Pliny, *HN* 16.94). This sexualized conception of plants should not, however, lead us to believe that the ancients had any understanding of pollination and plant sexuality. Even though the ancients had observed that some plants shed a type of dust (wild figtrees: Theophrastus, *Hist. pl.* 2.8.3), that some insects could bring forth fructification in trees (a type of wasp in the case of wild figtrees), and that flower gardens attracted bees (Pliny, *HN* 21.70), they did not come anywhere near an understanding of the process of pollination.[16]

Throughout their lives, plants could suffer from diseases which could lead to sterility, weakening, or death. To name and describe these plant ailments and their causes, botanical writers borrowed vocabulary from treatises on human diseases, a fact which Pliny (*HN* 17.218) noted. In the case of diseases as in other aspects of plant life, Pliny stressed the similarity between plants and animals and humans, as in the following example:

> Sometimes trees die from contagious diseases according to their kind, just as among humans, diseases sometimes attack slaves, and sometimes the urban or rural population. (*HN* 17.219)

If plants had not been killed by illnesses, they could live to old age, when they grew faster, their bark became more wrinkled, and they gave out a stronger scent (Pliny, *HN* 16.126; 21.36). The life of a plant ended with death, the cause of which could be either natural or unnatural according to Theophrastus (*Caus. pl.* 5.11.1).

CONCLUSIONS: USEFUL PLANTS

Ancient philosophers might have written in the most abstract of terms about the status of plants amongst other living beings ("animals"), but even in their works, the utility of plants to higher animals, and in particular to humans, is stressed. Plants existed to feed, heal, offer shelter, clothe, poison enemies, and delight humans. The ancients had no conception that the vegetable kingdom had its own rules and an independent existence from other animals. In this botanical context where plant utility was paramount, discovering new useful plants and working out how they could enhance the life of man was not the preserve of a few "scientific" authors; it was spread

[16] The fructification of the figtree, often compared to that of the palmtree, was much discussed in antiquity, starting with Herodotus 1.193. See, L. Georgi, "Pollination Ecology of the Date Palm and Fig Tree: Herodotus 1.193.4–5," *Classical Philology* 77 (1982), 224–8; Negbi, "Male and Female"; Torzten, "Male and Female."

through society. The simplest country-man could discover a plant, name it, and learn its properties. That information would most likely be transmitted orally (and sometimes lost) unless it came to the attention of a literate author. In writing, that information could find its place in botanical works (such as those of Theophrastus), in medical treatises (such as those of Dioscorides or Galen), in agronomical works (such as those of Columella), in learned poems (such as those of Nicander), or in historical/geographical treatises (such as those of Arrian). Technical or scientific knowledge – however one may wish to label it – of plants in antiquity could find many forms of expression. In any event, there was no homogeneous body of botanical literature in the Greek and Roman world.

14

SCIENCE AFTER ARISTOTLE: HELLENISTIC AND ROMAN SCIENCE

Liba Taub

The beginning of the Hellenistic period is conventionally dated to the death of Alexander the Great in 323 BCE. For our purposes, the death of his tutor, Aristotle, in 322, is at least as significant an event. This chapter considers science in the Greco-Roman world after Aristotle's death, focusing on a number of characteristic features of Hellenistic and Roman science: (1) the importance of allegiance to a philosophical school; (2) the public face of scientific work; and (3) interest in measuring and calculating, as well as a fascination with really big numbers.[1]

Philosophical "schools" provided the setting and identity for a great deal of work in natural philosophy, the study and explanation of nature (*physis*). These schools of thought or sects (*hairesis* = sect) were typically associated with a particular founding figure and approach; the word *hairesis* also has the meaning "choice," suggesting that members had chosen affiliation with a particular school.[2] Aristotle had been a student of Plato's school in Athens, the Academy. In turn, he founded his own philosophical school, the Lyceum. In the period following his death, allegiance to a particular school remained important, and this is the first characteristic considered here.[3] Even some individuals who might not have been regarded, in the first instance, as philosophers made a point of advertising their links to a specific school (as did the Roman orator Cicero by self-identifying (*Orator* 12) as a product of the Academy).

1 The conventional dating of the Hellenistic period is political, with the end of the Hellenistic period dated to 31 BCE (the Roman victory at Actium), and followed by the Roman imperial period.

2 D. T. Runia, "Hairesis," in Hubert Cancik and Helmuth Schneider (eds.), *Brill's New Pauly*, Antiquity volumes (Leiden: Brill, 2006). See also T. Engberg-Pedersen, "Introduction: A Historiographical Essay," in T. Engberg-Pedersen (ed.), *From Stoicism to Platonism: The Development of Philosophy, 100 BCE–100 CE* (Cambridge: Cambridge University Press, 2017), pp. 1–26, esp. p. 6, note 2 on issues involved in using the term "school."

3 "School" affiliation was similarly important for medical practitioners; see the chapters on medicine in this volume (2, 6, 16, 17, 27, 29).

The Hellenistic period saw an increase in literacy and the collecting of texts. Aristotle is credited with having assembled a significant number of books (papyrus rolls); one of our key sources, Strabo (ca. 64 BCE–after 21 CE), highlights his role: "Aristotle bequeathed his own library to Theophrastus, to whom he also left his school; and he is the first man, so far as I know, to have collected books."[4] The collection of texts provided members of the Lyceum with all sorts of information, including primary sources, historical materials, and an archive of earlier answers to philosophical questions.[5] Ptolemy I (ca. 367–ca. 283 BCE), a former general under Alexander the Great, became ruler of Egypt; it was he, or his son, Ptolemy II Philadelphus (309–246 BCE), who established the Library at Alexandria. Other rulers, notably the kings of Pergamon, competed in building impressive collections of texts. These royal collections were symbols of power and influence that supported scholarly activity, as was the Museum (or Mouseion) at Alexandria, paying tribute to the Muses. Institutions such as the Library and the Museum were associated with various intellectual endeavors, including natural philosophical and mathematical work, and the collection, sharing, and study of the writings of earlier thinkers were an important part of those intellectual activities, whilst the size of the collections reinforced images of power and influence.[6] There is evidence of other, smaller, libraries being assembled by various sorts of specialist practitioners, including physicians.[7]

Scientific work was presented in the public sphere in a number of different ways; this is the second characteristic considered here. Some individuals achieved a sort of celebrity status as mathematicians or philosophers particularly interested in studying nature: Archimedes and Posidonius are two such figures who achieved great fame in antiquity. Some sought to preserve and celebrate their own reputations: Archimedes gave instructions that his tomb should indicate his achievements. In other cases, the work and ideas of heroic thinkers were publicized by followers and admirers, including through the erection of monuments. And there were other public displays of

[4] Strabo *Geography* 13.1.54 (C608), on books bequeathed to Theophrastus, translated by H. L. Jones (Loeb Classical Library; Cambridge, MA: Harvard University Press, 1929), vol. 6, 111.

[5] C. Jacob, "Fragments of a History of Ancient Libraries," in J. König, K. Oikonomopoulou, and G. Woolf (eds.), *Ancient Libraries* (Cambridge: Cambridge University Press, 2013), pp. 57–81, on pp. 69 and 72. See also T. Hendrickson, "The Invention of the Greek Library," *Transactions of the American Philological Association* 144.2 (2014), 371–413, emphasizing book collections and the link to increasing literacy, rather than the spaces which housed such collections.

[6] Strabo 2.1.5 cites Hipparchus as commenting on the size of the library to which Eratosthenes had access. See also S. Johnstone, "A New History of Libraries and Books in the Hellenistic Period," *Classical Antiquity* 33.2 (2014), 347–93, on the book collections sponsored by the rich and powerful, as political institutions in which books were regarded as valuable objects, irrespective of their content.

[7] See, for example, V. M. Martínez and M. F. Senseney, "The Professional and his Books: Special Libraries in the Ancient World," in König et al. (eds.), *Ancient Libraries*, pp. 401–17. On ancient libraries more generally, see G. Woolf, "Introduction: Approaching the Ancient Library," in the same volume, pp. 1–20.

Figure 14.1. Roman mosaic described as depicting the Academy of Plato (Inv. No. 124545). From the Villa of Titus Siminius Stephanus, Pompeii; 1st century BCE–1st century CE. Reproduced by permission of the Ministero dei Beni e delle Attività Culturali e del Turismo – Museo Archeologico Nazionale di Napoli.

objects that could be understood as "scientific", such as sundials and calendars, as well as artwork – including mosaics depicting individuals associated with natural philosophy and mathematics.

Furthermore, literacy became more widespread, and some scientific and mathematical work was communicated through writings aimed at non-specialist audiences, such as the many "didactic" poems – both in Greek and Latin – dealing with the explanation of natural phenomena. The attraction of discussing nature and mathematics is evident in the work of authors such as Plutarch (for example, his dialogue *On the Face of the Moon*), and demonstrates the public face and wide appeal of scientific work.

Work on measuring and calculating the size of features of the natural world (and of the Earth itself), as well as large numbers, is the subject of

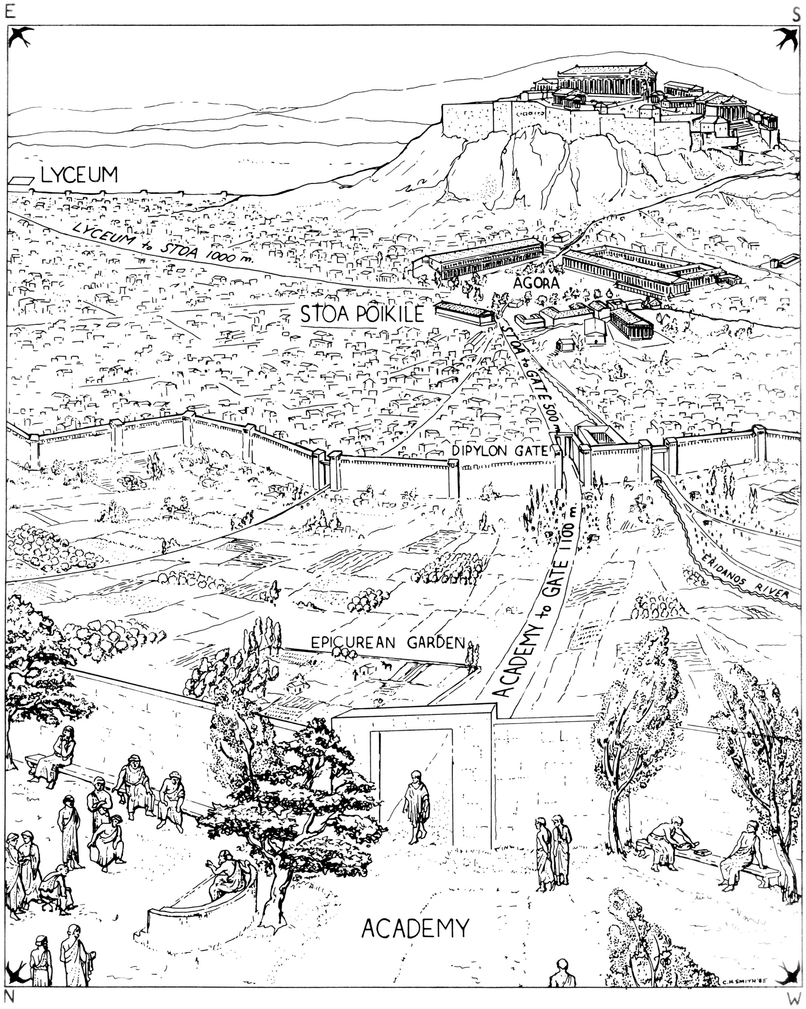

Figure 14.2. The philosophical schools of Hellenistic Athens. Drawing copyright C. H. Smith, reproduced with permission under Creative Commons Attribution-NoDerivatives 4.0 International License.

the final section of this chapter. One author writing in Greek under the Roman Empire, who was apparently widely read in antiquity, Lucian of Samosata (born ca. 120 CE), imagined a journey to the Moon in *A True Story*. As Lucian's work indicates, measuring and calculating the height or depth or distance of objects in nature (such as the Moon, mountains, or the sea) attracted a good deal of interest, as did counting more generally; indeed, the contemplation of really big numbers held a particular fascination.

PHILOSOPHICAL ALLEGIANCES

The Academy was founded by Plato, and Platonism in various forms had continuing influence. Aristotle's Lyceum passed after his death to Theophrastus (ca. 372–287 BCE), who became the new head. This school was also known as the Peripatos, because of the walkway around the Lyceum; members of the school were known as Peripatetics.[8] New philosophical schools were founded and flourished in Athens: the Garden of Epicurus (and the Epicureans) and the Stoa (or Porch), whose members were known as Stoics. Natural philosophy was an important area of inquiry for these schools.

A. A. Long and D. N. Sedley, in describing the Hellenistic philosophical schools, stated that each was "not, in general, a formally established institution, but a group of like-minded philosophers with an agreed leader and a regular meeting place, sometimes on private premises but normally in public."[9] The public setting of their philosophical activity was significant, having important historical roots (think here of Socrates in the Agora). Furthermore, the public activities of the philosophical schools enabled individuals who were not members of elite Athenian society – including in some cases women – to participate. Allegiance to a particular philosophical group and approach gave adherents an identity as members of a "school" or "sect," indicated by the names "Peripatetic," "Epicurean," "Stoic," or "Academic."

Loyalty to a school normally entailed loyalty to the founder: Plato for the Academy, Zeno for the Stoa, and Epicurus for the Garden. It was characteristic of the philosophical schools to offer interpretations of the founder's views, rather than criticisms. There were varying degrees of intellectual independence; for example, Peripatetics did challenge the ideas of Aristotle, whilst still often working within a Peripatetic tradition. In some cases members of the Stoa had philosophical differences, whereas followers of Epicurus – including Lucretius and Diogenes of Oinoanda – tended to present themselves as being focused on the views of Epicurus himself. In addition to Greek-language members, philosophical schools had adherents who were Roman, some of whom wrote about philosophical allegiances and ideas in Latin; indeed, these Latin authors did much to promulgate ideas about nature and the world.

Individual philosophers – indeed, the different philosophical schools – did not agree about the purpose of studying nature, or how such study should be undertaken. Questions about what constituted knowledge (*epistēmē*), or how it could be attained, did not always yield the same

8 See also chapters 10 and 11 on Aristotle in this volume.

9 A. A. Long and D. N. Sedley, *The Hellenistic Philosophers*, 2 vols. (Cambridge: Cambridge University Press, 1989), vol. 1, 5. This two-volume work is an invaluable collection of relevant texts and commentary.

answers. Nevertheless, almost everyone agreed that physics was important. (In the *Metaphysics* (1026a6 ff.), Aristotle had described theoretical knowledge as being comprised of three distinct subject areas: mathematics, physics, and theology.)

THE LYCEUM AND THE PERIPATETICS

There are particular characteristics of Aristotle's approach to study – for example, exhorting readers to take notes and the systematic collecting of documents and opinions, which reinforce the impression of the Lyceum as being focused on inquiry. The collection of the ideas of other thinkers was a feature of the coordinated work that characterized the Lyceum during Aristotle's time. His own scientific writings (for example, *History of Animals* and *Meteorology*) show that he was adept at collecting data, a task in which he made use of the contributions of others. He was also credited with having gathered the constitutions of many different city-states; this particular collection directly informed his work on politics. Having assembled data – through the constitutions themselves – about the political arrangements adopted in various places, Aristotle and his colleagues could use this information to inform their own ideas about the best constitution (cf. *Nicomachean Ethics* 10.9 (1181b16–22)).[10] This method of collecting ideas and then comparing them, as part of an intellectual project, was also associated with the Peripatetic school under Theophrastus, whose work as a philosopher followed that of Aristotle in many respects.

Similarly to Aristotle, Theophrastus' interests were wide-ranging, and he wrote on numerous topics related to the natural world. Diogenes Laertius (probably first half of the third century CE) reports that he wrote over 200 works on various subjects; of these a relatively small number survive, including one on the principles of nature, another *On the Senses*, and a number of shorter works on specific subjects, including *On Odours* (treating perfumes, including their medicinal properties), *On Fire*, *On Winds*, *On Stones*, *On Fish*, *On Sweat*, *On Fatigue*, and *On Dizziness*. He is also credited with *On Weather Signs*.[11] In particular, Theophrastus carved out an important area for himself through his study of plants in two works on botany, *Causes of Plant Phenomena* and *Enquiry into Plants* (apparently used to provide information for the writing of the former).[12] He furthered the

[10] Diogenes Laertius 5.27 states that Aristotle wrote a work on 158 constitutions of Greek cities. Scholars have taken this to mean that he was responsible, with others, for the collection and recording of these constitutions; see, for example, R. Brock and S. Hodkinson, "Introduction: Alternatives to the Democratic Polis," in their *Alternatives to Athens: Varieties of Political Organisation and Community in Ancient Greece* (Oxford: Oxford University Press, 2002), p. 1.

[11] See Theophrastus of Eresus, *On Weather Signs*, ed. D. Sider and C. W. Brunschön (Leiden: Brill, 2007), pp. 40–3.

[12] See chapter 13 on botany in this volume.

Peripatetic practice of collecting information and ideas from others, to inform intellectual inquiry.

As part of this practice, Theophrastus, as well as several unknown authors, produced texts that are essentially lists of the opinions (*doxai*) of natural philosophers and physicians, typically organized by topic, but sometimes by a particular individual or philosophical school. While collecting and recording the opinions of earlier thinkers may have begun in the fifth century BCE, during the fourth century the practice became integrated into the pursuit of knowledge about nature (as well as philosophy more broadly). One motivation for collecting opinions seems to have been to produce intellectual histories, in some cases to enable authors and their readers to identify with other thinkers as well as particular philosophical schools. In some cases, opinions are listed so that the author can argue against them, in favor of his own view. Aristotle himself often took this approach; see, for example, his *Meteorology*.

Scholars have identified a group of ancient texts that preserved opinions for later "mining" and use by ancient authors; these texts have been referred to as the *placita* literature (*placita* is a Latin term for "doctrines"). Such texts collected and organized opinions and tenets on a variety of topics, including natural philosophy, mathematics, and medicine; normally, they do not contain other sorts of information, such as biographical details about the people whose opinions are preserved. (Biographical details were sometimes included in other texts, such as the *Lives of Eminent Philosophers* presented by Diogenes Laertius.)

Aristotle emphasized the use of opinions (*endoxa*) as a starting point for inquiry, as part of his dialectical method, one of the types of deductive reasoning he advocated. At the very beginning of the *Topics* (100a18–100b24), he defined dialectic as the method by which we reason from generally accepted opinions; for this reason, familiarity with accepted opinions is crucial.

Theophrastus was credited by Diogenes Laertius (5.48) with having produced a work entitled *Physikōn Doxōn*, in sixteen books, the title interpreted to mean either *Tenets of the Natural Philosophers* or *Tenets in Natural Philosophy*. The work does not survive, and its exact nature has been a topic of extended debate. The nineteenth-century scholar Hermann Diels preferred the first meaning, emphasizing the role of individual thinkers, but some scholars opt for the second, focusing on the tenets, ideas, or opinions themselves.[13]

[13] See H. Diels, *Doxographi Graeci* (Berlin: Teubner, 1879), J. Mansfeld, "*Physikai doxai* and *Problemata physica* from Aristotle to Aëtius (and Beyond)," in W. W. Fortenbaugh and D. Gutas (eds.), *Theophrastus: His Psychological, Doxographical and Scientific Writings* (New Brunswick, NJ: Transaction Publishers, 1992), pp. 63–111, esp. pp. 63–7, and L. Zhmud, "Revising Doxography: Hermann Diels and his Critics," *Philologus* 145 (2001), 219–43.

Theophrastus' now-lost collection of *doxai* appears to have been collated in a systematic way, perhaps following a model similar to that used by Aristotle when he organized the collection of the constitutions of city-states in the Lyceum.[14] Diels argued that the *placita* texts are excerpts from this lost work of Theophrastus. His collection of ideas and opinions would have provided a sizable source of material to be used in philosophical, scientific, and even historical work by himself, students, colleagues, and later readers.

Doxographies – that is, collections of opinions and ideas – served their readers as a sort of archive, which – like book collections – could function as a resource or tool to aid further work. Some ancient authors relied on secondary sources, including summaries of the works of earlier thinkers in handbooks and epitomes. The collections of *doxai* can be seen as a special type of text produced as part of an intellectual practice that incorporated and built upon earlier work. One prominent example of the genre is the second-century CE work falsely attributed to Plutarch known as *On the Placita of the Philosophers concerning Physical Doctrines* (*Placita philosophorum*); the attribution to Plutarch, particularly well-known for his biographical writings on prominent Greeks and Romans (including politicians and emperors), is an indication of the broad range of interests he was acknowledged to have pursued. Plutarch was also responsible for a number of works that can be considered examples of scientific literature, including the *Dialogue on the Face on the Moon*, a text that Johannes Kepler (1571–1630) translated into Latin and upon which he also commented.

Another type of text was presented as a series of questions and answers. The work known as the *Problemata* is an intriguing example of a Peripatetic compilation of such questions-and-answers, many relating to natural philosophical or medical issues; the work may have been compiled as a group project. It is one of a number of texts associated with Aristotle, but most likely not written or compiled by him, but rather by members of his school, under the leadership of Theophrastus and his successors. The work is composed of thirty-eight books, covering a wide range of subjects, from problems connected with medicine, to problems associated with justice and injustice. The thirty-eight books of the work contain nearly 900 questions, organized by topic. Book 1 deals with problems connected with medicine, such as: "Why are great excesses disease-producing?," "Why do the changes of season and the winds intensify or check, and bring to a crisis and produce

[14] L. Taub, "Archiving Scientific Ideas in Greco-Roman Antiquity," in L. Daston (ed.), *Science in the Archives: Pasts, Presents, Futures* (Chicago, IL: University of Chicago Press, 2017), pp. 113–35; J. Mansfeld, "Doxography of Ancient Philosophy," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2012 Edition), section 5 "Using Doxographies" (http://plato.stanford.edu/archives/sum2012/entries/doxography-ancient/).

diseases?"[15] Elsewhere in the work, we find problems concerned with mathematics (Book 15) and problems with shrubs and plants (Book 20); Book 25 deals with air and Book 26 with wind. The problems are posed in a particular fashion; many begin with the question: "Why?" The answers are often in the form of a question: "Is it because . . .?"

The question-and-answer texts may have served as a list of active and interesting problems of inquiry to be tackled in the future; these texts also recorded topics thought worthy of consideration and discussion, in some context, including the philosophical community. It is easy to imagine that the questions posed reflected opinions (*doxai* or *endoxa*) and as such were related to doxographies, which also served a type of archival function, allowing retrieval and re-use of earlier ideas. Some of the texts attributed to Peripatetics, including Theophrastus himself, indicate an approach to doing natural philosophy that emphasized the continuing attraction of certain topics for inquiry, and a tradition of collective activity utilizing a methodology reliant on shared data, accumulated over time.

THE GARDEN: EPICURUS AND EPICUREANISM

Epicurus (341–270 BCE) was the founder of the school known as the Garden. The ultimate goal of his philosophy was the attainment of freedom from anxiety (that is, *ataraxia* or "being undisturbed"). To aid this, Epicurus developed a materialist explanation of the world and various phenomena: the world is composed of atoms and empty space (known as "the void"). For Epicurus, physics offered opportunities to explain potentially frightening phenomena – such as thunder and lightning – as due to natural causes, rather than to traditional gods; his strict materialist natural philosophy denied any causal role for the gods. Epicurean physics was motivated by the desire to alleviate fear of gods and death: if people can be freed from fear, they then achieve *ataraxia*.

His school, the Garden, was open to all, including women. This openness was a special feature of the Garden, perhaps contributing to its appeal, as well as the proselytizing zeal of some of its members, not least Diogenes of Oinoanda, who erected an enormous stone "billboard" publicizing the teachings of Epicurus. (As we will see, followers of other schools, including the Stoa, also spread the word about their philosophies; Cicero, Seneca, and Plutarch all wrote about philosophical ideas for a wider audience.)

Epicurean philosophy included epistemology, physics, and ethics. Epicurus set out his method in a work that no longer survives, but was summarized by Diogenes Laertius (10.31). The title, *Kanôn*, may be understood as referring metaphorically to a ruler or standard of measurement.

[15] Aristotle, *Problems* 1.1 and 1.3, translated by R. Mayhew (Loeb Classical Library; Cambridge, MA: Harvard University Press, 2011), p. 7.

It focused on epistemology and three criteria of truth, through which knowledge claims are formed and judged: our sensations, preconceptions, and feelings. This valorization of sensory experience underpins Epicurean empiricism.

Epicurus strongly advocated the use of empirical observation. He emphasizes that any causes proposed and accounts offered must not contradict experience. Agreement with the phenomena is imperative; even though what we see may be explained by a number of causes, none of these explanations may contradict sensory perception. And, even though sense perception contributes to knowledge, Epicurus acknowledges that direct perception is not always possible; in these cases, analogy is useful.

Analogy is one of the principal ways in which we come to have thoughts, for it allows us to extend knowledge from what is perceived to that which cannot be perceived. For example, Epicurus suggests that thunder may be due to wind rolling around in the hollows of clouds, in a way that is similar to when air is imprisoned in jars and similar vessels (*Letter to Pythocles* 100); we cannot access clouds directly, but we are familiar with household objects. The use of analogy is especially important for the explanation of distant phenomena, including the meteorological and astronomical.

Whilst Epicurus' work *On Nature* has been preserved only in fragmentary form, several of his surviving shorter works provide information about his views on the explanation of natural phenomena. Within the collection of *Key Doctrines*, one of the maxims sums up his stance on the utility of physics: "Were we not upset by the worries that celestial phenomena and death might matter to us, and also by failure to appreciate the limits of pains and desires, we would have no need for natural science."[16] For Epicurus, physics fulfilled a therapeutic end.

Some of his key ideas about physics are known through material preserved by Diogenes Laertius, including letters. These letters were not intended to give full details of his natural philosophy, but to serve as aides-mémoire for his followers and, as such, were designed to be brief and concise. This underscores the importance, for Epicurus, of making his ideas understandable to a wide audience.

In the letter to someone named Herodotus, Epicurus outlines his views on the material constitution of the world. All matter is composed of discrete "uncuttable" units (atoms) which exist in empty space; this void allows movement to occur. In the *Letter to Herodotus* (58–9), the use of analogy is key to how we can understand the atom, as direct observation is impossible; the word *analogia* is specifically used there.

[16] *Key Doctrines* 11, trans. in Long and Sedley, *Hellenistic Philosophy*, vol. 1, 155 (B), slightly modified.

The *Letter to Pythocles* offers Epicurus' explanations about astronomical and meteorological phenomena, and many analogies are offered to understand these.

Epicurus was committed to the view that it is not necessary to provide only one "correct" explanation of a phenomenon: rather, being able to offer a variety or multiplicity of explanations suffices to assure us that, in principle, such phenomena occur naturally, without the intervention of gods. By advocating a number of possible causes for phenomena, Epicurus may have been building on the work of Theophrastus, who held the view that for some phenomena a number of different causes exist. But Epicurus is not especially concerned with correlating specific causes with particular phenomena. Rather, he aims to establish only the general principle that multiple causes may result in the same phenomenon; detailed expositions do not add anything, from his point of view, and may unnecessarily complicate matters, causing further worry (*Letter to Herodotus* 79).

The language used by Epicurus (and by his follower, Lucretius) sometimes draws "biological" analogies, alluding to generation and corruption, two important processes undergone by living beings. Many passages in Epicurus and Lucretius are replete with imagery of living things and processes that affect them: seeds, irrigation, creation, and extinction. These analogies offer one sort of explanation for how worlds (*kosmoi*) come to be and end. Lucretius, who advocated Epicurean philosophy in his poem, *On the Nature of Things*, used "biological" metaphors to describe the world (*mundus*) itself. Epicurus offered a brief analogy between a living thing (*zōion*) and the cosmos. Lucretius presented an elaborate metaphor, describing the "life cycle" of our world – as well as the other worlds, or *kosmoi*, which exist – which, in a similar manner to living organisms, is subject to growth and decline, eventually coming to an end; it is not immortal and everlasting, but will cease to exist (2.1105–74).[17] This mortality of Epicurean *kosmoi* is in sharp contrast to the immortal and unchanging nature of, for example, the Aristotelian cosmos. And, while Epicureanism was in competition with Stoicism within the philosophical marketplace, Lucretius' allusions to the cosmic life cycle resonated with certain Stoic ideas, including their view that the world is a living being.

[17] On "biological" imagery in Epicurus and Lucretius, see D. Furley, "Cosmology," in K. Algra, J. Barnes, J. Mansfeld, and M. Schofield (eds.), *The Cambridge History of Hellenistic Philosophy* (Cambridge: Cambridge University Press, 1999), pp. 412–51, esp. pp. 424–6; P. H. Schrijvers, "Seeing the Invisible: A Study of Lucretius' Use of Analogy in the *De rerum natura*," in M. R. Gale (ed.), *Oxford Readings in Classical Studies. Lucretius* (Oxford: Oxford University Press, 2007), pp. 255–88; L. Taub, "Cosmology and Meteorology," in J. Warren (ed.), *The Cambridge Companion to Epicureanism* (Cambridge: Cambridge University Press, 2009), pp. 105–24, esp. 116–18; and F. Solmsen, "Epicurus on the Growth and Decline of the Cosmos," *American Journal of Philology* 74 (1953) 34–51.

THE STOA AND THE STOICS

Stoicism emphasized the importance of physics as a key part of philosophy, alongside logic and ethics. While both Epicurean and Stoic physics addressed the material composition of the world, unlike the situation for Epicurean theory, we have no Stoic text that addresses physics from first principles.

Whereas matter was composed of discrete units for the Epicureans, Stoics regarded matter as continuous, with no empty space within the universe. Their worldview emphasized the permeation of the cosmos by the divine.[18] Different Stoic accounts (and traditions) emphasized the roles of *pneuma* (breath) or *aither* (fiery material in the heavens) in the composition of the world. For the Stoics, the cosmos is a living being that is rational and has intellect (Diogenes Laertius 7.142, referring to the views of Chrysippus, ca. 279–ca. 206 BCE, the third head of the Stoa).

Whilst the Epicurean and Stoic schools shared some features – including philosophical systems aimed at achieving *ataraxia* – in many ways their ideas and approaches were in direct opposition. For example, the Epicurean gods are not concerned with the workings of the world (Lucretius 3.18–24), while the Stoic god is immanent throughout the universe, providing direction for all things. Furthermore, with regards to the value of physics, as already mentioned, Epicurus had warned against getting too caught up in the details of explanations, lest they lead to unnecessary worry. For Epicurus, knowledge of physics is useful (and sufficient) in that it dispels our fears, for example, that capricious gods cause violent weather. Stoics held that since happiness is obtained by following nature, the study of nature enables us to pursue happiness. As rational beings, the study of nature informs our understanding not only of the world, but of our own nature. Furthermore, the study of nature enables a practical, moral benefit that can be applied to our own lives.[19]

The impact and influence of the two schools was also somewhat distinct: Epicureanism was relatively marginalized; culturally, Stoicism achieved greater respectability. Lucretius' poem demonstrates the reach of Epicureanism in the Roman world, and Stoicism easily adapted to Roman culture, as seen in the writings of many different authors, operating at different levels of society. Indeed, the history of the Stoa is by convention divided into three periods: the "early," "middle," and "Roman," signaling the importance of Stoic philosophy within imperial Rome.

Our knowledge of the physics of the early members of the Stoa, including the founder of the school, Zeno (of Citium in Cyprus, 335–263 BCE), and his successors Cleanthes (of Assos in Asia Minor, 331–232 BCE) and Chrysippus

[18] See, for example, Cicero *On the Nature of the Gods* 2.22–3, 28–30; Diogenes Laertius 7.139.

[19] B. Inwood, "Why Physics?," in R. Salles (ed.), *God and Cosmos in Stoicism* (Oxford: Oxford University Press, 2009), pp. 201–23.

(of Soli in Cyprus, ca. 280–207 BCE), is based on reports preserved by later writers. Even though the early Stoics were prolific authors, for the most part only fragments and testimony of their writings survive. For information about founders and members of the Old Stoa, we are often reliant on doxographical writings, including pseudo-Plutarch's *Placita of the Philosophers*, Diogenes Laertius' *Lives of Eminent Philosophers*, and Stobaeus' *Excerpts* (fifth century CE). (Those Stoic works that survive in their entirety are from the Imperial period, and are largely concerned with ethics.) The Roman author Seneca (ca. 4 BCE–65 CE) identified himself as a Stoic; his *Natural Questions* is concerned with ethical questions, largely positioned as epilogues to his detailed discussion of the explanation of various meteorological phenomena.

Logic, physics, and ethics were not completely distinct areas within Stoic philosophy, as philosophy was regarded as an interconnected unity. The principal Stoic criterion of truth operates through an impression (*phantasia*) upon us; impressions may, for example, come through the senses. When we irresistibly assent to a *phantasia*, it is a cognitive impression, which is taken to be clear and reliable; for Stoics, the ability to describe impressions through language and discourse (*logos*) is essential. For this reason, Stoics valued grammar and rhetoric, as well as dialectic, in which the format of question-and-answer retained influence.

The Stoic conception of the universe relied on two principles interacting to make up the world. Matter is the passive principle acted upon by the active principle, variously referred to as "reason" (*logos*), "god," and even a sort of fire (or fiery *aither*). The cosmic fire has a special place in the Stoic cosmos, which experiences a never-ending cycle of coming-to-be and conflagration; paradoxically, the cosmos created in each of these cycles is identical to that which has preceded, and that which will follow. Following each conflagration, the four elements (fire, air, water, and earth) develop; the two active elements (fire and air) combine as *pneuma* or breath.

For the Stoics, matter is continuous, not atomic. Because there is no void, substances are able to interpenetrate each other: this enables *pneuma* to permeate the entire world. *Pneuma* imparts cohesion in various types of bodies through differing gradations; Chrysippus thought that *pneuma* is identical with semen (Diogenes Laertius 7.158–9).

Stoic ideas about the natural world had broad reach in antiquity, not least through the writings and reputation of Posidonius (of Apamea in Syria, ca. 135–ca. 51 BCE; aka "of Rhodes," where he died), who Strabo stated was the most learned philosopher of his time.[20] Today only reports and quotations of his writings survive, even though his ideas were widely cited by ancient authors, including Seneca as well as Strabo. He studied with the head of the Stoa, Panaetius (ca. 185–109 BCE), and wrote on many topics,

[20] Strabo *Geography* 16.2.10.

including logic, ethics, physics, and geography. Titles of about thirty books by him are named by ancient authors, covering a wide range of subjects, with titles such as *Physical Discourse*, *On the Cosmos*, *Meteorology*, *On the Size of the Sun*, and *On Gods*. In antiquity Posidonius had an impressive reputation.

The Roman Marcus Tullius Cicero (106–43 BCE), who studied oratory and philosophy in Athens and Rhodes, reported what he understood to be a Stoic conception of the value of physics:

> Physics too not unreasonably has earned the same distinction [being named a virtue], precisely because anyone who is going to live consistently with nature must start out from the cosmos as a whole and the providential care taken for it. Nor can anyone make true judgments about what is good and bad except on the basis of a knowledge of the nature and life of the gods too and of whether or not man's nature is congruent with that of the universe. No one can see the significance (and it is very great) of the ancient precepts of the sages . . . without physics . . . Nor can one understand piety to the gods nor how grateful we should be to them without a detailed account of nature.[21]

Physics played a therapeutic role within both Epicurean and Stoic philosophy. However, in contrast to Epicureans, for Stoics, understanding nature is a way to understand ourselves and our place in the cosmos.

PLATONISM AND NEO-PYTHAGOREANISM

The Academy continued to be important and influential. Aristotle had spent his formative years as a member of the Academy, and others famous for their scientific work, including Heraclides of Pontus (fourth century BCE), were associated with the Academy as well. Even in antiquity differing accounts of the history of the Academy were offered by various authors, some of whom identified themselves as members. Cicero, a student of the Academy, distinguished two Academies, the Old and the New, describing the latter as beginning with Arcesilaus (316/315–242/1 BCE), the sixth head of the school, whereas Diogenes Laertius reported three periods in the development of the Academy: the Old, the Middle, and the New. He regarded Arcesilaus as the head of the Middle Academy, associated with Academic skepticism, questioning the value of the senses in contributing to knowledge. Antiochus of Ascalon (born ca. 130 BCE) moved away from skepticism, adopting ideas of both the Peripatetic and Stoic schools and possibly contributing to the rise of what has been termed by some as Middle Platonism, sometimes regarded as a form of eclecticism.[22] Cicero – responsible for some of our information

[21] Cicerto *De finibus* 3.73, trans. Inwood, "Why Physics?," pp. 203–4.

[22] On the use of the term "eclectic," see Engberg-Pedersen, "Introduction," pp. 1–26; J. M. Dillon and A. A. Long, *The Question of "Eclecticism": Studies in Later Greek Philosophy* (Berkeley, CA:

Figure 14.3. Pythagoras, depicted on a coin produced at Samos, under Trajan Decius, Emperor 249–51 CE (Object ref.: 1844,0425.222) © The Trustees of the British Museum. All rights reserved.

about Hellenistic natural philosophy – was taught by Antiochus in Athens. Plutarch – another important source regarding Hellenistic natural philosophy and mathematics – identified as a Platonist. Theon of Smyrna (fl. ca. 115–40 CE), who wrote a work known as *Mathematics useful for Reading Plato*, and Alcinous (probably second century CE), the author of the *Didaskalikos* (*Handbook of Platonism*) with an interest in matter theory, were also both Platonists.

Platonism of the period was in some instances linked to neo-Pythagoreanism; both were seen to emphasize the role of mathematics in the cosmos. Plato's *Timaeus* taught that the world had been created incorporating mathematical ideas, and neo-Pythagoreanism regarded mathematics – including arithmetic, geometry, harmonics, and astronomy – as crucial to understanding its nature. Nicomachus of Gerasa (fl. ca. 100 CE) wrote an *Introduction to Arithmetic*, describing, amongst other things, the role of number in relation to philosophy and the world-order, as well as a *Handbook of Music*; both works survive, although his *Introduction to Harmonics*, *Theology of Arithmetic*, and *Life of Pythagoras* do not. What does survive of Nicomachus' writings makes clear that he regarded Pythagoras as a pre-eminent thinker. For some neo-Pythagoreans, Pythagoras acquired the status of a divine "hero." Apollonius of Tyana (first century CE) may have written a hagiographical account of his life

University of California Press, 1988). See also D. Sedley (ed.), *The Philosophy of Antiochus* (Cambridge: Cambridge University Press, 2012).

story; this approach was taken up by others in late antiquity (perhaps in competition with the Christian gospels).[23] Pythagoras' fame was celebrated in various ways, including on coins depicting his portrait.

THE PUBLIC FACE OF SCIENTIFIC WORK

Some individuals associated with scientific and mathematical work achieved a sort of celebrity status, capturing the interest of the wider public. And there were displays in public of various objects associated with scientific work and knowledge. Furthermore, scientific ideas and achievements were presented in a variety of written formats, including poetry.

Some founders of philosophical schools achieved extraordinary fame and even heroic status. Certain famous intellectuals felt obliged to take steps to protect their identity and reputation. The Alexandrian physician Galen complained that his name was falsely attributed to various books actually written by others, who took his name to authenticate their own work (*On my own works*).[24] Yet, there are others who were highly esteemed in antiquity but of whom we know very little today; especially notable here is Posidonius. As previously mentioned, although he was renowned in his own time, today his work only survives in fragments. Archimedes was so well known that Cicero describes his own pilgrimage to visit his grave; yet, as he recounts the story, the grave had been neglected and overgrown, unvisited. Plutarch had also described Archimedes' tomb, noting that it was Archimedes' wish that his grave be marked by a monument depicting a sphere inside a cylinder, a reference to his work *On the Sphere and Cylinder*, in which he demonstrated that the surface area and volume of the sphere are respectively two-thirds of the surface area and volume of the cylinder in which it is inscribed; he regarded this as one of his greatest achievements.[25]

[23] M. Hadas and M. Smith, *Heroes and Gods: Spiritual Biographies in Antiquity* (London: Routledge and Kegan Paul, 1965); C. Macris, "Porphyry's Life of Pythagoras," in C. Huffman (ed.), *A History of Pythagoreanism* (Cambridge: Cambridge University Press, 2014), pp. 381–98; L. Taub, *Science Writing in Greco-Roman Antiquity* (Cambridge: Cambridge University Press, 2017), Chapter 5. See also J. Dillon, "Pythagoreanism in the Academic Tradition: The Early Academy to Numenius," in Huffman (ed.), *History of Pythagoreanism*, pp. 250–73.

[24] Galen is discussed at length in chapter 17 in this volume. Also, see L. Totelin, "Mithradates' Antidote: A Pharmacological Ghost," *Early Science and Medicine* 9 (2004), 1–19 on "branding" scientific output in antiquity.

[25] Cicero reports his own search to find the grave of Archimedes in his *Tusculan Disputations* 5.64–6; see also M. Jaeger, "Cicero and Archimedes' Tomb," *Journal of Roman Studies* 92 (2002), 49–61; M. Jaeger, *Archimedes and the Roman Imagination* (Ann Arbor, MI: University of Michigan Press, 2008). Plutarch gives us a detailed description (*Marcellus* 17) of Archimedes' tomb. Historians disagree about what was actually depicted on Archimedes' tomb: see D. L. Simms, "The Trail for Archimedes' Tomb," *Journal of the Warburg and Courtauld Institutes* 53 (1990), 281–6; R. Netz and W. Noel, *The Archimedes Codex* (London: Weidenfeld and Nicolson, 2007), pp. 104–6.

Various objects associated with scientific work were displayed in public spaces, including stone monuments celebrating philosophical and mathematical achievements. Some individuals sought to erect their own testaments to their work. For example, in his *Letter to King Ptolemy* detailing his solution to the problem of doubling the cube, Eratosthenes (ca. 285–194 BCE) describes the votive monument he had erected to announce and celebrate his solution, presumably intended to be permanently and publicly visible. Claudius Ptolemy presented results from his astronomical work on a votive inscription in Canopus, near Alexandria, in the tenth year of Antoninus Pius' reign (146 or 147 CE). Although the original inscription does not survive, a number of medieval manuscripts containing his *Mathematical Syntaxis* (also known as the *Almagest*) preserve a (late) ancient transcription of the text, possibly based on one made by Olympiodorus (sixth century CE). Ptolemy's monument is an example of a specific form of astronomical inscription; another survives from Keskintos, Rhodes. Both are dedicated to gods. The first words of the Canobic inscription announce: "To the Saviour God Claudius Ptolemy [dedicates] first principles and hypotheses of mathematics." Ptolemy gives precise numbers for various elements of his astronomical models (including the radius of each epicycle). The text of the inscription ends by listing the harmonic pitches of the cosmic tuning associated with each astronomical body, and the ratios of musical concords.[26]

Whilst Eratosthenes and Ptolemy used publicly visible stone inscriptions to promote their own achievements, Diogenes of Oinoanda paid for an enormous (the largest currently known) inscription on the wall of a stoa in his hometown (in present-day southwestern Turkey), summarizing Epicurus' ideas. The entire inscription was apparently 2.37 meters high and extended about 80 meters, originally including around 25,000 words and occupying approximately 260 square meters of wall, presenting various works of Epicurus on subjects including epistemology, physics, and ethics. In the opening words, Diogenes of Oinoanda explains that, being at the end of his life, he wished to help others:

[26] A. Jones, "Canobic Inscription," in R. S. Bagnall, K. Brodersen, C. B. Champion, A. Erskine, and S. R. Huebner (eds.), *The Encyclopedia of Ancient History* (Malden, MA: Wiley-Blackwell, 2013), 1299; A. Jones, "Ptolemy's *Canobic Inscription* and Heliodorus' Observation Reports," *SCIAMVS: Sources and Commentaries in Exact Sciences* 6 (2005), 53–97; N. T. Hamilton, N. M. Swerdlow, and G. T. Toomer, "The 'Canobic Inscription': Ptolemy's Earliest Work," in J. L. Berggren and B. R. Goldstein (eds.), *From Ancient Omens to Statistical Mechanics: Essays on the Exact Sciences presented to Asger Aaboe* (Copenhagen: University Library, 1987), pp. 55–73; A. Jones, "The Keskintos Astronomical Inscription: Text and Interpretations, *SCIAMVS* 7 (2006), 3–41. On inscriptions as votive offerings, see B. H. McLean, *An Introduction to Greek Epigraphy of the Hellenistic and Roman Periods from Alexander the Great Down to the Reign of Constantine (323* BC–AD *337)* (Ann Arbor, MI: University of Michigan Press, 2002), p. 252. Serapis was presumably the god intended, the protector of Alexandria, whose Greco-Egyptian cult temple was located there.

> the majority of people suffer from a common disease, as in a plague, with their false notions about things, and their number is increasing (for in mutual emulation they catch the disease from one another, like sheep); moreover, [it is] right to help also generations to come (for they too belong to us, though they are still unborn); and, besides, love of humanity prompts us to aid also the foreigners who come here. Now, since the remedies of the inscription reach a larger number of people, I wished to use this stoa to advertise publicly the [medicines] that bring salvation. These medicines we have put [fully] to the test; for we have dispelled the fears [that grip] us without justification, and, as for pains, those that are groundless we have completely excised, while those that are natural we have reduced to an absolute minimum, making their magnitude minute.[27]

Diogenes aimed to enable others, both natives of and visitors to Oinoanda, to familiarize themselves with Epicurean ideas, as a remedy for fear and pain. His project was designed for therapeutic public benefit, to educate others in the teachings of Epicurus, including his scientific ideas.

The responsibility for the creation of certain other monumental objects associated with scientific work is unknown, even while some of these make reference to specific experts. For example, in the early twentieth century, fragments of inscriptions were found at Miletus, which were identified as stone examples of the astro-meteorological texts known as *parapēgmata*. The term *parapēgma* describes "something on which you fix something next to something else." The fragments (dated to the late second and early first century BCE) have peg-holes, representing days in the year; a peg was placed in the appropriate hole to mark the day. *Parapēgmata*, tools to track astronomical events (such as star phases) and seasonal weather, could be used in conjunction with local civic and astronomical calendars. The fragments that survive are relatively large, suggesting a public display. The provision and inscription of stone *parapēgmata* would have been laborious and costly; it is worth remembering the investment they represent. (And, although scholars stress their usefulness for agriculture and navigation, this presumes that farmers or bailiffs and sailors or captains would have had the literacy necessary to read them.) The *parapēgma* fragments attribute statements about astronomical phenomena to specific individuals (including Eudoxus and Kallaneos the Indian) as well as groups (the Egyptians).[28]

[27] Ed. and trans. M. F. Smith, *The Epicurean Inscription: Diogenes of Oinoanda* (Naples: Bibliopolis, 1993), p. 368. Square brackets indicate letters restored by Smith, where the text is damaged.

[28] D. Lehoux, *Astronomy, Weather, and Calendars in the Ancient World: Parapegmata and Related Texts in Classical and Near-Eastern Societies* (Cambridge: Cambridge University Press, 2007); L. Taub, *Ancient Meteorology* (London: Routledge, 2003); R. Hannah "From Orality to Literacy? The Case of the Parapegma," in Janet Watson (ed.), *Speaking Volumes: Orality and Literacy in the Greek and*

Ptolemy reports other astronomical objects on public display. In the *Mathematical Syntaxis*, he briefly alludes to two bronze astronomical instruments – equinoctial rings – that had been installed in a public area in Alexandria sometime earlier, possibly even in the time of Hipparchus (fl. second half of the second century BCE). Ptolemy implies that the rings were fairly useless, at least in the state in which he found them, as they did not give accurate readings and even sometimes indicated two equinoxes during the same day. The survival of these instruments in a public space for what may have been over 200 years is mysterious. It is possible that the equinoctial rings served several public functions – as sundials, if not also specialist instruments – reflecting a broad range of astronomical interests and practices alive in Alexandria, and contributing an air of prestige.[29]

Inscribed stone sundials were widespread in the Greco-Roman world: over 500 examples survive. Scholars have wondered whether most ancient sundials were merely decorative, adorning the gardens, for example, of Pompeii, as some do not seem to have been constructed for accurate time-finding. Recent work has used modern technology to analyze how these sundials were designed and made, and how technical knowledge was shared across the ancient Mediterranean world. In antiquity sundials were evidence of scientific and mathematical expertise: Vitruvius offered a brief history of the invention of distinct sundial designs in his work *On Architecture* (Book 9). Whilst many of these "garden" dials would have adorned private villas, there are also large examples of time-finding instruments on public display, notably the meridian line or sundial (the so-called Solarium Augusti, dedicated in 10 BCE) on the Campus Martius in Rome and, arguably, the Pantheon itself (dedicated during Hadrian's reign, probably 126 CE).[30] Such monumental displays involving the application of mathematical techniques to track natural phenomena (sunlight, indicating time of day and solar year) may have contributed to and been an expression of the appeal of science and mathematics to non-specialists.

Roman World (Leiden: Brill, 2001), pp. 139–59. In addition to the inscribed stone versions of *parapēgmata*, written forms were also produced; for example, Claudius Ptolemy was responsible for *Phases of the Fixed Stars and Collection of Weather Changes*, which also named particular astronomers. This recording of information attributed to others has parallels to the doxographical literature discussed above.

29 Ptolemy quotes Hipparchus referring to a bronze ring in Alexandria, at the "Square Stoa," although it is uncertain whether this was the same as one of those known to Ptolemy (*Mathematical Syntaxis* [*Almagest*] 3.1). L. Taub, "Instruments of Alexandrian Astronomy: The Uses of the Equinoctial Rings," in C. J. Tuplin and T. E. Rihll (eds.), *Science and Mathematics in Ancient Greek Culture* (Oxford: Oxford University Press, 2002), pp. 133–49.

30 R. Hannah, *Time in Antiquity* (London: Routledge, 2009), esp. pp. 131–4, 147–55; B. Fritsch, E. Rinner, and G. Graßhoff, "3D Models of Ancient Sundials: A Comparison," *International Journal of Heritage in the Digital Era* 2.3 (2013), 361–73; and the database of ancient sundials at http://repository.edition-topoi.org/collection/BSDP.

SCIENTIFIC AND MATHEMATICAL WRITINGS

The proliferation in antiquity of sundials in gardens and on estates is evidence of the appeal of such objects to members of elite classes; such objects were not only evidence of wealth, but would also have functioned as symbols of scientific and mathematical interests and understanding. Broader interest in scientific topics is also demonstrated in such works as Pliny the Elder's *Natural History*, written by a non-specialist, for non-specialists.

Intellectual work was undertaken across cultural and linguistic communities in the ancient Mediterranean world; some individuals, for example Cicero, had a working knowledge of both Greek and Latin. The ideas of many Greek philosophers were taken up, discussed, developed, and spread more widely by Roman authors, including Lucretius, Cicero, Seneca, and Pliny, to name a few. Some bilingual Roman authors translated Greek works (including Aratus' *Phainomena*) into Latin (as did Cicero, whose translation is largely now lost). Through a variety of "accessible" texts, scientific and mathematical ideas were also presented by Greek authors for non-specialist audiences, effectively serving to broaden the reach and appeal of science. Scientific and mathematical writings were not only produced by philosophers writing in prose; a number of authors, including Lucretius, Manilius, and possibly Archimedes wrote poetry on topics in physics, astronomy, and mathematics.

While in many instances, activities pursued by those writing in Latin were closely related to work undertaken by Greeks, this is not to say that some Romans, who had their own interests, agendas, and approaches, did not on some occasions deliberately contrast what they regarded as "Greek" ideas with Roman values.[31] Latin authors were often very self-conscious about their Greek predecessors and contemporaries. And, whilst during the period under consideration the different societies were undergoing substantial political and other changes, there were also important elements of shared culture across Greek and Roman linguistic and cultural communities. The texts we have are not simply a matter of linguistic translation and/or attempts to reach new audiences, but are also examples of cross-cultural interaction, involving the communication, transmission, and reappropriation of scientific and mathematical ideas and methods.

We find a number of Roman authors continuing an earlier practice of writing what may be regarded as histories of science, including Seneca (in his *Natural Questions*) and Vitruvius (whose work *On architecture* also contains accounts of the history of explaining nature, as well as the invention of sundials). In his *Natural History*, Pliny the Elder provides

[31] See, for example, A. J. S. Spawforth, *Greece and the Augustan Cultural Revolution: Greek Culture in the Roman World* (Cambridge: Cambridge University Press, 2012).

accounts of the history of striving to understand natural phenomena. Furthermore, then (as now) a premium was often placed on identifying who was the "first" to come up with a particular idea, or make a discovery.[32] These Roman authors were also situating their work and interests within a historical tradition that included Greek ideas about nature.

There is a good deal of evidence of non-specialist interest in engaging with ideas about the natural world. A number of politically active individuals devoted a fair bit of time to natural philosophy, data collection, and other "scientific" activities, sharing and debating ideas. Notably, in their dialogues, both Cicero and Plutarch show us how educated men were actively engaged with "scientific" ideas.

The Roman Stoic Lucius Annaeus Seneca was a prolific author, and many of his works are extant. He served as a tutor to Nero, and was politically engaged. While his writings cover a range of topics, his primary goal was to inspire ethical improvement. Toward the end of his life, in retirement from public life, he wrote the *Natural Questions*, which does not survive completely. Seneca's approach to explaining natural phenomena shares much with his predecessors whose ideas influenced his choice of topics. By and large, he addressed similar subjects to those covered by Aristotle in his *Meteorology*. As Aristotle, Theophrastus, Epicurus, and Lucretius had done before him, in his explanations of natural phenomena Seneca employed analogies to familiar experiences, some of which in his case provided details about contemporary Roman life (including the fashion for cooling drinks). He discussed and criticized others' explanations in the course of arguing for his own views; in the process, he provided information on many thinkers whose works are largely or entirely lost to us, including Posidonius.

Seneca makes it clear that one of his motivations for studying nature was the sheer pleasure involved:

> The investigation of this subject [nature] has many benefits, but none is finer than the fact that it captivates people with its own magnificence, and their motives for studying it are not gain but wonder. So let us examine the causes of these phenomena. I find this inquiry so enjoyable that, though I once wrote a book on earthquakes in my youth, I still wanted to test myself and explore whether age has added anything to my knowledge, or at least to my thoroughness.[33]

For Seneca, the study of the physical world is not separable from human concerns and activities, including morality. He encouraged others to study nature in order to leave behind sordid things, to keep the mind separate from

[32] See L. Zhmud, *The Origin of the History of Science in Classical Antiquity* (Berlin: de Gruyter, 2006).

[33] Seneca *NQ* 6.4.2, translated in Lucius Annaeus Seneca, *Natural Questions*, trans. H. M. Hine (Chicago, IL: University of Chicago Press, 2010), p. 92.

the body, and to exercise the mind on hidden matters, so as to deal with everyday, ordinary ones successfully (*Natural Questions* 3, preface 18). For Seneca, as a Stoic (and, similarly for Epicureans), the study and explanation of nature was a means to an end: a calm frame of mind.

Indeed, Epicurean natural philosophy was communicated in a number of different ways to a wide, non-specialist audience that was warned that too much specialist knowledge could lead to increased anxiety. Access to Epicurean philosophy was available through a variety of means: Epicurus' own letters and maxims were preserved in Diogenes Laertius' account of his life; and (as mentioned above) Diogenes of Oinoanda erected a huge inscription detailing his teachings. Lucretius' Latin poem *On the Nature of Things* was also apparently widely read in antiquity. Lucretius made it clear that he thought the "honeyed cup" of poetry made philosophical ideas more palatable (1:921–50 and 4:1–25, using almost the same words).

Poetry conveyed special authority within the Greco-Roman world, even in texts presenting scientific and mathematical ideas and methods. A number of "didactic" poems survive, including the so-called *Aetna* poem, and the *Astronomica* by Marcus Manilius (early first century CE), whose work shows Stoic influences. An important example of the wide reach of "scientific" poetry is provided by Aratus of Soli's (in Cilicia; ca. 315 to before 240 BCE) astronomical poem, *Phainomena*, written in Greek and subsequently translated repeatedly into Latin in antiquity. Aratus adapted the prose *Phainomena* of Eudoxus of Cnidus (ca. 390–ca. 340 BCE), which described the constellations and provided information about stellar risings and settings. Aratus also included information on weather-signs, as well as astronomy. In antiquity alone, Aratus' poem inspired at least twenty-seven commentaries, including one by Hipparchus, and several Latin translations, including those by Cicero as well as an author presumed to be Iulius Caesar Germanicus (15 or 16 BCE–19 CE). The various translators had differing interests and may have been appealing to different readerships. In some cases what was produced was not "merely" a translation, but an updating of scientific data or a re-working of the poem for other, non-scientific, purposes. The translations and adaptations of Aratus' poem are an indication of its interest and appeal, overshadowing Eudoxus' work upon which it was based; Eudoxus' text did not itself survive, whereas we have a number of versions (some fragmentary) of ancient Latin translations of Aratus's poem.[34]

[34] See E. Gee, *Aratus and the Astronomical Tradition* (New York: Oxford University Press, 2013); L. Taub, "Translating the Phainomena across Genre, Language and Culture," in Annette Imhausen and Tanja Pommerening (eds.). *Writings of Early Scholars in the Ancient Near East, Egypt and Greece: Zur Übersetzbarkeit von Wissenschaftssprachen des Altertums* (Berlin: de Gruyter, 2010), pp. 119–37.

The use of poetry to convey technical material was not confined to those who primarily identified as poets: a number of authors well known for their scientific and mathematical work are also credited with using poetry to promote their work, in particular Eratosthenes (whose *Letter to King Ptolemy* closes with an elegant epigram celebrating his achievement in providing a means to double the cube) and Archimedes (credited with the so-called *Cattle Problem*). In contrast to "didactic" poems, these epigrammatic poems convey technical expertise in a way that is simultaneously concise, erudite, and playful, exactly in keeping with the features expected of this particular genre of poetry. The *Cattle Problem* is a seemingly simple and practical problem about counting herds of different colored cattle, presented in epigrammatic form, as are the numerous mathematical problem-poems included in Book 14 of the *Greek Anthology* (*Anthologia Graeca*), a collection dating from the classical to the Byzantine periods.[35] Just as didactic poems aim to enlighten and often entertain, mathematical epigrams encourage intellectual engagement, offering challenging and clever riddles. Mathematical epigrams may have been composed and recited as part of the Hellenistic culture of the symposium, providing evidence of an early form of recreational mathematics.[36]

COUNTING, MEASURING AND BIG NUMBERS

Estimates of the height of mountains and depths of the seas, distances to the Sun and Moon, and size of the Earth and the cosmos itself were undertaken and discussed by various Hellenistic and Roman authors. Aristotle had referred to calculations made by others of the size of heavenly bodies, suggesting that they are much greater in size than they appear (*Meteorology* 339b8–9 and 32–6); he gave a figure for the Earth's circumference as 400,000 stades (*On the Heavens* 298a15–17). A number of Hellenistic and Roman authors produced their own estimates, including Hipparchus, who calculated the distance from the Earth to the Sun and Moon; his method was discussed by Ptolemy (*Almagest* 5.11). Aristarchus of Samos (fl. 280 BCE) devoted a work to the distances to and sizes of the Sun and Moon. Posidonius is credited with having given figures for the

[35] L. Taub, "'Eratosthenes Sends Greetings to King Ptolemy': Reading the Contents of a 'Mathematical' Letter," *Acta Historica Leopoldina* 54 (2008): *Mathematics Celestial and Terrestial – Festschrift für Menso Folkerts zum 65. Geburtstag*, ed. Joseph W. Dauben et al., 285–302. The *Cattle Problem* requires that eight unknown quantities be found, and is noteworthy for the difficulty of the mathematics involved. The problem was satisfactorily solved only in the twentieth century, and the solution involved the use of computers; it is not known whether the ancient author of the poem thought it was soluble and, if so, by what means. See M. Leventhal, "Counting on Epic: Mathematical Poetry and Homeric Epic in Archimedes' *Cattle Problem*," *Ramus* 44.1–2 (2015), 200–21.

[36] Taub, *Science Writing*, Chapter 1.

distances of the Sun and Moon (by Pliny 2.21.85–7, who notes that such estimates involve conjecture) and the size of the Sun (Cleomedes 2.1. 269–85), as well as a measurement of the circumference of the Earth (Cleomedes 1.7.1–48).[37] Eratosthenes estimated the size of the Earth (250,000 stades), applying geometrical methods to observations made at Alexandria and Syene. (The calculations of "sizes" often included the volumes of heavenly bodies; such calculations highlight the unmodernness of ancient interest in these quantifications.)

These individuals had specialist mathematical skills and knowledge. Yet, such topics held a broader fascination for the wider reading public, as evidenced by Pliny's account of one attempt at the estimate of the circumference of the Earth:

> Dionysodorus ... belonged to Melos, and was a celebrated geometrician; his old age came to its term in his native place; his female relations who were his heirs escorted his obsequies. It is said that while these women on the following days were carrying out the due rites they found in the tomb a letter signed with his name and addressed to those on earth, which stated that he had passed from his tomb to the bottom of the earth and that it was a distance of 42,000 stades. Geometricians were forthcoming who construed this to mean that the letter had been sent from the centre of the earth's globe, which was the longest space downward from the surface and was also the centre of the sphere. From this the calculation followed that led them to pronounce the circumference of the globe to be 252,000 stades.[38]

This calculation based on a fictional journey to the center of the Earth was strikingly similar to that obtained by other means. Furthermore, the account of a sort of imaginary scientific expedition resonates with other works from the period, including Lucian of Samosata's jocular account of journeying to the Moon. While modern critics, including Kingsley Amis, have debated whether such works can be read as science fiction, the appeal of writings

[37] N. Swerdlow, "Hipparchus on the Distance of the Sun," *Centaurus* 14 (1969), 287–305; T. Heath, *Aristarchus of Samos, the Ancient Copernicus: A History of Greek Astronomy to Aristarchus, together with Aristarchus's Treatise on the Sizes and Distances of the Sun and Moon* (Oxford: Clarendon Press, 1913); J. L. Berggren and N. Sidoli, "Aristarchus's On the Sizes and Distances of the Sun and the Moon: Greek and Arabic Texts," *Archive for History of Exact Sciences* 61.3 (2007), 213–54; K. Geus, *Eratosthenes von Kyrene: Studien zur hellenistischen Kultur- und Wissenschaftsgeschichte* (Munich: C. H. Beck, 2002). On Posidonius, see Cleomedes, *Lectures on Astronomy: A Translation of* The Heavens, trans. A. C. Bowen and R. B. Todd (Berkeley, CA: University of California Press, 2004), pp. 114–15; I. G. Kidd, *Posidonius,* vol. 2: *The Commentary* (Cambridge: Cambridge University Press, 1988), pp. 443–54 (on Cleomedes) and 464–6 (on Pliny). I thank John Hall for sharing his unpublished work on Posidonius.

[38] Pliny *Natural History* 2.112.248, translated by H. Rackham (Loeb Classical Library; Cambridge, MA: Harvard University Press, 1949), vol. 1, 373. Intriguingly, Pliny adds "to this measurement the principle of uniformity, which leads to the conclusion that the nature of things is self-consistent, adds 12,000 stades, making the earth the 1/96th part of the whole world."

centered on scientific subjects is demonstrated by the numerous texts that survive, as well as the titles and fragments of those no longer extant.[39]

Furthermore, it is through the work of some of these authors apparently writing for a broader readership that we have access to scientific ideas of the time. Pliny the Elder is an important source for our knowledge of the work of many whose writings do not survive. He reports Posidonius' views about the height at which winds and clouds occur (*NH* 2.21.85), and others' on the height of mountains (e.g. 2.65.162) and the depths of the sea (2.105.224), the sorts of things we can imagine all kinds of people – not only distinguished mathematicians – contemplating, if not actually calculating.[40]

Mountains were also of interest for various reasons, and there were numerous projects to measure, or estimate, their height. Dicaearchus of Messana (fl. ca. 320–300 BCE), reportedly a student of Aristotle, wrote on the measurement of mountains; his work may have been the earliest on the subject. Several authors report his measurements, including Pliny (*NH* 2.65.162), who explained that, with royal support, Dicaearchus measured the height of Mount Pelion, in Thessaly, at 1,250 paces.[41] Geminus (first century BCE), in his *Introduction to the Phenomena*, offered Dicaearchus' measurements to support his argument that clouds and wind formation occur below mountain summits.[42]

Several authors describe the use of instruments to help determine the height of mountains. Plutarch, in his *Life of Aemilius Paulus*, describes how Aemilius halted at the Pythium, to let his army rest prior to battle. Plutarch explains that "from this point Olympus rises to a height of more than ten stadia, as is signified in an inscription by the man [Xenagoras] who measured it," and provides what is apparently a quotation of the inscription:

> The sacred peak of Olympus, at Apollo's Pythium, has a height, in perpendicular measurement, of ten full stadia, and besides, a hundred feet lacking only four. It was the son of Eumelus who measured the

[39] Photius (ca. 810–ca. 893), in his *Bibliotheca* (summaries of 279 books composed for his brother) provided a synopsis of Antonius Diogenes' (possibly second century CE) *The Wonders Beyond Thule*, a fictionalized account of exotic travel, with references to Pythagoreanism; on possible connections to Lucian's work, see J. R. Morgan, "Lucian's *True Histories* and the *Wonders beyond Thule* of Antonius Diogenes," *The Classical Quarterly* 35.2 (1985), 475–90. See also K. ní Mheallaigh, *Reading Fiction with Lucian: Fakes, Freaks and Hyperreality* (Cambridge: Cambridge University Press, 2014), esp. Chapter 6, "*True Stories*: Travels in Hyperreality," and K. Amis, *New Maps of Hell: A Survey of Science Fiction* (Milton Keynes: Penguin, 2012; first published by Victor Golancz, 1960), pp. 11–12.

[40] Pliny reports Posidonius' estimates of height and depth; the figures appear to have been rounded, and the text itself may be corrupt. See Kidd, *Posidonius*, vol. 2, 464–6.

[41] On this measurement, see Rackham's note a, *Natural History*, vol. 1, 296.

[42] Geminus 17.5; see J. Evans and J. L. Berggren, *Geminos's* Introduction to the Phenomena: *A Translation and Study of a Hellenistic Survey of Astronomy* (Princeton, NJ: Princeton University Press, 2006), p. 218. Florian Cajori published an article in *Isis* 12 (1929), 482–514, "History of Determinations of the Heights of Mountains," in response to a query from Lynn Thorndike (published in *Isis* 9 (1927), 425, asking "When, where, and how did exact measurement mountains, by scientific method and on a cooperative basis, begin?".

> distance, Xenagoras; so fare thee well, O King, and be propitious in thy gifts.

Plutarch goes on to note that

> the geometricians say that no mountain has a height, and no sea a depth, of more than ten stadia. It would seem, however, that Xenagoras took his measurement, not carelessly, but according to rule and with instruments.[43]

Hero of Alexandria, in the work the *Dioptra* (13), described how a sighting instrument (a dioptra) can be used with geometry to determine a mountain's height from some distance away.[44] Plutarch's report provides evidence not only about the measurement of mountains but also – once again – about the sorts of scientific and technical achievements that were celebrated with monuments and inscriptions.

Mountains themselves served as a tool of measurement, and for delineating space, for geographical authors in antiquity.[45] All of the named mountains for which measured or estimated heights are given by ancient authors are in Greece,[46] but the manner in which their height is reported varies. For example, Strabo (8.8.1) mentions the height of Cyllene as part of a more general discussion of (not only topographical) features of the region:

> Arcadia lies in the middle of the Peloponnesus; and most of the country which it includes is mountainous. The greatest mountain in it is Cyllēnē; at any rate some say that its perpendicular height is twenty stadia, though others say about fifteen. The Arcadian tribes – the Azanes, the Parrhasians, and other such peoples – are reputed to be the most ancient tribes of the Greeks.[47]

Interest in the sizes, distances, and heights of natural objects also extended to estimates, and possible measurements, of the depths of rivers and the sea. Strabo cites Posidonius, noting that the Sardinian Sea is, according to him, the deepest of seas measured, at one thousand fathoms.[48] Other

[43] Plutarch *Life of Aemilius Paulus* 15, translated by B. Perrin, *Plutarch Lives* (Cambridge, MA: Harvard University Press, 1918), vol. 6, 395, slightly modified.

[44] See J. J. Coulton, "The Dioptra of Hero of Alexandria," in Tuplin and Rihll (eds.), *Science and Mathematics*, pp. 150–64. See also M. J. T. Lewis, *Surveying Instruments of Greece and Rome* (Cambridge: Cambridge University Press, 2001).

[45] See, for example, J. König, "Strabo's Mountains," in Jeremy McInerney and Ineke Sluiter, with Bob Corthals (eds.), *Valuing Landscape in Classical Antiquity: Natural Environment and Cultural Imagination* (Leiden: Brill, 2016), pp. 46–69.

[46] Namely, Cyllene, Pelion, Olympus, Atabyrius, and Acrocorinth; W. Capelle, *Berges- und Wolkenhöhen bei griechischen Physikern* (Berlin: Teubner, 1916), p. 34 n. 5.

[47] Strabo 8.8.1 (388C), translated by Jones, vol. 4, 227.

[48] Strabo 1.3.9. *The Online Liddell-Scott-Jones Greek-English Lexicon* defines *orguia* as "the length of the outstretched arms, about 6 feet or 1 fathom." One thousand *orguiai* (translated here as "fathoms") presumably equals ten stades, or about 1,850 meters; see Kidd, *Posidonius*, vol. 2, 795.

Hellenistic authors offer similar figures for sea depths: Plutarch (*Life of Aemilius Paulus* 15) mentions ten stades as a figure given for the deepest sea depths (he also mentions the measurement of the height of Mount Olympus), while Fabianus prefers fifteen stades (Pliny 2.105.224). Strabo implies that measurements of deep sea depths were undertaken, but it is not known how these figures (which appear to have been rounded) were achieved.

There was not always a firm division between what might be regarded as purely theoretical and more banausic concerns; Archimedes, for example, was known to be interested in both. Measurements (and estimates) of the heights of mountains and the depth of the sea were not only of interest in themselves, but were sometimes motivated by practical, including military, concerns, and also informed cosmology and physics.

While many educated Greeks and Romans would likely have understood the Earth to be spherical, mountains rising above the plains showed that the sphere was not actually perfect. Estimations and calculations of the height of mountains provided support for arguments that their altitude did not really alter the essential sphericity of the Earth. As part of his argument against the denial of the Earth's sphericity, Cleomedes (*ca.* 200 CE) stated that

> [t]hose who say that the Earth cannot be spherical because of the hollows occupied by the sea and the mountainous protrusions, express a quite irrational doctrine. For neither is there a mountain determined higher, nor a depth of sea [greater], than 15 stades.

Because the Earth has a diameter of over 80,000 stades (according to Eratosthenes, as Cleomedes had previously explained), even a protrusion of 30 stades would appear to be "just like a speck of dust would be on a ball."[49] The height of the highest mountains and the depth of the deepest sea are insignificant when compared to the size of the Earth itself, therefore such deviations on the surface do not provide evidence denying the Earth's sphericity.

Whilst measuring, calculating, and determining the sizes, distances, heights, and depths of features of the landscape and the cosmos itself were topics of interest, we find a special fascination with very large numbers. Mathematical authors, such as Hipparchus, Archimedes, and Eratosthenes, wrote about the number of fixed stars (approximately 1,000, according to Hipparchus) and calculated the number of cattle belonging to the Sun (in the *Cattle Problem*, attributed to Archimedes and addressed to Eratosthenes) and the number of grains of sand required to fill the cosmos (again,

[49] Cleomedes 1.7.121–6, translated by Bowen and Todd, pp. 84–5, who (note 25) list relevant passages in Strabo, Pliny, Plutarch, Seneca, and Theon of Smyrna.

Archimedes), whilst Aristarchus and Eratosthenes also contemplated the size of the cosmos and the number of prime numbers.[50]

Most of the works that deal with these large numbers are now lost. Only four survive, by two authors: Aristarchus' *On the Sizes and Distances of the Sun and Moon*, as well as Archimedes' *Measurement of the Circle*, the *Sand-Reckoner*, and the *Cattle Problem*. Yet, even work attributed to experts is sometimes cast in formats – such as poetry – to appeal to wider, non-specialist audiences; recall that the *Cattle Problem* was presented as an epigrammatic poem. The *Sand-Reckoner* (*Arenarius*) – concerned with how to express very large numbers – was addressed to the Syracusan king Gelon, not himself a mathematician, but possibly a patron.[51]

We find "big numbers" invoked by all sorts of authors, including Pliny the Elder, who detailed the vast number of sources he consulted in compiling the *Natural History*. The first book of the *Natural History* is a dense table of contents (*summarium*) listing the topics covered in the remaining thirty-six books of the work. In addition to noting the topics discussed and the sources consulted, Pliny also proudly reports the total number of facts, descriptions, and observations offered in each book. For example, Book 2 (on cosmology and astronomy, amongst other topics) has 417 facts, investigations, and observations, while Book 6, on various places, lists "1195 towns; 576 races, 115 famous rivers, 38 famous mountains, 108 islands, 95 extinct towns and races; 2214 facts and investigations and observations."[52] One gets the sense that – for Pliny and possibly some others, including his intended readers – more is more.

Pliny celebrated minutely detailed – and quantified – knowledge of the world. He also complained about the lack of serious scientific inquiry being done in his own time, despite it being a period of peace.[53] At one point, before launching into his own in-depth description of the winds, he provides (*NH* 2.45.117–18) a somewhat moralizing account of previous work on the subject, marveling at the efforts of over twenty (20!) Greek authors – working under adverse conditions – to study the winds. Pliny despairs that his fellow Roman contemporaries are not as knowledgeable as earlier Greeks:

[50] See R. Netz, *Ludic Proof: Greek Mathematics and the Alexandrian Aesthetic* (Cambridge: Cambridge University Press, 2009), pp. 55–62, on what he calls the fascination with size and the "carnival of calculation" undertaken by Hellenistic thinkers.

[51] Gelon (prior to 266 BCE–216 BCE) was co-regent with his father, Hiero II (ca. 308 BCE–215 BCE). Vitruvius (*On Architecture* 9.Preface 9–12) reports that it was Hiero who had asked Archimedes to find a way to ascertain whether the craftsman contracted to make a crown had removed some of the gold supplied for the project, substituting silver instead. Vitruvius reports Archimedes' experience in the bath (and shout of "Eureka"), his subsequent experiments, and confirmation of the attempted deception.

[52] Translated by Rackham, vol. 1, 37.

[53] Cf. Ptolemy's complaint in *Geography* Book 1.4 that there was so little data-collection of latitudes and longitudinal intervals determined from astronomical observations.

> This makes me all the more surprised that, although when the world was at variance, and split up into kingdoms, that is, sundered limb from limb, so many people devoted themselves to these abstruse researches, especially when wars surrounded them and hosts were untrustworthy, and also when rumours of pirates, the foes of all mankind, terrified intending travellers – so that now-a-days a person may learn some facts about his own region from the notebooks of people who have never been there more truly than from the knowledge of the natives.

He bemoans the lack of interest on the part of contemporary Romans in scientific work, complaining that

> now in these glad times of peace under an emperor who so delights in productions of literature and science, no addition whatever is being made to knowledge by means of original research, and in fact even the discoveries of our predecessors are not being thoroughly studied.

According to Pliny, his contemporaries are interested only in gaining wealth:

> now that every sea has been opened up and every coast offers a hospitable landing, an immense multitude goes on voyages – but their object is profit not knowledge; and in their blind engrossment with avarice they do not reflect that knowledge is a more reliable means even of making profit.

He contrasts his avaricious contemporaries with the scholars of the past, who pursued the study of natural phenomena mainly for intellectual reasons, approvingly noting that

> the rewards were not greater when the ample successes were spread out over many students, and in fact the majority of these made the discoveries in question with no other reward at all save the consciousness of benefiting posterity.[54]

The benefit of posterity is, in Pliny's view, the most admirable aim of scientific inquiry. Other authors, including Claudius Ptolemy, emphasized that they wished their scientific and mathematical work to be useful. A number of celebrated thinkers, including Archimedes, were especially concerned that they be remembered in the future, and took steps to ensure their reputations for posterity. One of the most important steps taken was the sharing of ideas and methods through written texts. Hellenistic and

[54] Pliny 2.45.117–18, translated by Rackham, vol. 1, 259–61. Recall that Thales allegedly profited financially through his astronomical knowledge.

Roman thinkers wrote on scientific and mathematical subjects in a wide range of types of texts, some of which were literally written on stone; not all of these writings survive today. By sharing their work, these authors made it more likely that it would be used, and that they would be remembered by posterity.

15

LATE ANTIQUITY: SCIENCE IN THE PHILOSOPHICAL SCHOOLS

Miira Tuominen

INTRODUCTION

Late antiquity has, during much of the twentieth century, been considered an era dominated by Neoplatonism with a mystical bend and not congenial to the science of nature. However, historians of science have also referred back to one late ancient philosopher, Philoponus (ca. 490–570), as a precursor of classical dynamics.[1] As has become increasingly clear over the past few decades, late ancient discussions of science and philosophy of nature include, even if not anticipate, many important developments, ideas, and arguments that were systematically furthered much later. The study of the late ancient philosophical schools is a rapidly growing field, and new discoveries are constantly made. The aim of the following discussion is to pinpoint some of the most conspicuous developments in the Platonic-Aristotelian schools. This limitation has been made to achieve some degree of internal unity, but it needs to be borne in mind that the Stoic and Epicurean schools also furthered sciences such as astronomy and meteorology in late antiquity.[2]

The scientific activity of the Platonic-Aristotelian schools was connected to the practice of commenting on the works of Plato and Aristotle but also of other authors on the theory of music (Porphyry (ca. 234–305 CE) on Ptolemy's *Harmonics*) and arithmetic (Ammonius (ca. 435/445–517/526) as written down by Philoponus on *Nicomachus's Arithmetic*), to mention some examples. In the context of the commentaries, questions and problems that we nowadays understand as scientific were discussed together with the problems that we classify as philosophical. The late ancient commentators

[1] See Michael Wolff, "Philoponus and the Rise of Preclassical Dynamics," in Richard Sorabji (ed.), *Philoponus and the Rejection of Aristotelian Science* (Bulletin of the Institute of Classical Studies Supplement 103; London: The Institute of Classical Studies, 2010), pp. 125–60.

[2] For Epicurean and Stoic meteorology, see Liba Taub, *Ancient Meteorology* (London: Routledge, 2003), pp. 127–61.

were mainly – with the important exception of Alexander of Aphrodisias (late second to early third century CE) – Platonists or Neoplatonists, and this affected the framing of the questions of natural philosophy as well as how nature and the methodology of natural philosophy were understood within a larger hierarchy of being and becoming (see below, "Unchanging Knowledge of Changeable Nature").

The philosophical schools also taught such disciplines as rhetoric, grammar, and mathematics (arithmetic and geometry), as well as astronomy/astrology and music theory, and had evolved slowly and gradually from an unsystematic private education practice.[3] There were chairs for philosophy in late antiquity, most importantly those established by Marcus Aurelius in 175/6 CE in Athens, and some evidence is found of philosophy teachers having been financed by municipalities. In the course of the fourth and fifth centuries, tension developed between the pagan schools and the Christian authorities, and the Athenian school came to closure due to Justinian's edicts given in 529 and 531.[4] In Alexandria, the Christian authorities financed the teaching of pagan philosophy, and during the turbulent time of the late 400s Ammonius, the head of the school, managed to reach a deal with the authorities to continue the activities.[5] After Ammonius, Eutocius, of whom we know little, concentrated on logic and mathematics, whereas the last pagan head of the school, Olympiodorus (late sixth century CE) wrote rather freely on issues more likely to provoke conflict with the officials.[6] The last schoolheads in Alexandria, Elias and David, were Christian and again concentrated on the safe logic of Aristotle – thus Christian faith did not entail greater freedom in studied topics.

Teaching in the late ancient Platonic-Aristotelian context followed a curriculum, established by Iamblichus (ca. 245–325), in which a certain order of learning was central. First the student read Aristotle's works in an order from logic to ethics, politics, physics, and theology. Then they followed Plato's dialogues in two cycles forming a succession from ethics to logic, physics, and theology. The second cycle of Plato's works concentrated on physics and theology and consisted of the *Timaeus* (physical) and

[3] See Ilsetraut Hadot, *Arts libéraux et philosophie dans la pensée antique. Contribution à l'histoire de l'éducation et de la culture dans l'Antiquité* (Paris: Vrin, 2006).

[4] See Edward Watts, *City and School in Late Antique Athens and Alexandria* (Berkeley, CA: University of California Press, 2006), pp. 130–42.

[5] For a suggestion of what the deal was, see Richard Sorabji, "Divine Names and Sordid Deals in Ammonius' Alexandria," in Andrew Smith (ed.), *The Philosopher and Society in Late Antiquity: Essays in Honour of Peter Brown* (Swansea: The Classical Press of Wales, 2005), pp. 203–14; for the teaching practices and the tension between the Christians and the pagans, see Edward Watts, *Riot in Alexandria* (Berkeley, CA: University of California Press, 2010), pp. 1–22.

[6] He wrote a commentary on Aristotle's *Meteorology* as well as on the Platonic *Alcibiades I* as well as on Plato's *Gorgias* and *Phaedo*.

the *Parmenides* (theological).[7] After Plato, the final stages included mystical texts, the *Orphic Poems* and the *Chaldean Oracles*, a collection of obscure hexameter verses from the late second century CE.

As mentioned, much of the work was done in the format of a commentary, and Plato and Aristotle were the most commented-upon authors, but other kinds of works were also written (for instance, Proclus' *Elements of Theology* and Philoponus' *On the Astrolabe*) and other authors were commented on (for example, Simplicius (sixth century) commented on the Stoic Epictetus' *Handbook*). In a sense, the commentary activity itself had, by the sixth century CE, become scientific.[8] The object texts were systematically followed on the basis of articulated principles of interpretation with discussions of controversial topics by other authors; Simplicius' commentaries provide a good example. The principles of interpretation varied from one commentator to another and were subject to vehement debate.[9] One of the main dividing lines between the commentators' methodology was whether or not they supposed that the disagreements between Plato and Aristotle should be taken as merely verbal – in which case these authors would be, in their deeper insights, in agreement. This so-called "harmony thesis" was especially important for Porphyry and Simplicius, whereas the Athenians (especially Proclus) argued that Plato must be preferred over Aristotle. Philoponus became more and more critical of Aristotle but did not accept Platonic tenets either (see below, "Philoponus' Unification of Dynamics").[10]

In the late ancient tradition, Aristotle's biological works were not commented on – except for *On Generation and Corruption* that concerns the more general principles of change. The Neoplatonists might have not written on biology because they were more concerned with metaphysics and theology, but even the Aristotelian Alexander of Aphrodisias did not write commentaries on Aristotle's biological works such as the *History of Animals*, *Parts of Animals*, or the *Generation of Animals*. One might speculate as to whether Aristotle's findings in biology were considered comprehensive and clear enough even without commentaries.

[7] For the two cycles, see *Anonymous Prolegomena to Platonic Philosophy*, chapter 26; translated in Richard Sorabji (ed.), *The Philosophy of the Commentators: A Sourcebook*, 3 vols. (London: Duckworth, 2004), vol. 3, 42.

[8] See Philippe Hoffmann, "Simplicius' Polemics. Some Aspects of Simplicius' Polemical Writings against John Philoponus: From Invective to Reaffirmation of the Transcendency of the Heavens," in Sorabji (ed.), *Philoponus and the Rejection*, pp. 97–123, p. 112; Ilsetraut Hadot, *Le problème du néoplatonisme alexandrin: Hiéroclès et Simplicius* (Paris: Études Augustiniennes, 1978), p. 194.

[9] Consider, for example, Simplicius' criticism of Philoponus discussed in Hoffmann, "Simplicius' Polemics."

[10] For references concerning Philoponus' philosophical development, see Sorabji (ed.), *Philoponus and the Rejection*, pp. 4–18.

MODIFICATIONS OF AN ARISTOTELIAN CONCEPTION OF SCIENCE

For the science of nature, the commentators to some extent took over the Aristotelian syllogistic framework, but also introduced important modifications. For Proclus, for instance, the dominant model was Euclid rather than Aristotle.[11] Platonism in general took a "Pythagorean" or a "mathematical" turn after Iamblichus,[12] and this development is important for Syrianus[13] (late fourth and early fifth century) and his student Proclus (411–485). The influence was carried over to Alexandria with Proclus' student Ammonius and to his followers, in particular Philoponus whose commentary on Aristotle's *Posterior Analytics* has much more mathematical examples than the treatise it comments on.

Aristotle's conception of science as demonstrative might seem suitable for articulating mathematical proofs in the style of Euclid. However, as Aristotle's examples show, the conception is better suited to expressing mathematical definitions in the form of genus and differentia. Triangle C, for example, is a plane figure A with the sum of its internal angles equal to two right angles B. This can be expressed as a syllogism. All figures that have the sum of their internal angles equal to two right angles are plane figures (A belongs to all B). All triangles have the sum of their internal angles equal to two right angles (B belongs to all C). Therefore, all triangles are plane figures (A belongs to all C). However, it is less clear how actual geometrical proofs could be expressed in a similar way. Philoponus' commentary suggests that the late ancient Platonic commentators made adjustments to allow more elaborate geometrical proofs in the synthesis and analysis form.[14]

The most important such adjustment was to take conversion in a geometric rather than in a syllogistic sense. In geometry, conversion is used to produce theorems; for example, the theorem "if a triangle is an isosceles, it has its base angles equal" converts into a new theorem "if a triangle has its base angles equal, it is an isosceles."[15] Proclus used similar conversion in his commentary on the *Timaeus* to produce theorems from hypotheses or axioms. He, for example, converted the definition of becoming into a demonstration of the universe as existent. The definition

[11] See Marije Martijn, *Proclus on Nature* (Leiden: Brill, 2010), especially chapter 3, pp. 67–192.

[12] For an earlier history of this turn, see Frans A. J. De Haas, "Philoponus and the Mathematisation of Logic," *Documenti e Studi sulla tradizione filosofica medievale* 20 (2009), pp. 193–210.

[13] See Angela Longo (ed.), *Syrianus et la métaphysique de l'antiquité tardive: Actes du colloque international, université de Genève, 29 septembre–1er octobre 2006* (Naples: Bibliopolis, 2009); Angela Longo, "Les 'Seconds Analytiques' dans le commentaire de Syrianus sur la 'Métaphysique' d'Aristote," in Frans A. J. de Haas, Mariska Leunissen, and Marije Martijn (eds.), *Interpreting Aristotle's Posterior Analytics in Late Antiquity and Beyond* (Leiden: Brill, 2010), pp. 123–33.

[14] De Haas, "Philoponus and the Mathematisation."

[15] The examples are from Euclid's *Elements*, propositions 1.5 and 1.6 as explained and quoted in Martijn, *Proclus on Nature*, pp. 140–1.

"becoming is what is grasped through perception and opinion" becomes by conversion "what is grasped through perception and opinion is becoming". With the additional premise that the universe is perceptible, this is developed into a demonstration that the universe is becoming (and has become) and is, in this sense, existent.[16]

Another, earlier modification is found in the Aristotelian commentator Alexander of Aphrodisias, who supposes that the syllogistic model of science from Aristotle's *Posterior Analytics*, together with the logic of categorical syllogisms of the *Prior Analytics*, applies much more broadly than Aristotle uses them in the preserved works. While Aristotle's treatises can be understood as taking the form of inquiries into the matters at hand, aiming at producing results that could be expressed as genus species relations or explanatory syllogisms, Alexander even occasionally attempts to force arguments in the treatises themselves into the straitjacket of categorical syllogisms (see for instance his cumbersome endeavor to formulate Aristotle's argument in *Topics* 1.8 as a categorical syllogism; Alexander *in Topics* 63.20–65.3).

Alexander also took the *Posterior Analytics* as a model for metaphysics, in Aristotle's terms "first philosophy" or "science of being *qua* being." Since Aristotle characterizes the science of being *qua* being as science,[17] it must, according to Alexander, also be demonstrative – a move that Aristotle himself does not make. Since, for Aristotle, all sciences must have a natural kind or genus as their proper object, and since being is not a genus (*Metaphysics* 3.3, 998b21–4), Alexander had to establish a proper object for science of being *qua* being. He did so by making a dubious move, interpreting substance as a genus in a more general sense (Alexander *in Metaph.* 245.1–5), on the basis of an ontological dependence of the other categories on substance, and claimed that the science of being *qua* being establishes the per se attributes of beings through demonstrative syllogisms on the basis of the common axioms in a manner analogous to the specific sciences. A crucial problem with this suggestion is that the category of substance does not include the other categories in the manner genera include the species.[18]

The commentators thus systematized the Platonic-Aristotelian framework for science in different ways and discussed topics we would classify as scientific. Let us next move to consider some of these discussions.

[16] Proclus, *in Tim.* 1.283.15–18 in Martijn, *Proclus on Nature*, pp. 140–2; De Haas "Philoponus and the Mathematisation."

[17] Aristotle, *Metaphysics* 4.1 (1003a21–1005a18); 6.1 (1025a3–1026a32).

[18] See Maddalena Bonelli, "Alexander of Aphrodisias on the Science of Ontology," in de Haas, Leunissen, and Martijn (eds.), *Interpreting Posterior Analytics*, pp. 101–21; Maddalena Bonelli, *Alessandro d'Aphrodisia e la metafisica come scienza dimostrativa* (Naples: Bibliopolis, 2001).

PHILOPONUS' UNIFICATION OF DYNAMICS

Aristotle's natural philosophy operates on the assumption that the sublunary and the superlunary regions of the universe move according to different dynamics. In the sublunary region, things move either according to their internal nature to their natural places (stones fall, water flows, fire rises upwards) if nothing prevents them or forces them to move otherwise, or in accordance with impulses originating from the soul's desires and deliberations and having external final causes. The heavens, by contrast, rotate in circles consisting of the fifth element, "ether," enabling perfect, eternal circular movement (*De caelo* 1.2). Aristotle had suggested up to fifty-five concentric spheres (*Metaphysics* 12.8), the fixed stars moving on the outermost one, the sun, the moon, and the five planets on the others with the earth at the center. The Aristotelian model also supposed that the celestial bodies are ensouled, intelligent beings that have the unmoved mover as their final cause of motion "as a beloved" (*De Caelo* 2.2 and *Metaphysics* 12.7).

Alexander modified Aristotle's explanation both with respect to the psychological explanation and the number of spheres, which he reduced to seven.[19] For Alexander, the outermost sphere on which the fixed stars move has a soul, and the final causality of the unmoved mover is explained by intellectual will (*boulêsis*) rather than desire (*erôs*). Even Aristotelians admitted that the model with concentric spheres fails to explain the approach and retreat of such planets as Venus and Mars,[20] and Ptolemy, for example, introduced eccentric spheres and epicycles to enhance the explanatory power of the model.[21] In the sixth century, the dominant conception in the philosophical schools was that the heavens have a dynamics of their own in contrast with the sublunary realm.

The Christian commentator and grammarian, John Philoponus, contested this division of dynamics. His new account was related to his criticism of Aristotle's remark concerning projectile motion (*Physics* 4.8, 215a14–17). Aristotle supposed that all forced motion in the sublunary realm has to be caused by a mover that is external to the moved object and in contact with it. These suppositions became problematic in the case of projectile motion, the stock example being a javelin that an athlete throws. Since no principle comparable to Newton's first law of motion belonged to Aristotle's discussion, the fact that the javelin keeps moving when the athlete has let go needed explanation. Aristotle's suggestion was based on the idea that air functions as the medium and keeps moving the body after it has lost contact with the mover. However, air also resists motion since it eventually dies out,

[19] For texts, see Sorabji (ed.), *Commentators: Sourcebook*, vol. 2, 337–40.

[20] Sosigenes, according to Simplicius *in Cael.* 504.17–506.3; in Sorabji (ed.), *Commentators: Sourcebook*, vol. 2, 377–9.

[21] See, Simplicius *in Cael.* 32.1–29 in Sorabji (ed.), *Commentators: Sourcebook*, vol. 2, 376–7. See also chapter 19 by Jones, in this volume.

and thus the medium (in this case air) ends up carrying two conflicting tendencies with respect to projectile motion.

Philoponus argued at length against Aristotle's brief remark and suggested that, were Aristotle's account correct, soldiers could launch their catapults by just moving the air behind them.[22] His own suggestion was that, instead of an external mover, the javelin receives an impressed moving force from the athlete's hand (*in Phys.* 642.4). The force then moves the javelin forward but since it is exhaustible, it dies out and the javelin falls. Such an impressed force came to be called *impetus* (*vis impressa* or *impetus impressus* by young Galileo), and Pierre Duhem coined the expression "*impetus* theory" for a theory that explains movement by such impressed forces.[23]

Philoponus also challenged Aristotle's arguments according to which movement is not possible in void and needs a medium (such as air, water, or ether) as its efficient cause. In this context Philoponus came close to the observation that Galileo later made concerning the speed of falling objects. Philoponus observed that "two unequal weights dropped from a given height strike the ground at *almost* the same time" (*Corollarium de inani, in Phys.* 683.16–25).[24] However, this did not lead him to draw Galileo's conclusion that bodies with different absolute weight fall with the *same* speed (in the same medium or in void). For Philoponus the natural heaviness of the bodies alone remained the efficient cause of fall (*in Phys.* 681.10–12).

The notion of an impressed force allowed Philoponus to unify the explanation of celestial motion with sublunary dynamics.[25] This means that he claimed the principles of the heavenly motion to be exactly the same as those governing forced motion in the sublunary realm. According to Philoponus, God has, when creating the universe, impressed a force into the celestial bodies to move (*De opificio mundi* 1.12). Since this force is exhaustible, celestial motion also comes to an end, a view congruent with the Christian view of the cosmos as created and perishable.

[22] See Sorabji (ed.), *Philoponus and the Rejection*, p. 48; Wolff, "Philoponus," pp. 129–32; Sorabji (ed.), *Commentators: Sourcebook*, vol. 2, 351–2.

[23] Pierre Duhem, "Liste de publications de P. Duhem et Notices sur les travaux scientifiques de Duhem," *Mémoires de la Société des Sciences physiques et naturelles de Bordeaux* 7 (1917), pp. 41–169; see pp. 162–3.

[24] See Michael Wolff, *Fallgesetz und Massebegriff. Zwei wissenschaftshistorische Untersuchungen zum Ursprung der klassischen Mechanik* (Berlin: de Gruyter, 1971), pp. 123–37. He points out (p. 136) that elsewhere Philoponus contradicts his observation and attributes to heavier bodies a greater capacity to divide air and thus a greater downward tendency than lighter bodies have (Philoponus, *in Phys.* 679.5–11).

[25] For the debate about the development of Philoponus' view, see Koenraad Verrycken, "The Development of Philoponus' Thought and Its Chronology," in Richard Sorabji (ed.), *Aristotle Transformed: The Ancient Commentators and their Influence* (London: Duckworth, 1990), pp. 233–74; Frans A. J. De Haas, *John Philoponus' New Definition of Prime Matter: Aspects of its Background in Neoplatonism and the Ancient Commentary Tradition* (Leiden: Brill, 1997); Pantelis Golitsis, *Les commentaires de Simplicius et de Jean Philopon à la Physique d'Aristote: tradition et innovation* (Berlin: de Gruyter, 2008). This paragraph concerns Philoponus' late views in *De opificio mundi*.

CREATION AND ETERNITY

Philoponus combined his dynamics of exhaustible impressed forces with a rejection of the Aristotelian ether as the constituent matter of the heavens and argued that the heavenly bodies consist of essentially the same elements as the sublunary world.[26] Therefore, heavenly bodies and the whole cosmos have been created and will perish. The claim that the cosmos is perishable and consists of the earthly elements was an anathema to the Neoplatonic pagan tradition,[27] and Philoponus also argued against this tradition in his interpretation of Plato's *Timaeus*.

In the late ancient (Neo-)Platonic tradition, the Demiurge of Plato's *Timaeus* was not assumed to entail an instance of time for the organization of the cosmos; several senses of "being generated" were distinguished to make that claim.[28] Cosmologically most important are the following two: (a) the universe is generated in the sense of being constantly in the state of becoming or change "as a mime, remaining one and the same person, takes on many different appearances";[29] (b) the cosmos is generated because "its being is derived from another source" (*Contra Proclum* 147.5–6), that is, from the Demiurge. Such distinctions allowed late ancient Platonists to maintain that even though the existence of the universe depends on the Demiurge as its efficient cause, demiurgic causation is constant and does not require temporal creation.

This interpretation was combined with the argument that if the universe had a temporal beginning, it would have to perish (Proclus, *in Tim.* 6.119. 14–120.14). Philoponus denied this argument and claimed that Plato's cosmology in the *Timaeus* is consistent even when combining temporal beginning with imperishability. Philoponus' argument was based on a distinction between natural and acquired properties: even though the cosmos is in its nature perishable, it can acquire imperishability from God (*Contra Proclum* 595.13–19).[30] Therefore, there is no contradiction in taking

[26] An Aristotelian, Xenarchus of Seleucia had already questioned ether (first century BCE; see Simplicius, *in Cael.* 21.33–22.17 in Sorabji (ed.), *Commentators: Sourcebook*, vol. 2, 358–9; Andrea Falcon, *Aristotelianism in the First Century* BCE*: Xenarchus of Seleucia* (Cambridge: Cambridge University Press, 2012)) as did the Platonist Taurus (second century CE; see Philoponus, *Contra Proclum* 520.18–20).

[27] See Simplicius *in Cael.* 119.7–13, 135.31–136.1; Hoffmann, "Simplicius' Polemics," p. 109. For Simplicius against Philoponus on the destructibility of the heavens, see *in Cael.* 35.31–3; 59.15–18; 73.5; 80.23–6; 81.10–11.

[28] The Platonist Taurus distinguished four such senses (according to Philoponus, *Contra Proclum* 145.13–147.25, translated in Sorabji (ed.), *Commentators: Sourcebook*, vol. 2, 162–4) and Porphyry added a further two (Philoponus, *Contra Proclum* 148.7–15 in Sorabji (ed.), *Commentators: Sourcebook*, vol. 2, 164).

[29] The Platonist Taurus in Philoponus, *Contra Proclum* 146.26–147.2; translated by Michael Share in Sorabji, *Commentators: Sourcebook*, vol. 2, 163.

[30] Proclus' argument and Philoponus' response to it is discussed in Lindsay Judson, "Generability and Perishability," in Sorabji (ed.), *Philoponus and the Rejection*, pp. 221–37, see especially pp. 228–30.

Plato to ascribe both a temporal beginning and imperishability to the orderly universe.

Philoponus also wrote extensively against Aristotle and Proclus on the claim that the universe is eternal in the sense of being ungenerated and imperishable (*De aeternitate contra Aristotelem* and *Contra Proclum*). The arguments usually aimed at showing that the adversaries' conceptions are internally inconsistent. Even though Philoponus' own position was Christian, and he argued against a pagan conception, the debate was not conducted in terms of Christian beliefs. One reason for this was probably that arguments referring to Christian beliefs or scriptures would not have worked against Philoponus' pagan adversaries.[31]

The crucial point in many of Philoponus' arguments was that the claim about the eternity of the universe conflicts with the claim that no actual infinity exists.[32] Philoponus argued that the Aristotelian notion of infinity as expandable finitude contradicts the assumption of an eternal universe without a beginning. According to Philoponus, the claim that we can always count further backwards from the present without reaching the beginning of time and the universe implies that the universe has already traversed an infinite number of years.[33] Philoponus did not accept the response according to which the infinity is not actual because the past years have already perished and thus do not exist. He also added that next year when the universe will have traversed one more year, the age of the universe is even greater than infinity. From a modern point of view this is not a valid objection because if the infinity of the number of years traversed by the universe is numerable, it will be no greater next year. However, a consistent analysis of the cardinalities of infinity was only developed by Georg Cantor in the late nineteeth century.

IS LIGHT A BODY? IS IT MOVEMENT?

The nature of light was discussed both on a cosmic scale and with respect to our everyday environment. Important discussions are found in the commentaries on the theory of perception, especially on Plato's *Timaeus* and Aristotle's *De anima*. From today's perspective, Plato's *Timaeus* and Aristotle's *De anima* concern both philosophical and scientific issues. Consequently, the commentaries on the works also concern both kinds of questions. In particular, the nature of light and the question of how

[31] See Dirk Baltzly, "Review of Helen Lang and A. D. Macro, *On the Eternity of the World (De aeternitate mundi) by Proclus* (Berkeley, CA and Los Angeles, CA: University of California Press, 2001)," *Bryn Mawr Classical Review* (2002), at http://bmcr.brynmawr.edu/2002/2002-10-19.html.

[32] See Richard Sorabji, "Infinity and Creation," in Sorabji (ed.), *Philoponus and the Rejection*, pp. 207–20.

[33] *Contra Proclum* 9.14–11.17; in Sorabji, *Commentators: Sourcebook*, vol. 2, 176–7.

perceptible qualities, especially colors, affect our sensory mechanism from a distance and activate our soul's capacity to perceive are examples of the latter. The commentators' views on what perception essentially is deviate on whether it is active or passive, a judgment of reason or a reception of qualities, for example. However, Aristotle's claim that perception involves the reception of perceptible forms without matter was widely accepted. The transmission of visible qualities seemed most difficult to explain as it might seem to involve efficient causation at a distance without any physical mediating mechanism – and such causation was not usually accepted.[34]

This debate reflected a controversy on the nature of light, focusing on the question of whether light is a body or not. Plato had suggested that light is a kind of fire (*Tim.* 45b2–d3, 58c5–6), whereas Aristotle denied this view and defined light as the active state of the medium, that is, such material that becomes transparent when illuminated by the presence of fire or something like fire (*De an.* 2.7, 418b9–10; 418b16–20; 419a9–11; *De sensu* 439a20). Another crucial question was whether light could be understood as movement, a view that Empedocles had proposed and the atomist Epicurus endorsed. The latter emphasized that the movement of light is so swift that it escapes our notice. He suggested that the speed of light is unsurpassed,[35] a view that is analogous to modern physics, taking the speed of light as the maximum speed for all energy, matter, and information in the universe. Aristotle had argued against the view that light is movement and claimed that it should rather be considered as an activity or actuality.[36]

Those who followed Plato in understanding light as a kind of fire and thus a kind of body needed to explain how light can pass through other bodies (such as glass and water) and be present in the air.[37] Philoponus formulated the challenge as follows:

> Moreover, [if light is fire] the air will not be ubiquitous, but where there was light there would not be air and where there was air there would not be light. As a result, we would turn out to breathe light and air and sometimes light only and not air, when our respiration was occurring in the illuminated area! All this is patently absurd and wildly contradictory to experience.[38]

[34] Cf., however, Plotinus who accepted causation without physical contact in visual transmission, even though he also subscribed to a version of the theory of reception of forms; see Eyjólfur K. Emilsson, *Plotinus on Sense Perception: A Philosophical Study* (Cambridge: Cambridge University Press, 1988), pp. 47–62, 133–40.

[35] See, for example, Epicurus, *Letter to Herodotus* 46–49.

[36] *De anima* 2.7, 418b20–6; *De sensu* 6, 446a26–b3 and 446b28. For the analysis of light as an activity or actuality, see Jean Christensen De Groot, "Philoponus on *de Anima* 2.5, *Physics* 3.3, and the Propagation of Light," *Phronesis* 28 (1983), 177–96.

[37] Aristotle, *De an.* 2.7, 418b17–20.

[38] Philoponus, *in De anima* 327.30–4; translated by Hugh Lawson-Tancred in Sorabji, *Commentators: Sourcebook*, vol. 2, 282.

Philoponus' formulation is not entirely charitable to those who suggested that light is a kind of fire, since they made careful distinctions between material or corporeal bodies and light. Syrianus, for example, clarified that if two bodies are material and resistant three-dimensional solids, they cannot occupy the same place (*in Metaph.* 85.18–22). However, such solid material bodies need to be distinguished from immaterial ones. One might suspect that this distinction makes light incorporeal, but, according to Syrianus, light nevertheless has the same extension as bodies do and stretches itself over three dimensions and occupies space.[39] Syrianus seems to have supposed that even though light is a body it is not solid and thus there is no harmful collision of bodies in the air that light occupies.

Proclus, Syrianus' student, stressed that we must distinguish between sublunary light and the divine light in the heavens (*in Tim.* 2.43.29–44.24). For Proclus, both these lights are bodies but the former is material and perishable whereas the latter is immaterial and eternal (*in Tim.* 2.9.8–18). However, the immaterial nature of the heavenly light did not imply that it would lack matter altogether but that its matter is more fine-grained than the matter of sublunary things. Proclus also distinguished sunlight and firelight in the sublunary region – the former being a body but to some extent immaterial, that is, supposedly more fine-grained than firelight. Thus the problem of two bodies occupying the same place became pertinent only to firelight. Proclus' solution seems to have been to argue that there are small pores in air and fire and they both go through those pores (according to "Simplicius" *in De anima* 134.6–8)[40] and thus only seem to coexist at the same place even though in reality they occupy different locations within their fine-grained structure.

Those who understood light as the actuality of the transparent medium as illuminated (for example, Philoponus *in De an.*, 330.19–20), to some extent modified Aristotle's conception. The main idea was that the medium (such as water or air) has the capacity to transmit colors only when it is illuminated. Aristotle had supposed that the transmission of colors in the illuminated medium differs from movement because the former is instantaneous and non-progressive (meaning that it involves a change of a whole at once), whereas local movement is progressive and takes time. Philoponus argued that even though the propagation of light through the medium takes no time, it is progressive. He explained this by saying that the sun as a luminous body possesses an illuminating power (*in De an.* 330.11) that acts on the neighboring body that, thus activated, also comes to possess the same power and acts on things adjacent to it.[41] Such progressive change is instantaneous

[39] For the text, see Sorabji (ed.), *Commentators: Sourcebook*, vol. 2, 276.

[40] For the texts, see Sorabji (ed.), *Commentators: Sourcebook*, vol. 2, 276–89; for "Simplicius" on Proclus, see p. 280.

[41] For Philoponus' conception, see De Groot, "Philoponus." For the view that light needs no time for its actualization, cf. Alexander *in De sensu* 132.1 quoted in De Groot, "Philoponus," p. 182.

because the actualization of the power is incorporeal (330.14–15).[42] Therefore, even though, according to Philoponus, light is not movement and thus does not have speed in the literal sense, it progressively covers enormous distances without taking any time (330.26).

The discussion of light in the commentaries makes occasional reference to experience and observations about light. One occurrence is Philoponus' appeal to experience in the passage quoted above: to suppose that we breathe light not air, is "wildly contradictory to experience." The Greek for "experience" here is *to enargês*, "that which is evident." The reference is not to an experiment chemically analyzing the material we inhale, but, rather, Philoponus points to the way in which we perceive and conceive of the situation. Supposedly the claim that we inhale light would entail that we should see light going into our mouths when we inhale and coming out when we exhale, which obviously is not the case. The arguments about light as movement also referred to experience. The atomists were empiricists but their view that light moves at an unsurpassed speed was not an empirical position in the sense of being based on experiments. Rather experiential evidence was used by those who claimed that light is not movement: since we do not see light moving and we see distant areas illuminated at the same time, for example, at sunrise, light is not movement.[43] Yet others pointed to experience when arguing against the conception that light is the actualized state of the medium: how could this conception explain why solid bodies block light? The suggestion that light is extended without being a solid body was apparently taken to explain this phenomenon.

UNCHANGING KNOWLEDGE OF CHANGEABLE NATURE

With respect to the science of nature, late ancient Platonists confronted a dilemma. Scientific knowledge requires universality and stability but nature is constantly changing.[44] Therefore, how can there be a science of nature? Further, given that natural processes are known to us to the extent they are known through perception, and that our knowledge of the (alleged) universal truths is not derived from perception, what is the relation between perceptual cognition and scientific knowledge? These

[42] Philoponus sometimes suggests that there is local movement in the medium as well. For example, when discussing hearing (*in De an.* 422b12), he talks about light as motion (*kinêsis*) in his explanation of why, in thunder, we see the lightning first and hear the noise only later. According to Philoponus, this is because the objects of hearing are coarse-grained and because the movement of light is swift (*in De an.* 413.6–7); for the observation that lightning reaches us more quickly than the sound of thunder, see also Aristotle, *De sensu* 6, 446b5–6.

[43] Aristotle, *De an.* 2.7, 418b20–6; Philoponus, *in De an.* 344.33–345.13.

[44] As Plato defines it, the essence of the universe is becoming, as opposed to being (see *Timaeus* 27d6–28a1; 28b7–c2); for Proclus on the distinction, see Martijn, *Proclus on Nature*, p. 72.

are persistent problems in the philosophy of science even without the (Neo-)Platonic metaphysical framework. As the Quine-Duhem thesis states, theories are underdetermined by perceptions.[45] Thus even though one might raise suspicions about the complex (Neo-)Platonic metaphysics as a framework for a philosophy of science, the core problems are by no means obsolete.

Proclus had a complex solution to the problem.[46] He built the scientific status of natural philosophy on a methodological parallel between science of nature and geometry. According to Proclus, both establish their theorems from definitions and axioms that must be conceived as hypotheses (see, for example, *in Tim*. 1.237.30–1). As we saw above, he also imported the notion of geometrical conversion to science of nature by turning it from a device of producing theorems from theorems into one that establishes theorems from axioms and definitions. For example, he introduced demonstrations of an efficient cause of the universe (Demiurge) and of its paradigmatic causes (the Platonic forms) from the definition of becoming (the essence of the universe). In the first case, the crucial move was to state that all becoming needs a cause for its becoming,[47] and this cause must be identified with the Demiurge.[48] The paradigmatic causes were then established on the basis of the paradigms being necessary for the Demiurge's arrangement of the universe (*in Tim*. 1.264.24–7).

According to Proclus, the definitions and axioms of natural philosophy were hypothetical, not in the sense that they would be altogether unprovable, but in the sense that they cannot be proved within natural philosophy (*in Tim*. 1.236.32–237.3).[49] Thus the principles of natural philosophy remain proper to that science but they can also be shown to be true on the basis of higher principles. Such endeavor, however, is not a part of the science of nature but belongs to the science of the One, theology (discussed in Proclus' commentary on the *Parmenides*).[50] Thus the parallel to geometry understood in a characteristically Platonic manner is preserved in Proclus' natural philosophy.

45 For the problem of underdetermination and references to the discussion, see Kyle Stanford, "Underdetermination of Scientific Theory," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2013 edn.), at <http://plato.stanford.edu/archives/win2013/entries/scientific-underdetermination/>.

46 Proclus' term is *phusiologia*, which I shall here render as "science of nature" or "natural philosophy." I shall mainly follow the analysis of Proclus' position in Martijn, *Proclus on Nature*, chapter 3, pp. 67–162. I only differ from her reading with respect to the modern scientific problems in terms of which the discussion is framed.

47 See also *Timaeus* 28a.

48 Formulated by Proclus as a syllogism (*in Tim*. 1.258.29) articulated by Martijn, *Proclus on Nature*, pp. 115–16).

49 Martijn, *Proclus on Nature*, pp. 91–7.

50 Martijn (*Proclus on Nature*, a summary on pp. 5–10) argues against Alain Lernould, *Physique et Théologie: Lecture du Timée de Platon par Proclus* (Villeneuve d'Asq: Presses Universitaires du Septentrion, 2001), according to whom natural philosophy is scientific only when reduced to theology.

Proclus also responded to the difficulty of how distinct cognitive faculties concerning changeable perceptible objects and their unchanging transcendent principles interact to make science of nature possible. The crucial move was to interpret our capacity to make judgments (*doxa*) as a capacity to identify perceptible objects by subsuming them under universals.[51] The particulars are perceived and the universals grasped by reason, and the connection is made by the capacity to make judgments that operates between the two. Proclus identified the rational capacity that is crucial for science of nature as *logos*, which he conceived as the lowest layer of reason. Thus scientific knowledge involves the combination of perceptible objects as identified through universals (function of judgment) with the Platonic forms as their causes (function of *logos*).

This layered conception of the human cognitive powers was introduced to bridge the gap between the cognition of natural perceptible things and their transcendent causes. Science of nature understood by comparison to geometry was thus seen as operating on the level intermediate between the transcendent principles and the perceptual layer of reality that is, from an ontological point of view, a reflection of those principles. From this perspective, the claim that the science of nature uses hypothetical axioms can be understood as follows. Someone who has already reached the higher level and grasped the principles of theology also understands how and why the definitions and axioms of natural philosophy are as they are and, presumably, that they cannot be otherwise. By contrast, for someone who is still learning the science of nature, such understanding is not yet available, and thus the *Timaeus* and natural philosophy also have a didactic function.[52]

CONCLUSION

In sum, late ancient philosophical schools produced sophisticated views on questions that we would nowadays understand as belonging to science or the philosophy of science. Many of those views have a strikingly modern ring to them, but we should not overestimate this likeness. Most of the late ancient discoveries were arrived at through philosophical analysis and argument, even though experiments were also conducted. However, we should not underestimate these discussions either. Intricate solutions to philosophical problems that have analogies in modern discussions were discovered. In addition, even though the Platonic-Aristotelian schools ascribed a dominant role to reason in the acquisition of knowledge, this feature did not distinguish them sharply from the programmatic

[51] Proclus, *in Tim.* 1.249.29, discussed in Martijn, *Proclus on Nature*, p. 146.
[52] Here I also follow Martijn, *Proclus on Nature*.

empiricists. For example, Epicurus' claim regarding the unsurpassed speed of light was not an observational claim, neither was his claim of the existence of atoms and void. Further, observations and scientific discoveries quite generally leave open philosophical problems, and with respect to such problems we can appreciate and enjoy the insights of the late ancient thinkers.

16

MEDICINE IN EARLY AND CLASSICAL GREECE

Philip van der Eijk

INTRODUCTION

The material and time frame covered in this chapter lends itself to various narratives, in part complementary, in part conflicting. More conventional accounts would emphasize the emergence and development of scientific, rational, and professional medicine in classical Greece, usually associated with the name of Hippocrates, thus stressing what has long been perceived as one of the most distinctive and influential features of Greek medicine. By contrast, a methodologically more up to date discussion would be pluralistic in orientation: it would highlight the diversity that also characterized Greek approaches to health and disease, and it would focus on how Greek medicine, seen in a global perspective, compares with medical cultures in other non-Western, pre-modern civilizations, how it differed from these, and what they had in common.

The latter type of discussion, which is more fashionable at the time of writing this chapter, ties in with methodological developments within the academic discipline of the history of medicine over the last decades.[1] Several of these are of particular relevance to early and classical Greece. One such development concerns the relationship between the study of the history of medicine and the study of the history of science.[2] By definition, the history of medicine is concerned with historical attempts at scientific and theoretical understanding of health and disease, and it examines historical attempts at practical management of health and at dealing with disease based on such scientific principles (even if what counted as scientific, rational, or professional itself varied over time); and in doing so, it tries to contextualize these

[1] For a recent survey see Mark A. Jackson (ed.), *The Oxford Handbook of the History of Medicine* (Oxford: Oxford University Press, 2011).

[2] See the discussion by Staffan Müller-Wille, "History of Science and Medicine," in Jackson (ed.), *Oxford Handbook*, pp. 469–83, with references to previous discussions of this relationship (e.g. by Henry Sigerist).

theories and practices in their social, cultural, and institutional environment. Yet within the current remit of the discipline, the term "medicine" has acquired a broader meaning and now covers a wide range of varying perceptions, representations, and understandings of phenomena in the sphere of health and disease;[3] it comprises a great variety of responses to pain, illness, and death, both on an individual and on a collective and institutional level; and it includes the patient's perspective (and that of his or her relatives or social circle) as well as the doctor's. This remit thus includes ideas of health and disease and related practices that lay no claim to "scientific" status, not even by the standards that obtained in the time frame in which they were developed. In early and classical Greece (and indeed throughout Greek and Roman antiquity), people's experiences of bodily and mental health, and of coping with disease, pain, and death, were not just the territory of professional doctors but also of priests, diviners, and philosophers; and even if patients' voices rarely surface directly in the surviving Greek and Latin evidence, the representation of such experiences was not just the business of medical writers but also of epic, lyric, tragic, and comic poets, of historians and writers of philosophical treatises, and of vase painters and sculptors.[4]

An account of early and classical Greek medicine along these lines would therefore cover not only the medical theories and activities described or advocated by medical writers such as the authors of the "Hippocratic" treatises, but also the representations of disease, suffering, and healing in the tragedies of Sophocles and Euripides and in Thucydides' account of the Athenian plague; and it would also cover the beliefs and expectations of patients seeking recovery at the healing shrines of Asclepius and Apollo, as testified by inscriptions associated with temple medicine. It would not be restricted to the activities of professional doctors but would also include the responses to issues of health and disease by religious and magical "healers," or by illiterate folk-healers and drug-sellers. And it would present these approaches as coexisting, stressing the presence of rivalry and competing claims to expertise and authority (in accordance with the well-known agonistic tendencies in Greek society) alongside peaceful cooperation and mutual respect.

[3] The historical reconstruction of these phenomena themselves is the subject matter of the history of disease, epidemiology, or more broadly environmental and demographic history, including the history of living patterns, nutrition patterns, and life expectancy. See the general account by Graham Mooney, "Historical Demography and Epidemiology: The Meta-Narrative Challenge," in Jackson (ed.), *Oxford Handbook*, pp. 373–92; for the Greco-Roman world see Mirko D. Grmek, *Diseases in the Ancient World* (Baltimore, MD: Johns Hopkins University Press, 1983); Robert Sallares, *The Ecology of the Ancient World* (London: Duckworth, 1991); and several chapters in Helen King (ed.), *Health in Antiquity* (London: Routledge, 2005).

[4] For an account of representations of disease in these various genres, see Geoffrey E. R. Lloyd, *In the Grip of Diseases. Studies in the Greek Imagination* (Oxford: Oxford University Press, 2003); for visual material see Mirko D. Grmek and Danielle Gourevitch, *Les maladies dans l'art antique* (Paris: Fayard, 1998).

Such an account would include a number of further topics that have been added to the historian of medicine's agenda over the last decades and that are likewise relevant to the ancient world. Thus it would stress the Greeks' engagement with the body, with bodily life, nutrition, sport, fitness, and sexuality, and with youth, old age, and disability;[5] and it would insist that representations of the body in literary and visual media, and archaeological finds relating to material culture and burial customs, are as important as sources of information about the ways the Greeks experienced their own bodies as theoretical discussions by doctors and philosophers.[6] Furthermore, in accordance with the current interest of historians of medicine in such topics as "flourishing," "wellness," and "quality of life," the discussion would no longer be restricted to attitudes to pain, disease, and death, but would also comprise Greek approaches to health, its varying experiences, understandings, and definitions, and the various attempts at management, preservation, and enhancement of health as made in the Greek world, notably in the context of Greek dietetics.

Finally, an account along these lines would point out that Greek medicine comprised the mental as well as the physical, not only in its cognitive aspects but also in the domain of emotions and behavior, and not only the area of mental disease and disorder but also that of mental well-being and flourishing. Indeed, during the period we are concerned with, we see "the mental" emerging as a distinct category and as a domain requiring attention, management, training, and treatment of its own;[7] we see distinctions between the cognitive and the ethical domains being made and challenged; and, again, we observe a variety of approaches and of claims to authority and expertise being made in these domains too, sometimes complementary, sometimes more adversarial.

Yet however appealing this modern style of discussion that I just outlined may be,[8] there is a risk that it ignores historical change and development and refrains from providing explanations for these changes. It is in danger of drawing a somewhat static picture of medicine in the ancient world, in which fundamentally nothing changed, or very little, and in which the features constituting the celebrated diversity of Greco-Roman medicine are represented as coexisting, regardless of whether they alternated or were in competition. In this respect, it would resemble a tendency that has often characterized historiographical accounts of medicine in other ancient cultures, such as Egypt and the Near East: the impression that for hundreds, if

[5] See Brooke A. Holmes, *The Symptom and the Subject. The Emergence of the Physical Body in Ancient Greece* (Princeton, NJ: Princeton University Press, 2010).

[6] For a recent discussion see Patricia A. Baker, *The Archaeology of Medicine in the Greco-Roman World* (Cambridge: Cambridge University Press, 2013).

[7] For a recent collection of papers see William V. Harris (ed.), *Mental Disorders in the Classical World* (Leiden: Brill, 2013).

[8] For a recent attempt see Philip van der Eijk, "Medicine and Health in the Graeco-Roman World," in Jackson (ed.), *Oxford Handbook*, pp. 2–39.

not thousands of years, no fundamental changes took place in medical thought and practice. Regardless of whether such a picture does justice to other medical cultures in the Eastern Mediterranean,[9] it is certainly misleading with regard to Greco-Roman medicine, especially to the period under discussion in this chapter, for no one will question that ideas, attitudes, and approaches to health and disease in the Greek world were very different by the end of the fourth century BCE from what they had been in Homeric times around the beginnings of the eighth century BCE. The challenge for the historian of medicine is, and remains, to describe and explain these changes.

THE DISCOVERY OF HEALTH AND THE EMERGENCE OF THE MEDICAL *TECHNÊ*

Before going into greater detail, it may therefore be appropriate to devote some discussion to what can perhaps be regarded as the most significant development in this time frame: health and disease were no longer regarded as matters entirely beyond human control, as questions of good or bad luck, divine (or demonic) dispensation or natural endowment. In the course of the fifth century, the idea established itself that the preservation and restoration of health were, at least to a certain extent, accessible to human influence, management, and control. The ensuing belief was that people were, again to a certain extent, responsible for their own health or that of their families, the members of their social networks, or the citizens of their *polis*; and that sense of responsibility translated itself in practical arrangements for the provision of medical care, either by means of specially appointed city doctors or by making use of the services of traveling doctors.

Several concepts are particularly relevant here in order to account for this development. First, on a more general level, there is the Greeks' growing fascination, in the late sixth and fifth century, with causation and responsibility (*aitia*): the idea that phenomena are caused by certain factors and mechanisms that are accessible to human understanding and, in some cases, manipulation. Interest in the identification of causes, and in the modus operandi of these causes, becomes manifest in several areas of Greek culture that show strong affinities with medicine, such as natural philosophy, historiography, and law.[10]

Secondly, and relatedly, there is the notion of *technê*, a word conveying study, practice, art, skill, expertise, discipline, or subject area, usually with

[9] For a recent account, stressing development and diversity, see Mark Geller, *Ancient Babylonian Medicine* (Oxford: Wiley-Blackwell, 2010).

[10] See Robert J. Hankinson, *Cause and Explanation in Ancient Greek Thought* (Oxford: Oxford University Press, 1998).

the connotations of practical applicability and teachability. The establishment of the medical *technê* was a process setting in during the fifth century as part of a wider cultural development in the Greek world, the emergence of the *technai*. The idea that there was a *technê* concerned with the specific field of health and disease implied the possibility of control, management, influence, and mastery over something that had long been regarded as being completely inaccessible to human intervention (unless by various sorts of magic). The person possessing that skill, the *technitês* or expert, would be someone who could apply it successfully to specific situations; and the preservation or recovery of health could accordingly be attributed (*aitiasthai*) to that specific activity rather than to chance or other factors. In Plato and Aristotle, we find the specification that the expert is the person who knows what he is doing and why he is doing what he is doing, the person who can give account (*logos*) of his actions, explain, justify, and convey them to others. In this way, it became possible to argue that medical treatment is not a matter of good or bad luck (*tuchê*) but of skillful application of knowledge (*technê*), a skill that can be taught and acquired by a combination of study and practice under the guidance of someone who already masters it. This argument is found in a number of medical writers of the fifth century, and the insistence with which it is made is to be understood in the face of critics who denied the very existence of an art of medicine.[11]

Thirdly, there is the very notion of health (*hugieia*), which is remarkably rare in Greek archaic literature, but prominent in medical and philosophical texts from the late fifth century onwards.[12] Health is seen here as a positive and productive state, comprising much more than just the absence of disease, and constituting an important condition for successful and virtuous life. Moreover, health is considered an *attainable* state, something which one can aim for and make efforts towards, if necessary with the help of an expert.

Expertise raises the question of the recognition of authority, and may bring with it competition for such recognition: who is considered the authoritative expert; how are claims to being an expert made and justified; and how successful are these claims in getting accepted and established? These questions were the subject of lively debates in various parts of the Greek world. Furthermore, expertise raises the question of the role of the non-expert, the *idiôtês*, a notion explicitly developed also in medical contexts, where the non-experts are envisaged as an important target group to whom medical knowledge should be communicated.[13] The preservation and

[11] For a recent discussion, with references to the older literature, see Joel E. Mann, *Hippocrates. On the Art of Medicine* (Leiden: Brill, 2012). See also Mark Schiefsky, *Hippocrates. On Ancient Medicine* (Leiden: Brill, 2005).

[12] See the discussion by Hynek Bartoš, *Philosophy and Dietetics in the Hippocratic On Regimen* (Leiden: Brill, 2015).

[13] The best example is the treatise *On Affections* attributed to Hippocrates, which in its first chapter indicates that the advice it provides is intended for the lay person (6.208 L.). Hippocratic passages

restoration of health was considered not just the expert's responsibility but also, again to a certain extent, that of lay people, the idea being that non-experts, if properly informed, should be able to distinguish situations where they could help themselves (and know what to do in such circumstances) from situations where one would have to seek treatment from experts.

Finally, expertise also raises the question of the boundaries of the relevant subject area, the distinctions between related areas of activity and their defining characteristics, and the extent to which these boundaries can be challenged or even crossed. In the period covered by this chapter, we observe attempts at ever clearer distinction, definition, and legitimization of various areas of activity in the domain of health and disease – the medical *technê*, philosophy, religion, and divination, and within medicine between therapeutics and prophylactics – and at the same time see attempts made to claim certain activities as forming part of one's own agenda.

Thus while doing justice to variety and diversity, and without ignoring or marginalizing the role of religion and magic, it is possible to write the history of Greek medicine in the early and classical period as a process of an increasing awareness of, and demand for, health as an attainable state, and of professionalization as a response to meet that demand. We observe the formation of a distinct discipline comprising not only knowledge of relevant phenomena but also theoretical concepts, methods and practical procedures, and reflection on its own principles; a discipline capable of being taught and communicated to apprentices and the wider public, and owned, shared, and debated by a number of people who regarded themselves as experts; and a discipline that developed a sophisticated technical language and elaborate literary genres in order to convey its ideas and practices to others. This, then, will be the background to the following account of medicine in early and classical Greece, which will attempt to steer a middle course between the two diverging narratives outlined above.

THE ARCHAIC AND EARLY CLASSICAL PERIOD

References in Greek literature to healers and to therapeutic activities of various kinds can be found as early as the Homeric epics. Idomeneus' often quoted statement "A healer (*iêtros*) is a man (*anêr*) equivalent to many others" (*Iliad* 11.514) gives us some idea of the importance attached to men possessing medical skills, even if its significance may be somewhat restricted by the fact that it is made in the context of warfare. Yet the *Iliad* and the *Odyssey* tell us not only of the treatment of battle wounds and casualties, performed by men such as Machaon and Podalirius, but also of

are cited by reference to the relevant volume and page number in the edition by E. Littré, *Œuvres complètes d'Hippocrate*, 10 vols. (Paris: Baillière, 1839–61).

soothing remedies and even of drugs (*pharmaka*) providing psychological comfort in times of grief over the loss of loved ones, prepared by women such as Helen (*Odyssey* 4.219–32). Passages such as these, in addition to archaeological finds,[14] testify to considerable levels of experience in surgery and pharmacology in Homeric times, though our evidence allows us only to guess how widespread this knowledge was. Parallels with Near Eastern and Egyptian medicine suggest that early Greek medical ideas and therapeutic practices drew from a common pool of knowledge spread across the Eastern Mediterranean, though the extent to which Greek medicine was dependent, in its early development, on foreign sources or managed to emancipate itself from this is disputed.[15]

Later in the archaic period, we hear of traveling doctors (*iatroi*), known by name, such as Democedes of Croton (sixth century BCE) or Epimenides of Crete, offering their services (or being invited to do so) to rulers and courts, being held in considerable regard and receiving, one may surmise, substantial remuneration. Information about the professional organization, training, and background of such healers, whether they were working on their own or in groups, families, or guilds (such as the "Asclepiads," mentioned in a famous Coan inscription),[16] and about the nature of apprenticeship is unfortunately scanty. Likewise intriguing is the question of the relationship of these healers to the various healing cults in the Greek world, and to sanctuaries in which deities and heroes were worshipped and approached for their healing powers, such as the cult of Asclepius (more on which below, p. 314). While some of these (semi-)divine healers seem quasi-mythical, such as Amphiaraus and the already mentioned Epimenides, for others, such as Pythagoras and Empedocles, there is a more solid historical basis for their therapeutic activity, which they combined with philosophical teaching, traveling, and involvement in social and political affairs. It seems at any rate clear that by the end of the fifth century BCE, the provision of medical care, i.e. diagnosis and prognosis as well as therapeutics, was not restricted to what we would call professional

[14] On medical ideas and practices as represented in the Homeric epics see Siegfried Laser, "Medizin und Körperpflege," in Hans-Günther Buchholz (ed.), *Archaeologica Homerica*, fascicle S (Göttingen: Vandenhoeck and Rupprecht, 1983); Christine Salazar, *The Treatment of War Wounds in Graeco-Roman Antiquity* (Leiden: Brill, 2000).

[15] For a recent discussion (with references to earlier work) see Markus Asper, "Medical Acculturation? Early Greek Texts and the Question of Near Eastern Influence," in Brooke A. Holmes and Klaus-Dietrich Fischer (eds.), *The Frontiers of Ancient Science. Studies in Honor of Heinrich von Staden* (Berlin: de Gruyter, 2015), pp. 19–46. See also several chapters in Manfred Horstmanshoff and Marten Stol (eds.), *Magic and Rationality in Greek and Near Eastern Medicine* (Leiden: Brill, 2004). For an older collection of parallels see Dietlinde Goltz, *Studien zur alt-orientalischen und griechischen Heilkunde. Therapie, Arzneibereitung, Rezeptstruktur* (Wiesbaden: Steiner Verlag, 1974).

[16] See Susan M. Sherwin-White, *Ancient Cos* (Göttingen: Vandenhoeck and Rupprecht, 1978), pp. 256–89, and the more skeptical account by Wesley D. Smith, *Hippocrates. Pseudepigraphic Writings* (Leiden: Brill, 1990), pp. 6–18.

doctors but also included shamanistic healers and religious officials in temples.

Our main sources of this early period are literary and historiographical texts, such as references to healers in the mythical sections of Pindar's victory odes and accounts of the activities of Greek and Egyptian doctors in Herodotus.[17] Another important source of information is Greek tragedy, which offers impressive (though no doubt stylized) accounts of human suffering, both physical and psychological, and of pathological behavior, such as the madness of Ajax and Heracles or the painful agonies of Philoctetes.[18]

These literary texts also show that in the Greek world there was considerable variety in attitudes and responses to disease. Even what counted as disease in the first place was by no means self-evident. The term *nosos* is both wider and narrower in significance than the English "disease," since it also covers psychological phenomena (such as Phaedra's lovesickness in Euripides' *Hippolytus*), while conversely not everything we would consider disease is covered by the Greek term. With the very definition of disease being in flux, it is not surprising that corresponding perceptions and approaches were so too. Thus it was far from clear that disease was in all cases something one should try to do something about; resignation, acceptance, or avoidance of the sick are equally well-attested attitudes, as is shown by the case of the Greek warrior Philoctetes, whose affliction (caused by the bite of a snake, but possibly indirectly by divine punishment for a religious offence) made him into an outcast.[19] Especially epidemic disease, *loimos*, was – and continued throughout antiquity to be – approached with caution, not only because of the obvious problems it posed in terms of management, but also because it continued to be associated with divine causation, even in times when naturalistic explanations of disease had become firmly established. And even if action was to be taken, the question was: what action? And who were the experts to answer that question? Doctors were by no means the only recognized authoritative people deciding on the course of action to be taken. Medicine was competing here with magical and religious approaches, and even within the medical domain there was no unanimous attitude or response to disease (see below section entitled "Medicine, Religion, and Magic").

17 See the passages assembled in Lloyd, *In the Grip of Diseases*.

18 See Jennifer Clarke-Kosak, *Heroic Measures. Hippocratic Medicine in the Making of Euripidean Tragedy* (Leiden: Brill, 2004); Alessia Guardasole, *Tragedia a medicina nell'Atene del V. secolo A.C.* (Naples: D'Auria, 2000); Jacques Jouanna, "Hippocratic Medicine and Greek Tragedy," in his *Greek Medicine from Hippocrates to Galen* (Leiden: Brill, 2012), pp. 55–80.

19 See the discussion by Fridolf Kudlien, *Der Beginn des medizinischen Denkens bei den Griechen* (Zurich and Stuttgart: Artemis, 1967), pp. 106–23.

MEDICAL WRITING IN THE CLASSICAL ERA: "HIPPOCRATIC" AND OTHER MEDICAL WORKS

We get a sense of this competitive climate from some of the so-called "Hippocratic" writings; and this takes us to the mid-late fifth century BCE, when the evidence for medical thought and practice suddenly becomes much more substantial and informative. A considerable number of technical prose texts devoted to medical topics and, as far as we can tell, written (in Ionic Greek dialect) by authors with considerable medical experience survive in their entirety. They include specialized treatises about various kinds of diseases, about therapeutic techniques, about food, drink, and medicaments, about the nature and structure of the human body, and about specific functions such as nutrition, reproduction, and sense perception. They also include long series of case histories of individual patients, recording day by day their symptoms and responses to treatment. Furthermore, there are works devoted to diagnosis and prognosis and to the interpretation of symptoms as signs pointing to states hidden deep down in the body. And we find works reflecting on the medical art, its methodological principles, and its ethical aspects.

Clearly, these texts not only served the purpose of storing information, but they also played an important role in medical teaching. There are also texts that seem to envisage a broader readership than just experts or apprentices and which show considerable levels of literary craftmanship and rhetorical skill, very similar to what we find in other literary genres of the same period, especially oratory. And indeed some of the Greek medical texts of the fifth and early fourth century may well have been presented as oral speeches in front of a listening audience.[20]

This rhetorical feature of classical Greek medical literature is closely related to the development of rhetoric and public speaking in Greek culture, both in the law courts and in political assemblies, and it is usually explained by reference to the already mentioned competitive climate in which medicine had to establish itself. We can read discussions about the medical art itself (*technê iatrikê)*, suggesting that this is a new, controversial phenomenon trying to gain acceptance in the face of skepticism and opposition. Furthermore, we get a strong sense of the modes and the extent of communication and dissemination of medical ideas by means of a specialized vocabulary and through the use of a number of different textual genres, clearly meant to reach audiences wider than just experts and also including lay people.[21] Medicine in classical Greece was not just a matter of skillful

[20] See Jacques Jouanna, "Rhetoric and Medicine in the Hippocratic Corpus. A Contribution to the History of Rhetoric in the Fifth Century," in his *Greek Medicine from Hippocrates to Galen*, pp. 39–54.

[21] See Philip J. van der Eijk, "Towards a Rhetoric of Ancient Scientific Discourse. Some Formal Characteristics of Greek Medical and Philosophical Texts (Hippocratic Corpus, Aristotle)," in

application of knowledge, it was also a linguistic and literary exercise requiring a thorough command of technical vocabulary and the ability to tell a good, plausible, and persuasive story to an audience of experts as well as non-experts. Moreover, parallels in some non-medical texts such as Euripides' tragedies and Thucydides' history of the Peloponnesian War testify to the spread of medical concepts, ideas, and specialized medical terminology beyond professional circles.

Who were the authors of these medical texts, and in what context or institutional setting did they work? It has long been customary in scholarship on ancient medicine to associate the above developments with the name of the Greek doctor Hippocrates and with his school of medicine, allegedly founded on the island of Cos. Yet it is by now widely accepted that Hippocrates was one among many doctors and medical writers, and that the development of the medical *technê* and the proliferation of medical ideas in the late fifth and fourth century BCE was more than just a one-man operation. Indeed, there is positive evidence for the existence and impact, in the fifth and fourth century, of a large number of medical doctors, many of whom are still known to us by name, who practiced medical care and wrote medical treatises, some of which survive in their entirety, others in fragments.[22] By contrast, in the case of Hippocrates, it is difficult to identify his individual contribution to the development of medicine. This is because very little information about his life survives,[23] and because we do not know which of the many writings handed down under his name were actually written by him. That a doctor named Hippocrates existed, that he held certain characteristic medical theories, which he put into practice and disseminated to others, that he taught medicine (and charged for this), and that already in his lifetime he was known as a prominent physician is widely accepted. Yet it is also universally agreed that Hippocrates, or indeed any single medical writer, cannot have been the author of the sixty-plus works that were attributed to him from antiquity onwards, let alone of the hundred-and-seventy-plus medical works attributed to him in later times,[24]

Egbert J. Bakker (ed.), *Grammar as Interpretation. Greek Literature in its Linguistic Contexts* (Leiden: Brill, 1997), pp. 77–129.

22 For a list of these names see Philip J. van der Eijk, "On 'Hippocratic' and 'Non-Hippocratic' Medical Writings," in L. Dean-Jones, R. J. Hankinson and R. Rosen (eds.), *What is Hippocratic about the Hippocratic Corpus?* (Leiden: Brill, 2015). The most important source for the existence and the theories of these doctors is the so-called Anonymus Londiniensis, a Greek text surviving on papyrus and dating in its present form probably to the first century CE but incorporating material of much earlier date, notably an Aristotelian account of the opinions of a large number of medical writers and philosophers on the causes of health and disease. For a discussion see Daniela Manetti, "'Aristotle' and the Role of Doxography on the Anonymus Londiniensis (PbrLibr Inv. 137)," in Philip J. van der Eijk (ed.), *Ancient Histories of Medicine* (Leiden: Brill, 1999), pp. 95–141.

23 For a discussion of the evidence see Vivian Nutton, *Ancient Medicine* (London: Routledge, 2004), pp. 53–60.

24 For a survey of all the works (preserved and lost) ever attributed to Hippocrates see Gerhard Fichtner, *Corpus Hippocraticum. Verzeichnis der hippokratischen und pseudohippokratischen Schriften* (Tübingen: Institut für Geschichte der Medizin, Universität Tübingen, 2011), with

for there is great diversity in style, doctrine, intellectual background, date, and geographical provenance in the writings usually assembled under the name "Hippocratic Corpus." While some scholars still believe that Hippocrates was the author of at least some of these writings, there is no unanimity as to which of these writings these are. This is why most scholars nowadays prefer to speak of "the Hippocratic writings" rather than the works of Hippocrates, "Hippocratic" in such contexts meaning nothing more than "attributed, at some point during its transmission in the ancient and/or early medieval period, to Hippocrates," without the use of that expression implying any commitment to the plausibility of that attribution.

These "Hippocratic" writings were later assembled into a collection or "Corpus," and it is in this collective form that they have been handed down, presumably already from late antiquity onwards, and presented as Hippocrates' *opera omnia* in the first printed editions of the Renaissance, which were the basis for subsequent editions such as the authoritative *Oeuvres Complètes d'Hippocrate* by Émile Littré in the nineteenth century. In reality, however, they were not conceived or written as part of a joint enterprise with the aim of producing a collective output.[25] While some of the works show great intellectual affinity with each other, there are also significant differences in doctrine and style, and sometimes there is even explicit disagreement or criticism in one "Hippocratic" work of ideas expressed in another. Some scholars have tried to identify a core of works representing a reasonably coherent set of ideas and practices, and these are usually believed to stem, if not from Hippocrates himself, at least from his intellectual circle; among these are well-known works such as *Epidemics* 1 and 3, *Prognostic, Airs Waters Places, Sacred Disease, Joints, Fractures, Ancient Medicine*, and *Nature of Man*. While some of these works appear to have been written by an author with a strong personality or intellectual identity, there are also indications suggesting collective, or at least multiple authorship. Thus there are good reasons to think of the medical information reflected in the *Epidemics* as the result of a collective gathering exercise by a number of doctors for the purposes of organization and storage of information and teaching. Other writings in the "Corpus" dating from the same period are intellectually, methodologically, and stylistically more remote from this core and do not breathe the same atmosphere. Among these are the works sometimes referred to as "Cnidian," since they were believed to derive from a rival medical school in Cnidos, a town on the west coast of Asia Minor not far from the island of Cos. While this belief in the existence of a Cnidian school of medicine has now been abandoned, it is undeniable that

updates online at http://cmg.bbaw.de/online-publikationen/hippokrates-und-galenbibliographie-fichtner.

[25] For a discussion see Elizabeth M. Craik, *The "Hippocratic" Corpus. Content and Context* (London: Routledge, 2014); Werner Golder, *Hippokrates und das Corpus Hippocraticum* (Würzburg: Königshausen und Neumann, 2007).

writings such as *Internal Affections*, or *Diseases* 2 and 3, are rather different in character from the works mentioned in the previous paragraph. How, when, and for what reason these writings came to have the label "Hippocrates" attached to them, and who was responsible for these attributions, is difficult to say.

The formation of the "Hippocratic Corpus" was a long process anyway, and there is good reason to think that over the centuries the Corpus grew, other treatises being added to it at a later stage. The formation of this Corpus started probably in Hellenistic Alexandria, though it is possible that Aristotle's school in Athens, the Lyceum, also had something to do with it.[26] Yet the establishment of this Corpus, and the canonization of the writings it included, had two paradoxical consequences. First, while these writings survived, possibly for the very reason that they were attributed to Hippocrates, their real authors' names disappeared from the record. Secondly, other medical writers from the fifth and fourth centuries BCE apart from Hippocrates were marginalized and their works were lost, though their names still survive.

MEDICINE, PHILOSOPHY, AND SCIENCE

For these reasons, the present account does not make a categorical division between "Hippocratic" and "non-Hippocratic" medical writings, and it treats relevant writings attributed to Hippocrates as individual works, not as part of a group or collection, and on a par with the works and fragments of other medical, scientific, and philosophical writers of the same time frame.

Furthermore, it includes also those writers who are generally believed not to have been practising doctors but who were nevertheless concerned, to a considerable extent, with medical theory. Many Greek philosophers developed an interest in medicine, the body, and the causes of health and disease, and devoted treatises to the subject.[27] The early Greek philosophers' "inquiry into nature" (*historiê peri phuseôs*), already mentioned above, extended also to the nature of the body and, indeed, the nature of disease, as testified by Democritus and Empedocles; and many early Greek philosophers were interested in questions about reproduction and inheritance. When we get to Plato, Aristotle, and Theophrastus, we have entire works preserved that testify to a profound interest in medical topics and, in the case

[26] The continued usage of the designation "Hippocratic Corpus" (up to the present day) makes an artificial separation between "Hippocratic" and "non-Hippocratic" medical writers from the same time frame, and silently presupposes that the writings assembled in the "Hippocratic Corpus," for all their acknowledged diversity, have a distinct identity or core that connects them more closely among each other than with the writings of these "non-Hippocratic" authors. For a discussion of all these issues see Philip J. van der Eijk, "On 'Hippocratic' and 'Non-Hippocratic' Medical Writings."

[27] See Philip J. van der Eijk, *Medicine and Philosophy in Classical Antiquity* (Cambridge: Cambridge University Press, 2005).

of Aristotle, explicit discussion of the close connections between medicine and the study of nature. This was because medicine, although ultimately – to put it into Aristotelian categories – a practical or even productive art, had a scientific, theoretical side to it, at least for some Greek thinkers: it was the knowledge of health and disease, their nature and causes, of bodily functioning and well being, and it was the knowledge of the natures and powers of therapeutic substances and remedies and of the causes and modes of their efficacy or nonefficacy. For some, it also included knowledge about preservation of health and the prevention of disease by prophylactic measures such as gymnastics, diet, and exercise. And for some, it was the knowledge not only of the body, whether in its healthy, functioning state or in its diseased, malfunctioning or non-functioning state, but also of the mental, psychological side of life. Moreover, it included not only humans but also animals and even plants, and thus required comprehensive knowledge of biology.

This whole area of medical knowledge was considered a sub-area of the broader field of physics, or knowledge of the natural world, and as such it was also included in Plato's discussion of the cosmos in the *Timaeus*, a substantial part of which is devoted to the structure and functioning of the human body and to the origins of disease, both physical and mental.[28] Medicine was also included in the various sciences that were studied in Aristotle's Lyceum.[29] While Aristotle himself alluded to a treatise on the causes of health and disease, several members of the Peripatetic school, such as Theophrastus and Strato, devoted specialized treatises to medical topics such as fatigue, dizziness, sweating, epilepsy, melancholy, respiration, and epidemic disease; and the (largely Pseudo-)Aristotelian *Problemata physica* testifies to extensive research in this domain. Even if there is no evidence that these Peripatetics were engaged in regular medical practice, it is hard to believe that they refrained from application of their theoretical ideas: we may suppose that structured empirical observation and, perhaps, a limited degree of empirical testing were practiced, though probably not with the purpose of treating patients but rather in order to assess the efficacy of certain therapeutic measures or the veracity of claims made about their alleged "powers" or "capacities" (*dunameis*) or "virtues" (*aretai*).

Furthermore, one of the characteristic features of Greek medicine is the presence of second-order knowledge, i.e. reflection upon the discipline itself,

[28] Plato's status as an authority in medicine in antiquity is testified by his prominent position in the Anonymus Londiniensis (above, footnote 22) and by Galen's extensive engagement with the *Timaeus*. On Plato's medical interests see Mario Vegetti, *La medicina in Platone* (Venice: Il cardo, 1995); James Longrigg, *Greek Rational Medicine* (London: Routledge, 1993), pp. 104–48; Lloyd, *In the Grip of Diseases*, pp. 142–75; Susan B. Levin, *Plato's Rivalry with Medicine* (Oxford: Oxford University Press, 2014).

[29] See Philip J. van der Eijk and Sarah R. Francis, "Aristoteles, Aristotelismus und antike Medizin," in Christian Brockmann, Wolfram Brunschön, and Oliver Overwien (eds.), *Antike Medizin im Schnittpunkt von Geistes- und Naturwissenschaften* (Berlin: de Gruyter, 2009), pp. 213–34.

its principles and methods, the ways in which it can gain knowledge and apply this knowledge in practice, and the levels of certainty it can attain. Greek medical writers developed not only a vocabulary to refer to bodily features, structures, and processes, but also "ordering" concepts such as nature, cause, sign, and power. Most prominent is the concept of nature, *phusis*, which appears usually in the sense of the nature of things or phenomena, such as the nature of the body, the nature of a particular therapeutic substance, or the nature of a disease, rather than in the sense of an overarching concept (Nature with a capital N). *Phusis* here conveys the sense of a normal pattern of origin and growth: thus the statement made by the author of *On the Sacred Disease* to the effect that epilepsy, like all other diseases, is not sent by the gods but "has a nature,"[30] is to be understood in the sense that diseases originate, develop, and behave in a certain, regular way that is open to human understanding and therapeutic correction. Likewise, Greek medical writers developed a sophisticated understanding of causation, distinguishing between structural and antecedent causes (*aitia, prophasis*), auxiliary causes (*sunaition*), and necessary conditions, and between "powers" or capacities (*dunameis*) and activities or effects (*energeiai, erga*).[31]

Yet the above areas of overlap between medicine and philosophy should not suggest that the relationship between them was entirely unproblematic. For alongside the story of the origins of medical inquiry in Ionic natural philosophy and the emergence of a medical *technê* that immediately shows the signs of heavy influence of early Greek philosophy, there is also the narrative of the rivalry between medicine and philosophy.[32] Critical reactions against certain strands within this philosophical medicine were expressed by the authors of *On Ancient Medicine* and *On the Nature of Man*, who rejected the simplifying schema of the four elementary qualities hot, cold, dry, and wet or the reduction of bodily processes to just one principle. Their criticism did not, however, arise from a hostile attitude to philosophical speculation as such, nor from an exclusively empiricist standpoint, but from a concern to do justice to what they perceived as the variety and complexity of bodily structures and processes and their interaction with the environment. The ambivalent attitude of doctors to theoretical ideas and methods continued throughout the fourth century and well into the Hellenistic period, with doctors such as Diocles of Carystus and Mnesitheus of Athens engaging with philosophical thought, yet also

[30] *On the Sacred Disease* 2.1 (6.364 L.).

[31] See Philip J. van der Eijk, "The Role of Medicine in the Formation of Early Greek Thought," in Patricia Curd and Daniel W. Graham (eds.), *The Oxford Handbook of Presocratic Philosophy* (Oxford: Oxford University Press, 2008), pp. 390–98.

[32] Interestingly, in Celsus' historiographical introduction to his *On Medicine* (first century CE), Hippocrates is presented as the one who separated medicine from philosophy (*Medicina*, Preface, section 8).

expressing dissent and criticism, e.g. Diocles' cautionings about the scope and relevance of causal explanation in medical practice.[33] We see here the foreshadowing of the divide between the medical sects of the Rationalists (or Dogmatists) and the Empiricists that was to follow in the early Hellenistic period.

OBSERVING AND CONCEPTUALIZING THE BODY

A major development in the relevant period, highlighted by recent scholarship, is the discovery of "the living body" (*to sôma*) as an entity in its own right, as distinct from other entities by which it is surrounded (such as the environment) or inspired (such as the soul), as an object of inquiry, and as a category under which a number of items and features can be subsumed.[34] In Greek medical thought, the body is substantivized, thematized, and individualized. It is the subject or carrier of a number of features and accidents; it is also the carrier or container of a number of physical qualities and substances. And the body is what the doctor's attention primarily focuses on as the most relevant category of medical diagnosis and treatment: it represents the unit, the most immediately relevant context in which diseases or complaints are believed to be located and to be accessible to understanding, and the locus that needs to be targeted in order for health to be restored.

Yet how was the body approached, and what ways were open to Greek doctors to obtain information about the body's condition? One of the most striking features of Greek medicine in the classical period is the development of systematic observation. The clinical case histories of the *Epidemics* provide hundreds of examples of observation of individual patients; furthermore, in one of the more theoretical sections of the *Epidemics* we have a checklist of things to look for when examining a patient, the categories or key words in a medical questionnaire.[35] The author presents his method as proceeding from observations of a number of individual cases (*idia*) towards the formation of a general view on the common (*koinos*) nature of all people. The text also provides more generalized accounts in terms of the so-called "constitutions" (*katastaseis*), in which information about climate, weather, and other environmental circumstances of specific areas in particular time frames is recorded. This reflects the underlying belief in the importance of the environment to the Greeks' understanding of health and disease.

[33] See Philip J. van der Eijk, "Between the Hippocratics and the Alexandrians: Medicine, Philosophy and Science in the Fourth Century BCE," in Robert W. Sharples (ed.), *Philosophy and the Sciences in Antiquity* (Aldershot: Ashgate, 2005), pp. 72–109.

[34] See Holmes, *The Symptom and the Subject*.

[35] *Epidemics* 1.23 (2.668–70 L.).

Evidence for dissection is more problematic. There certainly are passages in the Hippocratic writings that mention or suggest dissection as a source of information about internal structures and processes; and there is some evidence for dissection practice in Democritus. Yet it is not before Aristotle that we get a clear sense of more systematic dissection, practiced on animals.[36] Dissection of human bodies (corpses) was not allowed in the classical period, and knowledge of internal human anatomy and physiology was arrived at only superficially and accidentally, in the treatment of trauma, or indirectly through the examination of animals. Ideas about internal human anatomy[37] in the "Hippocratic" writings were accordingly vague and unspecific, and in most cases based on imagination or analogy rather than autopsy. Thus we find rough divisions between "the region of the bones" and "the cavities,"[38] or between "structures" (*schêmata*) and "capacities" (*dunameis*),[39] or between "solid parts, soft parts, and fluids."[40] In general, medical writers in their discussions of the body distinguished between parts (e.g. the heart, the brain, the blood vessels), substances (water, bile, phlegm), properties (hot, cold, sweet, salt), relationships (balance, imbalance, proportion, tension), processes (e.g. cooling, drying), and functions or capacities (e.g. sensation, nutrition, generation).

As we get into the fourth century, the evidence for dissection becomes more tangible. Aristotle discusses the importance (but also the disadvantages and the limitations) of dissection in the context of his zoological works, more specifically the study of the "parts" (*moria*) of the bodies of organisms; and there is no doubt that Aristotle and his pupils performed a number of dissections on a variety of animal bodies and drew inferences from this about the internal anatomy of the human body. It is therefore not surprising that the most developed account of human anatomy in this period can be found in Aristotle, especially in his *Parts of Animals*. Here, however, the study of anatomy is not part of a medical project but part of a larger investigation dealing with the "parts" of animals, looked at from a comparative perspective. Aristotle not only compares the parts of animal bodies in terms of their shape and structure but also viewed in relation to their natural end or function (*telos*). Moreover, Aristotle distinguishes a number of "activities and affections" (*praxeis kai pathê*) that all living beings have in common, such as sensation, memory, sleep and dreams, youth and old age, nutrition

[36] See Geoffrey E. R. Lloyd, *Magic, Reason and Experience* (Cambridge: Cambridge University Press, 1979), pp. 126–225.

[37] Here used in the sense of "anatomical structure," i.e. the bodily parts and structures that can be revealed by dissection; in Greek, the verb *anatemnein* refers to the activity of dissection, while the noun *anatomê* may refer both to that activity and to its results in terms of what is being revealed by the activity; the latter seems to be envisaged in the title of a lost Aristotelian work, *Anatomai*, which also appears to have contained diagrams and other visual representations of bodily structures.

[38] *On the Art* 10 (6.16 L.).

[39] *On Ancient Medicine* 22 (1.626 L.).

[40] *On Generation* 3 (7.474 L.).

and growth, respiration, generation, and locomotion. Aristotle's discussions include much that is relevant to medicine, such as observations about the anatomy and physiology of a large number of bodily organs, theoretical discussions about nutrition, growth, and reproduction, and theories about the functioning of the living body through the presence of a number of substances and entities, such as vital heat, *pneuma*, and blood. Occasionally, Aristotle also includes discussion of dysfunction and disease, though not for its own sake but because of the light it sheds on particular aspects of the healthy situation.

Likewise, other fourth-century medical writers such as Diocles and Praxagoras are reported to have written whole treatises on dissection(s), entitled *Anatomê* or *Anatomai*. Diocles refers to the dissection of the uterus of mules, and his later contemporary Mnesitheus of Athens refers to the intestines of an elephant, apparently also on the basis of empirical examination.[41]

DIAGNOSIS AND PROGNOSIS

On the basis of inference from observation and dissection, Greek medical writings of this period develop elaborate procedures for the diagnosis and prognosis of pathological conditions. The "Hippocratic" *Epidemics* and the *Prognosticon* present checklists of features in the patient's bodily condition to which the doctor should pay particular attention, either by direct physical examination or by questioning the patient (or his/her relatives), such as the color of the skin, urine, excrements, speech, breathing, sleep, and dreams. On the basis of this, and sometimes via an explicit procedure of inference from specific signs to an underlying (but invisible) physical state, the doctor forms an assessment (*krisis*) of the patient's condition, which is usually connected with a forecasting of the patient's chances of survival and recovery, or the opposite. This procedure was captured in the formula *opsis adêlôn ta phainomena*, "what is manifest offers a view of what is obscure."[42] This prognostic activity was considered of great importance, as successful prognosticating was regarded as evidence of the doctor's competence, which in turn was of major importance to a doctor's overall reputation.

[41] Diocles, fr. 24 in Philip J. van der Eijk, *Diocles of Carystus*, 2 vols. (Leiden: Brill, 2000–1), vol. 1: *Text and Translation*; Praxagoras, frs. 11, 12, and 67 in Fritz Steckerl, *The Fragments of Praxagoras of Cos and his School* (Leiden: Brill, 1958); Mnesitheus, frs. 17 and 52 in Janine Bertier, *Mnésithée et Dieuchès* (Leiden: Brill, 1972).

[42] The expression is variously attributed to Diocles (fr. 56b in van der Eijk, *Diocles*, vol. 1), Anaxagoras (H. Diels and W. Kranz, *Die Fragmente der Vorsokratiker* (10th edn; Berlin: Weidmann, 1961), 59 B 21), and Democritus (ibid., 68 A 11).

THEORIES OF HEALTH AND DISEASE

These observations were not made in isolation. They were often informed, implicitly or explicitly, by general theoretical concepts and medical theories. As Langholf has shown, implicit understandings of bodily structures, features, and processes as well as nosological entities and mechanisms are reflected in the selection of phenomena the doctors are interested in, the kinds of questions they ask, and the vocabulary in which they express their observations.[43] Here, terms such as *krisis* (the critical turning point in the course of a disease) and *pepsis* ("cooking" or "ripening," believed to apply to physical or pathological material in the body) come to the surface at an early stage.

As already indicated above, Greek doctors and philosophers developed concepts of health and ideas about what health consisted in, what brought it about, and how it might be disturbed. Likewise, they developed ideas about what constitutes disease and about its causes. There was great variety of opinion on these matters, as testified by the evidence provided by the already mentioned Anonymus Londiniensis (see above, footnotes 22 and 28), but some recurrent tendencies can be observed. Some medical writers identified a single substance, for example air (*aêr, pneuma, phusa*), as the key ingredient both to healthy and to pathological processes. In other writings, we see a binary model, with bile (*cholê*) and phlegm (*phlegma*) being identified as the key factors in constituting health and disease. In addition, health and disease were conceptualized in terms of "balance" or "imbalance" between constituents, such as the four humors blood, phlegm, yellow bile, and black bile in the influential treatise *On the Nature of Man*, or of one entity or element being in excess and "dominating" the others, or being segregated from others, or in terms of the formation of specific substances (such as *perittômata*, "residues"), or in terms of specific processes such as the obstruction of the flow of air, *pneuma*, or blood through bodily passages caused by phlegm, or the chilling of vital substances such as blood.

THERAPEUTICS

How did Greek doctors of the classical period respond to the phenomena of disease and death, pain and suffering?[44] As already mentioned, one important question that required answering before any kind of intervention was

[43] Volker Langholf, *Medical Theories in Hippocrates. Early Texts and the Epidemics* (Berlin: de Gruyter, 1974).

[44] See the collection of studies edited by Ivan Garofalo, Alessandro Lami, Daniela Manetti, and Amneris Roselli (eds.), *Aspetti della terapia nel Corpus Hippocraticum* (Florence: Olschki, 1999).

considered was whether to engage in treatment at all: the diagnosis of an individual patient consisted not only in a determination or classification of the condition he or she was suffering from but also in a prognostic assessment of the chances that the patient would survive or die. Yet this should not be understood in the sense that a determination of the hopelessness of the case always led to a decision not to treat the patient. It is worth stressing this, for it is sometimes argued that Hippocratic doctors did not treat hopeless cases. This claim is exaggerated, for we have many examples in the texts where specific kinds of treatment are recommended even in contexts where the disease is said to be "mortal."[45] It is of course possible that the use of the word "mortal" (*thanasimos, thanatôdês*) in these cases refers to a possibility or likelihood obtaining for the most part, without excluding the possibility that the patient may nevertheless survive; and it could be that treatment in such cases is recommended in the hope that the patient happens to belong to the small minority of cases that make a recovery. Yet there are good reasons to assume that Greek doctors treated patients of whom they were certain they would die; in such cases, treatment seems to aim for the most tolerable condition. This is related to the belief, held by a number of Greek doctors, that even in cases of incurable chronic disease, medicine was not irrelevant and could, and should, contribute to bringing about the most tolerable condition, even if this changed nothing about the lethal nature of the condition.

Methods of treatment were traditionally divided into the three categories of dietetics, pharmacology, and surgery, with cauterization sometimes added as a further category. In practice, the treatment of specific diseases (as elaborated in great detail in designated therapeutic texts) often consisted of a combination of different measures, with the details of these combinations sometimes being left to the discretion of the doctor, as indicated by phrases such as "if you wish," "if you judge this appropriate." Clearly, a gap was perceived between fixed, written instructions and practical implementation in ever changing individual circumstances, a gap that could not be covered adequately in words and where room for adaptation and adjustment had to be left to the practicing healer.

PREVENTION AND PROPHYLAXIS

Yet the response was not just reactive, but also pro-active: Greek doctors were as equally concerned with the prevention of disease, the preservation of

[45] For a discussion of this issue, and of the most important passages, see Heinrich von Staden, "Incurability and Hopelessness: The Hippocratic Corpus," in Paul Potter, Gilles Maloney, and Jacques Desautels (eds.), *Maladie et maladies dans la Collection Hippocratique* (Québec: Éditions du Sphinx, 1990), pp. 75–112 and Philip J. van der Eijk, "To Help, or To Do No Harm. Principles and Practices of Therapeutics in the Hippocratic Corpus and in the Work of Diocles of Carystus," in his *Medicine and Philosophy*, pp. 101–18.

health, and the enhancement of what we would call the quality of life as with the treatment of disease. Regimen in health (*Hugieina*) was developed as a sub-discipline within the medical art, later explicitly distinguished from therapeutic regimen; and extensive lifestyle programmes were developed. The main texts here are the "Hippocratic" work *On Regimen* and the *Matters of Health* by Diocles of Carystus.[46] We observe here a concern with healthy and wholesome nutrition, based on a catalog of foodstuffs with their corresponding "capacities" (*dunameis*), i.e. their powers to produce certain effects in the body of the organism taking in the nourishment. We also see a concern with exercise and fitness, and more generally with the spelling out of a pattern of rules for healthy, balanced living. Diocles also emphasized the appropriateness (*to harmotton*) of such rules to living conditions and personal circumstances, and he (and other dietetic writers, such as his younger contemporaries Mnesitheus and Dieuches) developed regimens for younger people, for the elderly, for people traveling, and for people sailing.

MEDICINE OF THE MIND

Finally, and relatedly, Greek medicine, especially in its more philosophical versions, also developed diagnostic, prophylactic, and therapeutic approaches to mental health and mental illness. In the "Hippocratic" writings, we find a rich spectrum of what we would call psychological disorders, usually presented alongside somatic complaints as part of a bigger picture of symptoms characterizing certain illnesses; we also find more explicitly developed clinical concepts such as *phrenitis*, *mania*, and *melancholia*.[47] The treatment of these conditions consisted mainly of dietary and pharmacological measures. A case in point is the author of *On Regimen*, who in chapters 35–6 sets out a wide-ranging account of various kinds of mental disturbance, grouped under the rubric of *phronêsis* and *aphrosunê*, terms that cover "intelligence" but also "consciousness," "cognitive awareness," and their corresponding dysfunctioning. His clinical accounts of these conditions are accompanied by physiological explanations in which the elements fire and water play a major role: variations in the proportional relationship between these two elements, and in their degrees of wetness and dryness, give rise to various complaints in the sphere of sensation, thinking, memory, concentration power, and emotions. The treatment recommended for these conditions is likewise cast in terms of a corrective regimen consisting of specific eating and drinking measures, specific forms of exercise, sexual

[46] See the discussion by Georg Wöhrle, *Studien zur Theorie der antiken Gesundheitslehre* (Stuttgart: Steiner, 1990).

[47] See Silvia Matenzoglu, *Zur Psychopathologie in den hippokratischen Schriften* (Berlin: dissertation.de, 2011), and Chiara Thumiger, *A History of the Mind and Mental Health in Classical Greek Medical Thought* (Cambridge: Cambridge University Press, 2017).

activity, and various other activities, all aimed at restoring the balance that has been disturbed.

Plato and Aristotle develop more philosophical approaches. Plato's *Timaeus* devotes a long section to what it calls "diseases of the soul" (86b ff.), and this term is not just a metaphor for ignorance or stupidity in the cognitive, philosophical sense of the word (as was to be the case in Stoic discussions of the treatment of emotions). In Plato, it also refers to clinical conditions such as epilepsy and mania, and to a pathological lack of self-control when it comes to bodily needs and desires. Aristotle in book 7 of his *Nicomachean Ethics* refers in a similar context to a condition called *akrasia*, "lack of self-control," a state that could be compared to what is nowadays referred to as addiction to various kinds of physical stimulation. Both Plato and Aristotle recommend specific programs of moral and intellectual formation and training, both as prophylactic safeguards and as corrective therapeutic agents, alongside a physical regime of exercise and a frugal diet.

MEDICINE, RELIGION, AND MAGIC

This account has so far largely concentrated on "secular" medicine. Yet, as already indicated above, it would be a gross mistake to believe that this was the whole story, or, even worse, that in the classical period secular medicine triumphantly and definitively marginalized religious medicine. First of all, the whole concept of secular medicine is presumably anachronistic, for no Greek doctor would have thought of his own activities as being completely independent of any relationship to the divine. Yet we need to distinguish tacit beliefs and presuppositions present in the background from explicit references to the divine in the medical texts; and of course what matters is what role is assigned to the divine in these contexts. On the latter front, it is noteworthy that very few medical texts make explicit reference to the divine, either as a cause or as a source of health and healing. The famous "Hippocratic" treatise *On the Sacred Disease* denies that the gods have any role in the causation of disease, an idea that he rejects as being "godless," *atheos*, and "impious," *asebês*.[48] However, that does not prevent him from positively saying that all diseases are divine,[49] for this latter point is meant in a different sense, i.e. that diseases, in so far as they are natural and display regularity, are divine to the same extent that all other natural phenomena are divine. Yet he also describes, with obvious endorsement, the rituals that people should perform when approaching the gods in temples for the forgiveness and cleansing of wrongdoings (*hamartêmata*).[50] This alone

[48] *On the Sacred Disease* 1.28–30 (6.358–60 L.).
[49] Ibid., 2.2 (6.364 L.).
[50] Ibid., 1.44–6 (6.362–64 L.).

indicates the compatibility of religious beliefs and practices on the one hand, and a medical practice that does not involve any reference to divine forces on the other.

The author of *On Regimen* goes one step further. He is concerned with prognosis and prophylactic regimen in order to preserve health and to avert imminent disease. In this connection, he devotes the whole of the fourth book of his work to the significance of dreams as signs announcing potential illness to the person having the dream. The rationale for this is that the dream has been brought about by the present state of the dreamer's body, a state that is, as it were, an early stage of disease. The dream thus reflects, in a symbolic manner, the state of the body of the dreamer; and the author presents himself as an expert in the correct medical interpretation of these dreams. In response to such prognostic dreams, he recommends prayer alongside dietetic treatment as measures to be taken in order to avert a disease that is imminent.[51] He mentions specific gods such as Apollo, Helios, Hermes, the Heroes, Hecate, and Hestia as the gods one should pray to, both in cases of good signs and in cases of bad signs. Interestingly, the gods are mentioned here as sources of the prevention of disease, not necessarily its cure; and it may be significant in this regard that Asclepius is not mentioned. This indicates that any suggestion that *On Regimen* may somehow be associated with, or be implicitly referring to, the healing cult of Asclepius (in which dreams played a major role) is unlikely to be to the point.[52]

The Asclepius cult developed more or less simultaneously with the rise of professional medicine, and established itself with remarkable success across the Hellenic world, with major centers in Cos, Epidaurus, and indeed Athens. Physically, the cult involved a sanctuary with a temple and a number of adjacent buildings, to which people would travel in order to seek treatment. Temple medicine included incubation, i.e. the patient would sleep in the temple and receive healing, either overnight or in the form of dreams indicating a subsequent course of action to be followed in order to receive treatment. One may suspect (though the relevant evidence is scarce for the classical period) that doctors were also involved in the cult.[53]

The simultaneity between the spread of the Asclepius cult (and other forms of temple medicine) and the developments testified by Hippocratic and other medical and philosophical writings described above, as well as the close geographical proximity (the Asclepius sanctuary on Cos is situated only

[51] *On Regimen* 4.87 (6.642 L.).

[52] See Philip J. van der Eijk, "Divination, Prognosis, Prophylaxis: The Hippocratic Work 'On Dreams' (De victu 4) and its Near Eastern Background," in Horstmanshoff and Stol (eds.), *Magic and Rationality*, pp. 187–218.

[53] For a recent discussion see Bronwyn Wickkiser, *Asklepios, Medicine and the Politics of Healing in Fifth-Century Greece* (Baltimore, MD: Johns Hopkins University Press, 2008); Jürgen W. Riethmüller, *Asklepios: Heiligtumer und Kulte*, 2 vols. (Heidelberg: Archäologie und Geschichte, 2005); Nutton, *Ancient Medicine*, pp. 103–14.

a few miles away from where Hippocrates is believed to have practiced) has sometimes been believed to be indicative of a causal relationship between the two. Already in antiquity, Roman sources (Pliny the Elder, Hyginus) suggest such connections. Yet the relationship, if there was any, is difficult to specify. It seems more prudent to regard them as two parallel developments, and to believe that both had a common origin in what I have referred to above as the quest for health, with the corresponding sense of scope for control and responsibility over matters of health and disease. This sense of taking control, I would argue, translated itself also into a more active search for healing even in situations that would hitherto, in the archaic period, have been deemed hopeless. The cult of Asclepius can be seen as responding to this growing demand. The kind of ailments with which people presented themselves in the cults of Asclepius were often (though not exclusively) chronic diseases that would have been difficult to cure. But the significance of that fact is not so much that Asclepius provided comfort and relief in cases where "secular" or professional medicine was believed to be impotent (for the Hippocratic treatises do claim the capacity to cure, or at least to provide comfort, in a number of severe chronic diseases), but rather that in the classical period people were demanding, expecting, or at any rate hoping to receive healing or relief from conditions that would hitherto have been deemed hopeless and for which no one would have sought recovery.

17

HELLENISTIC AND ROMAN MEDICINE

Vivian Nutton

INTRODUCTION

The centuries that followed the death of Aristotle in 322 BCE transformed Greek medicine. The Eastern conquests of Alexander the Great and, with even greater consequences, the subsequent penetration of Greek culture to Latin-speaking Italy and Rome, introduced the medicine of the Greeks into new areas.[1] The conquests of Roman armies brought this mixture of cultures to the chilly moors of Northumberland and the burning heat of the African desert, to the wave-beaten shores of Galicia as well as to the banks of the Tigris and the Nile. Greek medicine benefited from this expansion. A doctor trained at Alexandria in Egypt settled in Milan, a Syrian doctor treated patients in Spain, while among the Roman forces that invaded Britain in 43 CE were a bilingual medical writer from Italy, Scribonius Largus, and a doctor from Cos, Stertinius Xenophon, who claimed descent from the divine healers Asclepius and Hercules. Elite medicine, however defined, became identified with Greek medicine, written in Greek but also in Latin by 150 BCE and, later still, in some of the languages of the Middle East. Local traditions remained, to resurface in late antiquity, but they had little wider influence, except in introducing new drugs or non-Greek healing cults such as Isis from Egypt, Men from Phrygia, and Coventina from the wilds of Hadrian's Wall in Britain. Writings from Egypt and the Near East, including the New Testament, show the continued belief in demons and magic as a cause of illness.[2]

[1] For Egypt, see Marguerite Hirt Raj, *Médecins et Malades de l'Égypte romaine* (Leiden and Boston, MA: Brill, 2006); Philippa Lang, *Medicine and Society in Ptolemaic Egypt* (Leiden and Boston, MA: Brill, 2013).

[2] Manfred Horstmanshoff and Marten Stol (eds.), *Magic and Rationality in Ancient Near Eastern and Graeco-Roman Medicine* (Leiden: Brill, 2004); Lang, *Medicine and Society*, pp. 46–100; Frank Lebrun and Agnès Degrève (eds.), *Deus medicus* (Turnhout: Brepols, 2013).

But a distinction between Greek and non-Greek medicine does not depend on a rejection or acceptance of a role for the gods in the process of healing. Almost all doctors acknowledged divine intervention in health and illness. They made dedications to the gods and contributed to temple buildings. In big cities like Rome or Ephesus they marched in an annual procession to honour their divine patrons, and Galen at the end of his long exposition of anatomy, *The Usefulness of Parts*, hymned the wonders of the Creator. The divine healers of classical Greece spread further and swifter than their human counterparts. Well before 300 BCE Greek healing cults had made their appearance in central Italy, and a shrine to Asclepius was established in Rome on the Tiber island in 291 BCE with the full approval of the Roman senate.[3] In the Hellenistic and Roman periods local bigwigs and Roman emperors paid for the massive reconstruction of the greatest shrines of Asclepius, Epidaurus, Pergamum, Lebena in Crete, and Aegae in southeast Turkey. Elsewhere, libraries, theatres, and stoas were built alongside temples where sufferers awaited visitations from the gods in dreams.[4] Some, like the orator Aelius Aristeides, recorded their experiences in hymns and orations; others took back to distant homes souvenirs of the deities who had cured them. Traces of the healing god Glycon from Abonouteichos (Turkey) can be found around the Black Sea, and even in Rome. Although Lucian was scathing about this, in his view fraudulent, cult, doctors vied with each other to become the god's mouthpiece at the shrine.[5] Asclepius and "gentle-handed Panacea," as a Greek doctor addressed her at Chester, gained as much from the spread of Greek culture (and from Roman military might) as any secular healer.[6]

Vocal opposition to the introduction of Greek medicine is recorded only in Rome.[7] Archagathus, a surgeon from the Peloponnese, invited by the Roman senate to act as a civic doctor in 219 BCE, left after a brief residence. Stories of his harsh treatments gave him the soubriquet of "executioner," and encouraged politicians like the Elder Cato in their crusade against all things Greek, from medicine and philosophy to three-legged tables and silk dresses. Cato in his own tract on agriculture recommended a domestic agrarian medicine of herbs and charms, accessible to all, and advised his son to shun "filthy Greeks" with their high-flown verbiage. His complaint was echoed

[3] Emma J. Edelstein and Ludwig Edelstein, *Asclepius. Collection and Interpretation of the Fragments (*Baltimore, MD and London: Johns Hopkins University Press: 1998).

[4] Jürgen W. Riethmüller, *Asklepios. Heiligtümer und Kulte* (Heidelberg: Verlag Archäologie und Geschichte, 2015).

[5] Alexia Petsalis-Diomidis, *Truly beyond Wonders. Aelius Aristides and the Cult of Asklepios* (Oxford: Oxford University Press, 2010).

[6] Évelyne Samama, *Les Médecins dans le Monde grec: sources épigraphiques sur la Naissance d'un Corps médical* (Geneva: Droz, 2003), no. 524.

[7] Heinrich Von Staden, "Liminal Perils: Early Roman Reception of Greek Medicine," in F. Jami Ragep and Sally Ragep (eds.), *Tradition, Transmission, Transformation* (Leiden: Brill, 1996), pp. 369–409.

two centuries later by the Elder Pliny in his history of medicine within his *Natural History*, an influential saga of medical greed, sex, and corruption. Both men knew Greek medical writings, and Pliny's information on medicine is derived almost entirely from such sources. But their appeals fell on deaf ears. Most Romans, of all classes, disregarded their warnings. Cicero had no qualms about recommending Greek doctors to his friends, and the humble Sosicrates, son of Sosicrates, expected to find clients when he migrated from Nicaea to Rome around 10 CE.[8]

Nonetheless, medicine in Rome always carried with it a taint of something alien. Many of its practitioners were non-citizens, even slaves or ex-slaves, and few attained any local office of any importance. There was money to be made from rich patients in Rome, especially at court, but it did not bring with it status. Court physicians donated millions to rebuild the walls of their native cities, but wealthy Italians did not send their sons off, like several Greek contemporaries, to spend a year or more studying medicine. At Beneventum (central Italy), the doctor son of a Greek immigrant (himself possibly a doctor) made money, but it was his son, apparently not a doctor, who first became a local magistrate and benefactor.[9]

PHARMACOLOGY

The opening up of the world beyond the Aegean basin to the Greeks brought with it new possibilities for therapy and, in particular, new drugs. A comparison between the drugs mentioned in the Hippocratic Corpus and those of Dioscorides five hundred years later shows an enormous accession of previously unknown plants. Although most still come from the eastern Mediterranean, others are more exotic. Customs documents from Egypt attest the import of rare spices from Africa and India to the greatest entrepot of antiquity, Alexandria. Compound drugs, too, became more common and more complex. The simple herbs of the field or garden, like the humble cabbage, praised by Cato and Pliny, were supplemented by compound drugs sometimes containing over fifty ingredients. A knowledge of local herbs was still desirable, and Galen was not the only writer on simples or on the substitutes for others more distant or seasonally unobtainable, for the drug trade was not always efficient, or trustworthy. But most writers on drugs presume that even exotica are available somewhere in the largest towns. Galen's own recipe collection, obtained by persistence, gifts from friends, and swaps with owners of similar rarities, he believed was the finest in the world, and his period of residence in Rome, with access to the imperial stores

[8] Ralph Jackson, *Doctors and Diseases in the Roman Empire* (London: British Museum, 1988).

[9] Jukka Korpela, *Das Medizinopersonal im antikem Rom* (Helsinki: Finnish Academy of Sciences, 1987); Vivian Nutton, *Ancient Medicine* (2nd edn; London: Routledge, 2013), pp. 164–81.

and imports, especially from Crete, allowed him opportunities denied to others more humble. They had to rely on whatever was to hand, unless they profited from the largesse of a Galen, who generously gave them drugs (as well as medical instruments) from his own collection.[10]

Many of these new herbs were first described by Hellenistic writers, only fragments of whose works survive, with the exception of some didactic poems by Nicander of Colophon. They ranged from tracts on antipathy and sympathy, and quasi-magical remedies by those such as Bolus of Mende and Ostanes the Persian magus, through Pamphilus of Alexandria's lists of plants and animals, to learned writers on pharmacology such as the Empiricist Serapion of Alexandria or Sextius Niger.[11] The richness of the tradition is visible in Pliny's *Natural History* and in the *Materia medica* of his Greek contemporary, Pedanius Dioscorides of Anazarbus (southeast Turkey). Dioscorides was trained at nearby Tarsus, the home of a distinguished series of pharmacologists for almost a century. His book, which became the bible of botanists and pharmacologists down to the seventeenth century, aimed to describe the properties of all known herbs, and a few animal and mineral drugs, but it records few plants from Western Europe. The descriptions are clear and succinct, but they were not originally accompanied by the illustrations that adorn manuscripts of the work from the sixth century onwards. Dioscorides organized his book so as to allow his readers to locate herbs with similar medicinal properties close together, an arrangement that many found unsatisfactory, for at least one recension of his treatise replaces them in alphabetical order.[12]

Finding a recipe within a large book roll, or series of rolls, was not easy, and Pliny and Scribonius Largus provided indexes as a way to locate relevant material. Largus dedicated, ca. 48 CE, his collection of recipes to Callistus, a wealthy freedman of the emperor Claudius.[13] It contains recipes associated with members of the imperial family as well as some more popular remedies, often with ingredients that would be effective today such as opiates. His treatise was a wake-up call to fellow doctors who rejected drugs in favor of diets or whose complex and time-consuming prescriptions let in less expert and more dangerous practitioners who promised a swift and easy cure.[14]

[10] Galen, *Avoiding Distress* 31–6; Sabine Vogt, "Drugs and Pharmacology," in R. James Hankinson (ed.), *The Cambridge Companion to Galen* (Cambridge: Cambridge University Press, 2008), pp. 304–22.

[11] Alain Touwaide, "Pharmacology," in H. Cancik (ed.), *Brill's New Pauly* (Leiden: Brill, 2007), vol. 10, 927–32; John Scarborough, *Pharmacy and Drug Lore in Antiquity: Greece, Rome, Byzantium* (Farnham and Burlington, VT: Ashgate, 2010).

[12] John M. Riddle, *Dioscorides on Pharmacy and Medicine* (Austin, TX: University of Texas Press, 1985); Lily Y. Beck, *Dioscorides of Anazarbus, De materia medica* (Hildesheim and New York: Olms-Weidmann, 2005).

[13] Sergio Sconocchia, "L'opera di Scribonio Largo e la letteratura medica latina del I. sec. d. C.," in Hildegard Temporini and Wolfgang Haase (eds.), *Aufstieg und Niedergang der römischen Reich, Teil II, Band 37.1* (Hildesheim: Olms, 1993), pp. 845–922; Nutton, *Ancient Medicine*, pp. 175–8.

[14] Christoph R. Machold, *Scribonius Largus und die antike Iatromagie* (Hildesheim and New York: Olms, 2010).

In his preface Largus set his complaints into an unexpected context, a Latin reformulation of the Hippocratic *Oath*.[15] Like a soldier taking the military oath, he should make his *professio*, a word that originally meant a public pronouncement, but which here comes close to the modern use of "profession." By acknowledging the *Oath*, the true doctor accepts certain ethical or moral responsibilities which in turn determine how he should practice. Some remedies, such as gladiator's blood, fall outside the "profession of a doctor," even though they were often recommended, notably for chronic illness such as epilepsy, in works by respectable authors. Largus also used stories and language familiar to a Roman audience to strengthen his argument; a doctor should not kill an enemy through his medical knowledge, a reference to a famous incident in Rome's wars with King Pyrrhus of Epirus, although if this were done in battle, the doctor would be excused if he were acting as a soldier. Choosing to become a doctor and to follow the Hippocratic *Oath* thus guarantees certain moral standards. Largus' emphasis on the *Oath* is unusual, and not just because others, like Cato and Pliny, viewed it as evidence of a conspiracy on the part of doctors to deceive their patients. The fragments of a commentary ascribed to Galen on the *Oath* say nothing about its ethical prescriptions, and in his own formulations of Hippocratic ethics Galen relied instead upon stories of his great predecessor. Largus also antedates by more than half a century the earliest Greek evidence for the employment of the *Oath* in medical education, on a papyrus from Oxyrhynchus in Egypt, where it is recommended to be sworn by those beginning a medical education.

Whether Largus would have accepted as a fellow professional the anonymous author of a neglected tract on the properties of the centaury is a moot point. This man, a Greek living in the 170s, was educated in medicine in Rome with a "very respected doctor." From him he learned of the wonderful properties of this plant to cure almost anything from headaches and stomach complaints to wounds and even rabies. His short tract, written to inform his brother back home of the truth about this panacea, is a rare survivor of a type of literature that may have been widespread, an account of the uses of one single substance.[16]

The simplicity of this anonymous author's exposition contrasts with the medical poetry, written in Greek, by some of the leading doctors of the Roman imperial period. None of the works of Heraclitus of Rhodiapolis survive today to enable us to judge whether he deserved the title of the "Homer of medical poetry" and the honours from the major intellectual centers of the Greek world recorded on his tombstone. But the poems of

[15] Scribonius Largus, *Recipes*, Pref.; Jeffrey S. Hamilton, "Scribonius Largus on the Medical Profession," *Bulletin of the History of Medicine* 60 (1986), 209–16; Margherita Cassia, *Andromaco di Creta: medicina e potere nella Roma neroniana* (Rome: Bonano Editore, 2012).

[16] Vivian Nutton, "De virtutibus centaureae: A Pseudo-Galenic Text on Pharmacology," *Galenos* 9 (2015), 153–79.

Servilius Damocrates and the elder Andromachus, both contemporary with Largus and Dioscorides, display a considerable skill in putting into verse the many ingredients that made up their prescriptions for theriac, both an antidote against poisoning and a tonic. Their poetry not only displayed their high level of culture in a society that valued culture highly, but also, some believed, provided a check against the corruption of numbers and names as one recipe was copied by hand from exemplar to exemplar. Words, especially into meter, were harder to confuse than the letters used in both Greek and Latin to indicate numbers.[17]

MEDICINE AND THE CITY

The greater accessibility of medicinal substances was a direct consequence of the expansion of the Greek-speaking world in the Hellenistic period. Another was the growth of mega-cities. Although most people still lived in communities with fewer than two thousand inhabitants, others had migrated to cities with populations in tens, if not hundreds, of thousands. Carthage and Rome in the west, Athens, Antioch, Ephesus, Pergamum, and Alexandria in the East spread over several square kilometers; Rome soared skywards in apartment blocks several storeys high. Smaller cities, like Aspendus (southern Turkey), Gadara (Jordan), or Pompeii still astonish today's visitor by their size and grandeur. These cities, Rome and Alexandria above all, signaled a break with the agrarian past. A face-to-face society civilization in which everybody knew everybody was replaced by the anonymous big city, where, so Galen grumbled, a murderous doctor could continue to practice simply because he could always find patients who did not know him. These crowded cities brought with them new problems. They had to be provisioned, often from sources hundreds of miles away, for famine threatened the very stability of the state. The annual arrival of corn ships from Egypt or north Africa was a cause for rejoicing among the urban plebs of Rome, dependent on corn doles from the emperor. Many cities also invested in long aqueducts, bringing water for drinking and for cooling fountains, whose size and beauty displayed the wealth of the community. Galen was particularly proud of Pergamum's magnificent aqueducts, which brought water to the very summit of the acropolis in a remarkable feat of engineering. But, however much architectural and medical theorists dilated upon the need for broad, sunny streets and clean, wholesome air for all, in reality such amenities were largely confined to the upper classes, who migrated in summer away from the stifling heat and noisome smell to the

[17] Samama, *Les Médecins*, nos. 289–90; Sabine Vogt, *Servilius Damokrates. Iambische Pharmaka im Corpus Galenicum. Einleitung, Edition und kommentierte Übersetzung* (forthcoming); Cassia, *Andromaco di Creta*.

seaside villas of Campania or the nearby hills. In such a crowded environment diseases must have proliferated, and death in the city came far more swiftly than in the countryside.[18]

ALEXANDRIA

Of all the new cities it was Alexandria in Egypt that was to have the most profound effect on medicine.[19] An active tradition of medical education and innovation can be traced there for more than a millennium after the city's foundation by Alexander the Great in 331 BCE, and it richly deserved the appellation given to it by an ancient geographer of "the foundation of health for all men." Situated in the Nile delta, it became a great trading city, where drugs from Africa, Asia, and the Eastern Mediterranean arrived by land and sea. After Alexander's death in 323 BCE, one of his generals, Ptolemy I, made it his capital. Protected from attack by land, and controlling the fertile lands watered by the Nile, Alexandria proved a powerful base from which successive rulers sought to extend their control over Greece and the Near East. But Alexander's city was both a cultural and a political statement, a Greek city in an ancient non-Greek land, and within a short time Alexandria surpassed Athens as the cultural center of the Greek world.

Not that Athens was a scientific backwater. The philosophical groupings at Athens, especially the followers of Aristotle at the Lyceum, continued his broad interests in natural phenomena. Theophrastus in particular not only produced books on minerals and on plants but also touched on matters medical, such as sweating and giddiness. The shadowy Euenor, who lived at Athens in the 320s and is credited with giving a new name to the horns of the uterus, is hard to place, but dissection of living animals in the tradition of Aristotle and Diocles of Carystus may have continued there into the third century. The model of royal support for scientists, and indeed of royal involvement, was transferred to the courts of Alexander's successors – to Antioch, where Cleombrotus, the father of Erasistratus, was a royal physician (as his son may also have been); Pergamum, where Attalus II carried out experiments on herbs; and, later, Pontus, where Mithradates VI was famous for testing poisons and antidotes. An extremely unreliable Jewish tradition, but circulating before 200 CE, also attributed to Cleopatra of Egypt murderous investigations of pregnant slaves to show the growth of the foetus. More widespread, although less dangerous, was her reputation as

[18] Valerie M. Hope and Eireann Marshall (eds.), *Death and Disease in the Ancient City* (London: Routledge, 2000); Paul Erdkamp (ed.), *The Cambridge Companion to Ancient Rome* (Cambridge: Cambridge University Press, 2011).

[19] Peter M. Fraser, *Ptolemaic Alexandria* (Oxford: Clarendon Press, 1972).

a writer on cosmetics and alchemy, although none of the writings under her name will have been written by her.[20]

The legends around Cleopatra also emphasize the importance of Alexandria as a cultural capital. Ptolemy I founded two institutions that were to have a long-lasting influence. The Library housed copies of all of Greek literature, including science, obtained by means both regular and irregular: boats arriving at the port were searched for rare books, which were then confiscated for copying. A distinguished series of scholars and litterateurs presided over the production of detailed catalogs of all subjects, thereby setting a precedent for the founding of major libraries elsewhere, at Pergamum or, later, on the Palatine hill in Rome. Cities like Smyrna or Ephesus, and even provincial towns like Nysa (west Turkey), had their own libraries, to which wealthy citizens, such as Galen's friends, might donate copies of new books on medicine and science. The emperor Tiberius ordered a pain-killing recipe to be publicly displayed in such libraries, and Hadrian acted similarly with a work by Marcellus of Side, some of whose medical poetry still survives. But it was Ptolemy's library at Alexandria that was of greatest significance. The medical writings ascribed to Hippocrates may have been first assembled together here, although what they were and how they relate to the modern Hippocratic Corpus are contentious issues. Galen was heir to four centuries or more of the Alexandrian tradition of Hippocratic emendation and interpretation, even if it is far from clear that during his stay in Alexandria in the 150s he ever set foot in the Library.[21]

Equally obscure is the involvement of medical men in Ptolemy's second major institution, the Museum or Hall of the Muses. It provided a meeting place for intellectuals of all kinds, poets, mathematicians, musicians, geographers, and doctors – in the Roman period it also hosted local politicians and even champion athletes among those entitled to "dine free at the Museum" – and this scholarly community undoubtedly contributed to the intellectual ferment of early Alexandria. Its success was imitated elsewhere, particularly among the wealthy cities of Asia Minor in the Roman period. But evidence for any teaching carried on there is problematic. The doctor-historian Andreas wrote a book on the "House" of Herophilus, implying that he and his pupils studied outside the Museum in private houses, and although it is tempting to associate the introduction of human dissection with the Museum, no ancient source makes that link.[22]

Nonetheless, royal involvement would have been necessary from the start of what was a momentous development in medicine, the first clear and

[20] Geoffrey E. R. Lloyd, *Methods and Problems in Greek Science. Selected Papers* (Cambridge: Cambridge University Press, 1991), pp. 352–71; Mirko D. Grmek, *Il calderone di Medea* (Bari: Laterza, 1996).

[21] George W. Houston, *Inside Roman Libraries: Book Collections and Their Management in Antiquity* (Chapel Hill, NC: University of North Carolina Press, 2014).

[22] Fraser, *Ptolemaic Alexandria*, pp. 305–36.

systematic attempts to reveal, describe, and investigate the internal organization and working of the human body. This is traditionally associated with one scholar, Herophilus of Chalcedon, who certainly worked in Alexandria, and with another, Erasistratus of Ceos, who probably did so. A third anatomist, Eudemus, may also have been in Alexandria, and teaching of human anatomy continued at least down to late antiquity, even if this only rarely involved cutting into a (animal) body.

In Classical Greece, there was no systematic attempt to investigate the internal structures of the body, let alone of the human body. Writers from the fifth and fourth century BCE who describe them rely almost entirely on observations from animals or from dealing with wounds and fractures. A reasonable understanding of surface anatomy is accompanied by an extremely deficient one of internal anatomy. Human dissection was also hampered by religious prohibitions on touching (let alone cutting) a corpse, prohibitions that continued to be repeated long after the first dissections of humans.[23]

Although the chronological relationship between the two men is uncertain, historians are agreed that Aristotle and Diocles began systematic dissection of animals around 350 BCE. Aristotle's dissections were, as far as can be judged, far more extensive and intellectually challenging than those of Diocles. He was concerned with living things, animals, both in their structures and in their workings, and to this end he cut up a wide variety of animals, birds, and fish. He was concerned not just to describe what he, and his collaborators, saw, but also to seek out the extent to which some purpose could be observed in their creation. Unlike Plato, who had placed his three parts of the soul in three different parts of the body, brain, heart, and belly, Aristotle gave the primacy to the heart, the seat of his unitary soul. He concluded that the heart was the dominant organ, communicating with the rest of the body by a series of channels, and that the brain, which at first sight appears grey and almost bloodless, functioned to cool down the heat of the heart and the body. His animal dissections, like those of Diocles and a slightly younger contemporary, Praxagoras of Cos, went a long way towards distinguishing veins and arteries, both of which had been covered by the same name of *phlebs*, "vessel." Praxagoras is also credited with the discovery of nerves (or, to be more precise, distinguishing a nerve from the other cords and ligaments that had also been called *neuron*), but in a somewhat Pickwickian sense, for he gave the name to what he believed were the terminal channels of the arterial system that distributed sensation and motion to the individual parts of the body.[24]

[23] Geoffrey E. R. Lloyd, *Magic, Reason and Experience. Studies in the Origin and Development of Greek Science* (Cambridge: Cambridge University Press, 1979), pp. 146–168; Lloyd, *Methods and Problems*, pp. 164–93.

[24] C. Reginald S. Harris, *The Heart and the Vascular System in Ancient Greek Medicine* (Oxford: Clarendon Press, 1973); James N. Longrigg, *Greek Rational Medicine* (London: Routledge, 1993), pp. 149–76.

These discoveries helped to establish dissection as a valuable way of investigating the internal organs, as well as stimulating a wider interest in science. Both Aristotle and Praxagoras had pupils to whom they transmitted their ideas and techniques, and some of their discoveries were also known to scholars in other disciplines. But the confused political situation after Alexander's death was not conducive to research, and religious and social taboos prevented any immediate progress along the lines sketched out by Aristotle.

The success of Ptolemy in establishing control over Egypt, and in continuing the intellectual patronage of Alexander, opened up a new and unique opportunity. Unlike the Greeks, the Egyptians had had a long tradition of interfering with a corpse – mummification. This involved removing some of the internal organs and placing them in so-called Canopic jars, a religious ritual that accompanied what was, to a Greek, mutilation. At the very least, here was a society that ostensibly did not retain the same reverence for a corpse as did Athenians, and social differences between the all-conquering Greeks and the Egyptians, and, still more, between the bourgeois Alexandrians and the Egyptian peasants of the Nile valley, would have confirmed the view that the Egyptians cared little or nothing for the bodies of their families. What knowledge the Greeks obtained from any contact with the mummifiers is doubtful, given the Greeks' notorious reluctance to learn other languages and the degraded state of the organs themselves, particularly the brain, when they were removed from the body. Neither the evidence of the mummies themselves nor any surviving Egyptian medical text shows any sign of the systematic exploration carried out by the Greeks, although Egyptian trauma surgery was certainly as good as that of contemporary Greeks. If the next step towards human dissection was to be taken, Egypt offered far more likely prospects than Greece itself, and, at Alexandria, protected by the ruler and even by the Muses, Greek intellectual doctors could take the momentous step of cutting up a human body.[25]

ALEXANDRIAN ANATOMY

According to the detailed exposition of the Latin author Cornelius Celsus, Herophilus and Erasistratus carried out their researches on the living bodies of condemned criminals. The truth of this allegation is hard to substantiate, not least because Galen, who saw himself as an inheritor of the Alexandrian tradition, says nothing on this point. That criminals were involved is almost certain, for Celsus, reporting Hellenistic debates about the value of dissection, repeats the (minority) argument that it is not cruel to execute criminals

[25] Longrigg, *Greek Rational Medicine*, pp. 177–219; Lang, *Medicine and Society*, pp. 249–58; Nutton, *Ancient Medicine*, pp. 130–41.

in this way for the future benefit of the innocent – an argument used to justify deadly experiments even in the twentieth century. Herophilus and Erasistratus certainly dissected both human corpses and living animals, but the results of human vivisection cannot easily be distinguished from those of a combination of human dissection and animal vivisection. But two things are clear. Whether the subjects were alive or dead, dissection required official permission and protection; and opposition to the cutting up of human corpses contributed to its abandonment within a generation. Some objected to its cruelty, others to its irrelevance to the treatment of the living, for a dead body, ipso facto, could not be expected to respond in the same way as a living one. Even dissection, though not cruel, is messy, and although Celsus recommended dissection, he restricted it to beginners seeking to understand the positions and relations of bodily parts. No wonder, then, that Rufus and Galen looked back to a vanished age under the kings when dissection was possible. Even skeletal anatomy was confined to Alexandria by 100 CE, but animal dissection was reintroduced there soon after by Marinus, and fifty years later Galen described public anatomical displays (and private lessons) being put on in Rome by doctors trained at Alexandria.[26]

No writings by Hellenistic physicians connected with Alexandria survive in full, except possibly for the "meditation" on the Hippocratic *On Joints* by Apollonius of Citium around 90 BCE. This exposition, less formal than later commentaries, reveals a sound understanding of the anatomical structure of the joints, not least when Apollonius attempted to resolve the vexed question of how best to reduce permanently a dislocated hip joint, whose damaged ligaments made a repetition of the injury very likely. For the rest we rely on fragments mediated through others, principally Celsus and Galen, sometimes using the author's own words but more often summarizing or inserting them into a context that may not correspond to that in the original. At best we have dim echoes of contemporary debates.

Herophilus, a pupil of Praxagoras, based his physiology and pathology on a theory of humors and responded critically to some of the writings in our Hippocratic Corpus. He took an interest in the meaning of words, which he put to good use when naming previously unseen or undescribed parts of the body. Many of his striking coinages are still in use today: the *calamus scriptorius* for a pen-shaped groove in the brain; the *torcular* ("wine-press") for the junction of the veins of the head; the *styloid*, pen-shaped, process of the skull; and the *duodenum* ("twelve-fingers long"). His anatomical researches covered the whole of the body, from a careful examination of the coats of the eye and the first detailed account of the liver, and possibly the pancreas, to the male and female reproductive systems, where his description of the womb is based on animals. He continued Praxagoras' interest in the vascular system, distinguishing between veins and arteries by the greater

[26] Celsus, *On Medicine*, Pref.

thickness of the arterial coats, and noting the importance of the hepatic portal system. But he rejected Praxagoras' view that the arteries contained only *pneuma*, spirit, the vehicle of sensation, and he went further than his master in distinguishing the nervous from the vascular system. Centuries later, Galen was impressed with his understanding of the brain and with his distinction between sensory and motor nerves.[27]

But anatomy was only part of his wide-ranging activity as a doctor.[28] His dissection of the womb contributed data for a book on midwifery in which he discussed multiple births and uterine prolapse. His understanding of the pathways of some nerves had an obvious relevance to his treatment of paralysis, and his work on the arteries provided him with an anatomical explanation and justification for pulsation, which he considered crucial to diagnosis. He constructed a portable water-clock by which to measure the pulse, delineating different types of pulse by their speed, size, vehemence, or rhythm and describing them in words that have many parallels with Hellenistic discussions of music theory.

Galen, our major source of information, viewed Herophilus positively, partly because many of his ideas corresponded to what Galen believed the historical Hippocrates had taught. Herophilus sought to investigate the causes of illness, although he admitted that this might not always be possible, but to Galen his refusal to discuss therapy in terms of the four primary qualities, hot, cold, wet, and dry, was an unnecessary restriction on an otherwise sound methodology. He praised various ways of keeping fit, "for without health one's skill remains invisible, one's strength unused, one's wealth worthless, and one's speech powerless," and a statue in his memory displayed gymnastic equipment. He favored bloodletting and treatment by opposites, especially drugs. But not everything he did would find favor today. His most notorious remedy, and his most dangerous, was hellebore, which he compared to a general in its ability to stir up all the inward parts of the body. Merely to survive such treatment, long forbidden in Western human medicine, demanded a strong constitution. Pharmacology, along with the pulse, remained a major object of study among his followers, the so-called Herophileans, who can be traced in Asia Minor for several centuries. But, like dissection, it was only one part of the wide-ranging, innovative and thoughtful activity of Herophilus himself.

Far more controversial was Erasistratus of Ceos.[29] There is no reason to believe that he was affected by a law of his native island that ordered all those over the age of sixty to be put to death, or that Galen was right to claim that

[27] Heinrich Von Staden, *Herophilus. The Art of Medicine in Early Alexandria* (Cambridge: Cambridge University Press, 1989); Fraser, *Ptolemaic Alexandria*, pp. 348–58; Harris, *The Heart*, pp. 177–94; Longrigg, *Greek Rational Medicine*, pp. 189–204.

[28] Von Staden, *Herophilus*, provides the best overview and translates all the fragments.

[29] Ivan Garofalo, *Erasistrati Fragmenta* (Pisa: Giardini, 1988), lists the fragments; Nutton, *Ancient Medicine*, pp. 135–9; Harris, *The Heart*, pp. 195–233; Longrigg, *Greek Rational Medicine*, pp. 205–18.

he performed accurate dissection only at the end of his life, when he could devote more time to it. Although his father and one of his teachers, Chrysippus, were linked with the Seleucid court at Antioch, this need not mean that Erasistratus' practiced there, for Antioch is never mentioned as a site of dissection and his appearance in a famous story about Queen Stratonice is an obvious late addition to the saga. Similarly, reports that he was a student of Theophrastus at Athens or of another Peripatetic, Metrodorus, cannot be substantiated, and may have been introduced to explain his fascination with ideas and practices drawn from other philosophers and scientists. His restless imagination, his obvious willingness to consider new hypotheses and experiments, and his vigorous and combative style must have made him difficult to pin down, in real life as much as in later history, and, despite our abundant information, he remains an enigmatic figure.

Galen distinguished between Erasistratus' activity as an anatomist and his other work. The former he praised highly, the latter he regarded with suspicion – and the interpretations of later Erasistrateans with obvious scorn. But this says more about Galen's priorities than about those of Erasistratus, even if Erasistratus' comparison, fragment 247, of the neophyte investigator to an untrained athlete still resonates today:

> As soon as his mind begins to work, he becomes confused and bewildered, ready to withdraw from his investigation in a state of mental fatigue and exhaustion, like untrained runners in a race. But with constant practice, pursuing research, not just in an hour or so, but unceasingly throughout his whole life, he comes to penetrate whatever topic he chooses, reaching his goal by persistent enquiry into everything that might be relevant.

If this credo sounds remarkably modern, it is because Erasistratus employed techniques and ideas that would not be out of place in a modern laboratory. Working with the naked eye, he pursued a series of anatomical investigations that went beyond those of Herophilus. He investigated the workings of the heart valves, concluding that they were there to prevent any reflux as the heart expanded and contracted "like a smith's bellows." He may not have been the first to dissect a heart systematically, for the anonymous author of the tract *On the Heart* seems to have described at least the semi-lunar valves among what he terms the "membranes" of the heart, "a piece of craftsmanship most worthy of description." But his talk of "stays," "cobwebs," and "fibres" in what was almost certainly an animal heart (for he recommends a different experiment on a pig) remains at the level of description. That this anonymous author was writing in the Hellenistic era is clear from the language he uses, but he cannot be located in any obvious center, and his apparent ignorance of what Herophilus and Erasistratus had found is no clear proof – given the fragile transmission of information in the ancient world – that he must have been writing before, or at the same time as, them.[30]

Erasistratus' dissections required an expert eye. He traced the arteries and veins into tiny channels that he believed finally linked together but with so tiny an aperture that blood could not pass from one to the other under normal circumstances. His repeated examination of the brain compelled him to change his opinion of the origin of the nerves from the *dura mater* to within the ventricles of the brain itself. His skill with the knife was matched by his willingness to experiment. He inserted a cannula into an artery to discover whether pulsation was a property of the tunics of the artery or the mechanical result of *pneuma* being driven into the arteries by the beating heart, an experiment that produced contradictory results among those, from Galen onwards, who tried to repeat it. On another occasion, he showed that animals emitted imperceptible effluvia by keeping a bird in a vessel without food for some time, weighing it with its visible excreta, and then comparing it with its original weight. The crudity of the experiment does not detract from Erasistratus' belief in the heuristic value of experimentation.

But although several later writers placed Erasistratus' work in a canon of great physicians that went back to Hippocrates, Galen had little time for much of his non-anatomical work. His arguments are at times devastatingly unfair, sometimes deriving from the conflicting views of later followers, and at others interpreting ideas out of context. His disagreement arises from the fact that often Erasistratus goes against what in Galen's eyes were central elements in the medicine of Hippocrates. Erasistratus wrote no commentary on Hippocrates and his explanatory universe is more akin to that of Aristotle than to Plato's. He saw humors simply as fluids, causing harm through blockage or overflow, not as the essential qualitative building blocks of the body posited by the author of *The Nature of Man*. He also rejected teleology as an explanation for the organization of the body: the biceps muscles were not created strong in order to lift things; they became strong by lifting.

Particularly striking is Erasistratus' conception of the body as a living, functioning organism, explicable in terms of mechanics and physics. The kidneys, liver, and bladder act as filters. The stomach grinds and crushes ingested food like a corn mill, producing sensations of hunger if grinding continues after the food has been finely milled. Damage to, or failure in, an organ led to physical, and at times mental, illness. Dropsy, for example, was the result of cirrhosis, hardening, of the liver, which allowed only the watery part of the blood to pass into the veins. Growth and nutrition were mechanical processes, whereby the essential nutriment and *pneuma* were

[30] Iain M. Lonie, "The Paradoxical Text *On the Heart*," *Medical History* 17 (1973), 1–15, 136–53; Elizabeth M. Craik, *The "Hippocratic" Corpus. Content and Context* (London: Routledge, 2015), pp. 52–6.

distributed around the body by means of a three-fold rope of nerves, veins, and arteries, which gained strength and flexibility through their interweaving. Bone, flesh, and even the brain substance were produced as *parenchyma* from this nutriment, which was distributed by a sort of vacuum process. Since, according to Erasistratus, a continuous void was impossible, more nutriment must have been drawn in to fill the void left when other nutriment had been taken up.

This notion of a vacuum, whether derived from contemporary physicists like Strato of Lampsacus or from observing the machines made by contemporary engineers, plays a crucial role in Erasistratus' ideas on the vascular system. He distinguished arteries and veins by both their structures and their contents. In the normal body, blood was carried only in the veins, while the arteries carried only *pneuma*, a refined type of air or spirit. Blood moved into the arteries only in exceptional circumstances or when the arteries were cut. Just as blood or *pneuma* moved from one chamber of the heart to another as the valves opened and closed to allow it to pass outwards and new material to come in to replace it, so when an artery was cut, *pneuma* escaped and blood was drawn in through invisible connections with the veins to fill the vacuum. Likewise a build-up of excessive blood in the veins, "stirred into motion like the sea moved by a gale," seeped through into the arteries to cause inflammation or other kinds of disease. This was a mechanical process, not the result of qualitative changes within the body.

Galen saw this as a perversion of sound anatomical research for dubious clinical ends, and he viciously attacked weak points in Erasistratus' logic and therapy. Given Erasistratus' emphasis on the dangers of excess blood, he failed to see why he treated such conditions by diet, drugs, and the use of tourniquets to move the blood away from the affected area or prevent more blood accumulating rather than the swifter and more effective method of venesection. Not that Erasistratus entirely rejected venesection, as later followers believed, but he did not recommend it when less drastic therapies were to hand. The doctor was more likely to succeed by gradual intervention, to which the patient would become accustomed, than by vigorous and dangerous remedies. Nonetheless, when necessary, he was prepared to take radical action, promoting a massive, and possibly fatal, evacuation of fluid in order to cure a dropsical patient, and, on one occasion, "daringly" (Galen's comment) removing the covering flesh and membranes to apply drugs directly to the liver.[31]

Erasistratus' writings and influence survived for centuries. His views on diseases long continued to be reported in medical handbooks, and Galen had his own copies of some of his works wherewith to refute the ideas on bloodletting of contemporary Erasistrateans in Antonine Rome. But he was

[31] A major survey of Erasistratus is expected from Heinrich Von Staden. See also the studies listed in footnote 29 above.

a controversial figure even in his own lifetime through the very fecundity of his ideas. One can trace links with much earlier writers on medicine and science, but their opinions are developed by Erasistratus in unexpected ways. Unlike Herophilus or Eudemus, Erasistratus, an able dissector who investigated the bones, pancreas, nerves, and the channels in the umbilical cord, went beyond mere description to consider the body as a working organism. To call him the first physiologist is a pardonable anachronism.

SECTARIAN MEDICINE

Celsus, expounding Hellenistic medicine to a Latin audience, explained that in therapeutics there had long been a primary difference of opinion between those who held that all that was required was derived from experience by itself, the so-called Empiricists, and those who argued that this additionally required a reasoned knowledge of human bodies, and specifically of causes of disease, in order to determine a successful therapy. In this debate, the role of anatomical knowledge was crucial, even if its proponents rejected human vivisection in favor of a combination of knowledge gained from the anatomy of a corpse and the treatment of the wounded. But most of Celsus' discussion is taken up with wider questions of epistemology, and in particular with the relevance to medical practice of an understanding of causation, beyond the merely evident (a term ambiguous in itself). This argument, which Celsus traced back to Hippocrates, developed in the Hellenistic period in two ways. It became more sophisticated, drawing on ideas in logic and philosophy as well as medicine, and it was conducted along sectarian lines, pitting the so-called Rationalists or Dogmatics against the Empiricists and, later, the Methodists. This tripartite scheme was widely adopted as a classification of an apparently multifarious, even chaotic, medical profession, and, from late antiquity down to the twentieth century, it became the standard means to delineate post-Hippocratic medicine.[32]

The Greek word used to describe these three groupings, *haeresis*, literally means "choice," and is often found in religious writings to indicate an alternative view, and often a wrong one, hence "heresy." (Its Latin equivalent, *secta*, "sect," stresses following a leader or the doctrines of the group, and is often equally pejorative.) Its earliest certain appearance in medicine appears in the late third century BCE, when Serapion of Alexandria wrote a treatise entitled *Against the Haireseis.* The plural is significant, for Serapion will have been defending his preference for Empiricism over a variety of choices made by others. He was followed over the next three centuries by

[32] Philippe Mudry and Jackie Pigeaud (eds.), *Les Écoles médicales à Rome* (Geneva: Droz, 1991); Guy Sabbah and Philippe Mudry (eds.), *La médecine de Celse. Aspects historiques, scientifiques et littéraires* (St Etienne: Université de St Étienne, 1994), esp. pp. 63–122.

many authors expounding under similar titles the doctrines of their own sect. Some of them were extensive. *On the Haeresis of Herophilus* by Apollonius Mys ran to twenty-nine books, that of his fellow Herophilean Aristoxenus to at least thirteen, while Galen's summary of Heraclides of Tarentum's *On the Empiricist Haeresis* alone took up eight books. These treatises, containing apologetic as well as polemic material, became depersonalized; individuals are mentioned but, like the authors themselves, as part of a larger group. This had advantages and disadvantages. It allowed an author to show that his intellectual stance was no aberration but shared with others, but, as opponents enjoyed pointing out, there were individual differences within each group that indicated dissension. Maintaining doctrinal coherence in the general absence of formal institutions was difficult, but some groupings, and particularly the Empiricists and Methodists, developed doctrines and practices distinctive enough to warrant the title of "sect."

The Empiricists owed their foundation to a dissident pupil of Herophilus, Philinus of Cos, although they later claimed to trace it back beyond Hippocrates to his older contemporary, Acron of Acragas. Its adherents can be found all over the Greek-speaking world from Antioch, Alexandria, and Cyrene to Tarentum and Naples, and the last known member of the sect, the philosopher Sextus Empiricus, lived around 200 CE. They were not empirics in the common meaning of the word, but derived their name from their insistence on the primacy of experience, *empeiria*, over any search for the causes of disease. That was doubly unhelpful. Knowing a cause, even if it was evident, did not always provide a guide to treatment, and effective treatment could often be achieved more swiftly or effectively without bothering about causes, especially if a proper Galenic diagnosis might take days. Nor did knowing the cause necessarily mean knowing the therapy immediately. A sudden inspiration, a desire for a drink, or chance, the application of a leaf of a nearby plant to an insect bite, might provide a cure just as much as ratiocination. Hence the Empiricists' preference for the data of experience. Like the cobbler or the carpenter, for the doctor, practice made perfect.[33]

This experience was not just that of the individual doctor. Case histories, the codified record of previous success, were important in providing a wider data bank for future use, particularly as concerned drugs. Their opponents' criticisms that such a reliance on the past left the Empiricists unable to adapt to a new condition or new circumstances except by trial and error were countered by the principle of similarity. Something that had worked in similar circumstances provided a good starting point for intervention, and swift action, taken cautiously, might prove beneficial, even if neither

[33] Michael Frede, *Essays in Ancient Philosophy* (Oxford: Clarendon Press, 1987). Fundamental is Karl Deichgräber, *Die griechische Empirikerschule* (2nd edn; Berlin and Zurich: Weidmann, 1965).

attained the certainty that acquaintance with an identical case might give. The weaker the similarity, the less likely the cure, and hence the need to rely on the so-called "tripod": accurate observation; a well-stocked library or, more often, a collective memory of what had worked in the past; and an understanding of the virtues and limitations of similarity.

The Empiricists, through their insistence on careful, independent observation, and a detailed recording of symptoms, not to speak of their reverence for Hippocrates, the empirical observer par excellence, were among the most learned physicians of their day. Even Galen grudgingly allowed that an experienced and thinking doctor might be far more valuable than one who knew the proper theory but lacked the insight to apply it properly. He held Heraclides of Tarentum, who lived around 90 BCE, in high regard as a writer on therapy, drugs, and surgery, and he praised his Hippocratic commentaries for combining medical awareness with a fine sense of Greek style. Heraclides himself praised Hippocrates for his wide experience and passion for truth, a far cry from opponents such as Andreas of Carystus, whom he accused of copying down information about drugs like a town-crier announcing the escape of a slave but unable to recognize the slave himself even if he were standing before him. Simple collection of information by itself was not enough, as a variety of Empiricist authors insisted.[34]

By comparison with our fragmentary sources on the Empiricists, several texts survive from authors belonging to another of Celsus' groupings, the Methodists. They cover a period of more than five centuries. Some come from Rome or north Africa, others from the big cities of Asia Minor. Its adherents were arguably the most important and the most numerous in their day, often serving as doctors to the imperial house in Rome and to municipal elites. This was a Roman grouping, created in Rome, the capital of the Empire, and capable of expressing its doctrines as much in Latin as in Greek. As such, it must be considered against the wider background of the hellenization of Rome and Italy in the last three centuries BCE.[35]

According to Celsus, the Methodist sect was founded by Themison of Laodicea, a Greek in Rome who was, in some way, the pupil of the controversial Asclepiades of Bithynia. Asclepiades, also an immigrant, had gained a great reputation in Rome by the 90s BCE, although whether he had arrived there recently or many years before is disputed. Asclepiades rejected a qualitative humoral pathology in favor of one that explained disease in terms of the blockage of pores or channels by corpuscles or the movement of the corpuscles to places where they ought not to be. His body was constituted out of fragile "unlinked" corpuscles, a theory that has parallels in Epicurean philosophy, or even that of Democritus before him, but which may also be derived from Erasistratus' ideas on the body as a mechanical

[34] Alessia Guardasole, *Eraclide di Taranto. I frammenti* (Naples: D'Auria, 1997).
[35] Manuela Tecusan, *The Fragments of the Methodists* (Leiden: Brill, 2004), vol. 1.

organism. Whatever his intellectual ancestry, Asclepiades was placed firmly by later doctors among the rationalists, i.e. those who interpreted medicine as primarily a search for causes of disease. He enjoyed a great reputation as a therapist – one story has him almost raising a man from the dead – but it was probably his preference for gentle therapies, including wine, massage, passive exercise, and the use of music in treating mental illness, that gained him many patients. His memorable slogan, "Swiftly, safely, pleasantly," sums up his ideas on therapy, but also subtly denigrates his opponents. Not that he was averse to strong purgatives, and even pharyngotomy, if need be, and the early stages of one of his treatments for fever were described as torture. Whereas Archagathus, a century or more before, was already respected before arriving in Italy, Asclepiades was the first Greek doctor to make his name within Rome, an indication of that city's incorporation within the Greek cultural universe.[36]

Themison followed Asclepiades in insisting that good medicine was effective practice – he in particular had a reputation for his knowledge of herbs – but he went further in developing ideas about the body. Whereas Asclepiades had emphasized problems involving only its channels, Themison also blamed the size or shape of the globules in contributing to the three "common conditions" that encompassed all disease conditions – stricture, looseness, and a mixture of the two. Acute diseases were the result of stricture, chronic of looseness. These common conditions were immediately evident to the Methodist observer, and easily determined what type of opposite treatment was necessary. A complicated investigation into causes or a long search for the appropriate therapy was unnecessary; a knowledge of the body's anatomy might be valuable, but not the act of dissection itself.[37]

By emphasizing three basic "commonalities," Themison restricted Asclepiades' wider-ranging understanding of illness, and set out firm guidelines for what the doctor should consider as causes. Later Methodists went further in subdividing his common conditions, e.g. into those requiring dietetic treatment and those requiring surgery, which in turn were divided according to the site of the condition and the size of the globules or pores. Therapy could be prescribed according to the age, sex, or state of the patient, but in a way that avoided a complex and lengthy investigation into the nature of the individual patient. All this imparted a certain degree of flexibility to Methodist medicine, as can be seen in the writings of Soranus of Ephesus at the end of the first century CE and, much later, in those of Caelius Aurelianus, whose work on acute and chronic diseases depends heavily on Soranus. They combine sound observation and clear description, and one can see why they were influential. Soranus' *Gynaecology* is the best

[36] John T. Vallance, *The Lost Theory of Asclepiades of Bithynia* (Cambridge: Cambridge University Press, 1990); Nutton, *Ancient Medicine*, pp. 170–3.

[37] Tecusan, *Fragments of the Methodists*, pp. 101–2, lists the fragments of his writings.

tract on the subject to survive from antiquity. It shows shrewdness in its assessment of the problems involved in all aspects of childbirth and the treatment of very young children, and resembles a modern textbook in its clarity and, at times, the effectiveness of its recommendations. It is not surprising that, alone among Methodists, he gained the approval of Galen.[38]

The Methodists' opponents attacked them on a variety of grounds. Celsus thought their generalized approach suitable for treating the sick in a large barrack hospital, but lacked the understanding of the individual patient he thought essential for proper medicine. But as a Roman gentleman he could afford to pay highly for personal attention. Galen's particular bugbear was Thessalus of Tralles, a charismatic and flamboyant Methodist doctor of the time of Nero. Thessalus, a recent migrant, paraded around with a retinue like a star performer, and his claim to be the "champion doctor" has echoes of the public arena. Galen could never resist a jibe at this man and his following of Thessalian asses, a neat reference to Thessaly's reputation in producing horses. Galen's complaints were many. He grumbled about the Methodists' lack of precision in terminology and their refusal to concern themselves with questions of aetiology. He was scathing about Thessalus' simplistic therapies, especially his insistence on starving the patient for the first two or three days. No wonder that Thessalus could claim to teach medicine in six months, if all he could offer were large-scale generalizations and therapies that might kill as much as cure. Soranus is the only Methodist for whom he has any time.[39]

But Galen's polemics are devastatingly unfair, and he himself admits that contemporary Methodism had changed somewhat since the days of Thessalus. The therapies in Soranus and Caelius Aurelianus display a flexibility denied by Galen to Thessalus. Even the celebrated *diatritos*, Thessalus' injunction to break down treatment into three-day periods, i.e. treating on the first, third, fifth day, and so on, had the advantage of simplicity compared with Hippocratic numerology, and it also forced the doctor to continually observe and reassess the patient, modifying treatment if necessary. In its stress on generalities, Methodism suited a Roman environment, a city of immigrants and the poor, who could not afford the long-term acquaintance of doctor and patient Celsus and Galen thought essential for understanding the nature of the patient in both health and disease. Patients had other ideas. The emperor and his family and the haute bourgeoisie of Ephesus or Smyrna employed Methodists, and the sect continued to flourish well into the fifth century.

[38] Owsei Temkin, *Soranus' Gynaecology* (Baltimore, MD: Johns Hopkins University Press, 1956); Geoffrey E. R. Lloyd, *Science, Folklore and Ideology: Studies in the Life Sciences of Ancient Greece* (Cambridge: Cambridge University Press, 1983), pp. 168–200.

[39] Tecusan, *Fragments of the Methodists*, p. 103, lists the fragments of his writings; David Leith, "The diatritos and Therapy in Greco-Roman Medicine," *Classical Quarterly* 58 (2008), 581–600.

By contrast with Empiricists and Methodists, the third (and by far the largest) grouping of the traditional triad, the "Rationalists," is heterogeneous, united only in a belief in the importance of causation and of individualist therapy. But, leaving aside writers whose works no longer exist, such as the imperial doctor Tiberius Claudius Menecrates, who expounded his own "clear and logical sect" in 156 books, the range of possible causes they considered was almost limitless. Democritean doctors, one of them a friend of Plutarch, believed in a universe of atoms, the so-called Pneumatists in a body whose health was primarily determined by a balance of humors under the dominance of *pneuma* or spirit. At Elea, southern Italy, there flourished for several centuries a quasi-religious association of physicians that claimed descent from Parmenides the Presocratic philosopher.[40]

Many other doctors looked back to the great Hippocrates, and it is tempting to see the triumph of the theory of the four humors as inevitable. But, in a telling aside, Galen laments that this was a minority view, and even among followers of Hippocratics there was far from agreement on which Hippocratic texts should form the basis for modern medicine. Galen himself rejected as inauthentic over half of *The Nature of Man*, the crucial text for a belief in four humors, while at least two doctors combined this theory with a belief in a very different theory of flux found in *Diseases* I, a treatise scarcely mentioned by Galen. Even where there was a strong tradition of Hippocratic exegesis, notably at Alexandria, Galen found many reasons to distinguish his Hippocratism from that of others – and a survey of citations from Hippocratic texts reveals many titles, like the *Testament of Hippocrates*, that he does not deign to mention.[41]

This heterogeneity is best interpreted not, as Galen implied, as proof of the intellectual incompetence of his adversaries, but as evidence for the vitality of medicine in the first two centuries of the Roman Empire. Aretaeus of Cappadocia, for example, wrote an important treatise on acute and chronic diseases that contains some shrewd observations of diseases, including paralysis and diabetes. Many of his comments, like his distinctions between childhood and adult asthma, or between congenital asthma and one brought on by exposure to harmful fumes, remain valid today, and his graphic descriptions, modeled on those of Hippocrates and written deliberately in the same dialect of Greek, served as models for nosography down to the nineteenth century. Claudius Agathinus, who lived around 100 CE, wrote extensively on bathing, warning against the dangers of too hot or too cold baths – although the latter were useful in toughening up the body. (This was an appropriate topic for a society in which bathing was as much a social event as a hygienic pursuit. Public baths were often the largest

40 Nutton, *Ancient Medicine*, pp. 207–21; Samama, *Les Médecins*, nos. 461, 507.

41 Wesley D. Smith, *The Hippocratic Tradition* (Ithaca, NY: Cornell University Press, 1979).

building in a city, as in Trier or Paris, and their remains still inspire wonder today.[42] It was a poor city that could afford only one such establishment.) A pupil of Agathinus, Archigenes of Apamea, paid particular attention to the diagnostic significance of the pulse, attempting to classify its rhythms musically, and giving them names, "mouse-tail," or "double hammer," that endured for centuries.[43] Athenaeus of Attaleia, an older contemporary, was praised by Galen as the best modern author of a general medical treatise, dealing with topics as varied as embryology, town-planning, and water supply. Medicine was also becoming part of general culture. Athenaeus advocated that everyone should read books on medicine, not just because it might be impossible to find a doctor or, at any rate, because consulting him for minor ailments was tedious, but also because one might gain psychological benefit from consorting with great philosophers and doctors from the past. That his injunctions were often followed can be gauged by the number of non-medics who accompanied Galen on his rounds and by a passing comment by Aulus Gellius, living in Athens around 160 CE, that it was a social as well as an intellectual faux pas not to know the difference between veins and arteries.

SURGERY

Of all the therapies recommended by these authors, it was surgery that developed most over the Hellenistic and Roman periods. A variety of instruments, some made by specialist makers, others devised by the practitioners themselves, were used in operations, and many, including trepans and vaginal specula, resemble modern surgical tools closely in shape and sophistication. The increased anatomical knowledge of Alexandrian surgeons led to new operations, some very complex and now involving penetrating the body cavity. New techniques were devised for setting fractures or dealing with dislocations, while the increasing size and professionalization of the Roman army from the first century BCE onwards must also have led to a greater expertise in treating wounds. The large permanent fortresses, as at Neuss (Germany), Chester (England), or Windisch (Switzerland), and even some smaller forts on the Danube frontier, had hospitals, some with a large room that may have served as an operating theatre. All were carefully planned with special attention to lighting and even to the placing of beds so as to be out of the way of draughts. Some may have dealt with battle casualties, but most fighting took place some way away from the fortress, and

[42] Garrett Fagan, *Bathing in Public in the Roman World* (Ann Arbor, MI: University of Michigan Press, 1999); Fikret K. Yegül, *Baths and Bathing in the Roman World* (Cambridge: Cambridge University Press, 2010).

[43] Harris, *The Heart*, pp. 237–57.

the notion of an ancient triage system, in which only the seriously wounded were sent back for specialist treatment in a base hospital, is largely fantasy. Nonetheless, there must have been a constant stream of injuries from training, although most of the soldiers recorded on sick lists suffer from ordinary diseases, such as conjunctivitis.[44] But many of the complicated operations, such as a rib resection, rhinoplasty, tracheotomy, breast reduction, and herniotomy, ascribed to Greek surgeons of this period, took place in civilian life. Papyri from Egypt show some effective methods of dealing with fistulae and colobomata, and Galen expected the accomplished surgeon to be able to deal with hernias, aneurysms, bladder-stones, cataracts, tumors, and a wide range of diseases of the ear, eye, and throat.[45]

His expectations of the accomplished physician in *Examining the Physician* are even more extensive – and demand a patient with the time and money to spend, and the ability to pose the question and assess the answers from several possible therapists. (Numbers of doctors or healers are hard to determine, but a ratio of 1 per 1,000 of the population is not impossible; many doctors also held land, and some treated animals as well as humans. Several women doctors are known, some of whom possibly treated both men and women and children.) From his ordinary physician Galen demands a knowledge of the principal tenets of "the ancients," largely doctors from Hippocrates to Erasistratus, and especially with Hippocrates and his notion of prognosis.[46]

He says nothing in this treatise about the greatest Hippocratic before him, Rufus of Ephesus, although he is elsewhere always complimentary about him. Rufus, who was active at the end of the first century, studied in Alexandria; whether he ever visited Rome is less clear, as, unlike Soranus, he does not criticize what he may have found in the big city. His most quoted work, alas surviving today only in later quotations, was a huge medical encyclopedia, probably entitled *The Layman*, that ranged over the whole spectrum of medicine, from failing eyesight to kidney problems, and from child-care and advice for students to geriatric medicine. He gave advice on sexual dysfunction, and in his *Buying Slaves* sympathized with the sick slave in his painful affliction and with the owner on the consequent financial burden. His *Medical Questions*, the best surviving guide to ancient diagnosis, explains how the doctor, faced with a new patient, can discover the cause of

44 Patricia A. Baker, *Medical Care for the Roman Army on the Rhine, Danube and British Frontiers from the First through Third Centuries* AD (Oxford: British Archaeological Reports, 2004); Julianne C. Wilmanns, *Der Sanitätsdienst im römischen Reich* (Hildesheim: Olms Weidmann, 1995).

45 Sabbah and Mudry (eds.), *La médecine de Celse*, pp. 173–210; Ralph Jackson, "Holding on to Health? Bone Surgery and Instrumentation in the Roman Empire," in Helen King (ed.), *Health in Antiquity* (London: Routledge, 2005), pp. 97–119; Marie-Hélène Marganne, *La chirurgie dans l'Égypte gréco-romaine d'après les papyrus littéraires grecs* (Leiden: Brill, 1998).

46 Galen, *Examining the Physician. Corpus Medicorum Graecorum*, trans. A. Z. Iskandar (Supplementum Orientale 4; Berlin: Akademie Verlag, 1988); Rebecca Flemming, "Women, Writing and Medicine in the Classical World," *Classical Quarterly* 57 (2007), 257–79.

his disease. Through a combination of observation, even of the timbre of the patient's voice, and precise questions, he builds on Hippocratic comments in, for example, *Airs, Waters and Places*, to discover the individuality of the patient as well as the possible reason for the illness. His insistence on what today is called "the patient's voice" shows that medical communication was not a one-way street. His own vocabulary, often using terms that could be widely understood, forms a contrast with that of contemporaries who, Galen complains, resorted to neologisms or an over-technical terminology.[47]

Like Galen, Rufus recommended the study of anatomy, lamenting that in his day it was now confined to surface and skeletal observation even at Alexandria. But, if Galen is to be believed, there was soon afterwards a revival of anatomy instituted by Marinus, a teacher there who wrote a large treatise on anatomy in twenty books, dealing with the whole body and especially the vascular system. He transmitted this interest to pupils like Quintus in Rome and Numisianus, who worked at Corinth and, later, Alexandria. Others took up the challenge. The young Galen wrote a little tract on the anatomy of the womb while a student in Asia Minor around 148, where he was taught by teachers trained in anatomy at Alexandria; Martialis (or Martianus) the Erasistratean enjoyed a great reputation in Rome in the 160s; and his contemporary Lycus of Macedon wrote a large anatomical work in nineteen books. Lycus carried out both dissections and vivisections, and possibly even a post mortem, although the extent to which he operated on humans is unclear. He certainly devoted much space in his large treatise to the anatomy of the womb and the foetus, and Galen thought it worthwhile to prepare his own summary of the tract in two books. But elsewhere, perhaps because of the parallels between their careers, Galen is scathing about "what Lycus didn't know." Mistakes Lycus may have made, but one should always beware of the self-serving rhetoric of Galen (129–ca. 216), the most influential doctor of antiquity in the range of his later influence.[48]

GALEN

The son of a wealthy architect, Galen enjoyed a privileged upbringing before being turned to medicine as a result of a dream sent to his father by Asclepius.[49] Asclepius was one of the patron gods of Pergamum, and

[47] Peter Pormann (ed.), *Rufus of Ephesus: On Melancholy* (Tübingen: Mohr Siebeck, 2008); Lloyd, *Science, Folklore and Ideology*, pp. 149–67.

[48] Véronique Boudon-Millot, *Galien* (Paris: Les Belles Lettres, 2007), vol. 1, 145–54.

[49] Major studies of Galen include Owsei Temkin, *Galenism: Rise and Decline of a Medical Philosophy* (Baltimore, MD: Johns Hopkins University Press, 1973); Hankinson (ed.), *Cambridge Companion*; Susan P. Mattern, *Galen and the Rhetopric of Healing* (Baltimore, MD: Johns Hopkins University Press, 2008); Véronique Boudon-Millot, *Galien de Pergame: un médecin grec à Rome* (Paris: Les Belles Lettres, 2012); Susan P. Mattern, *The Prince of Medicine: Galen in the Roman Empire* (Oxford: Oxford University Press, 2013).

Galen was convinced that his whole career was directed by the god, even to the extent of occasionally performing treatments given him in a dream. Galen's medical studies at Pergamum, Smyrna, and Alexandria lasted a decade, the longest on record, and left him convinced of the superiority of the Hippocratic medicine of the four humors as well as of the importance of anatomy. He returned to Pergamum around 157 CE, where he quickly obtained the post as surgeon to the gladiators kept for the provincial games by the High Priest of Asia, evidence of his skill and his social contacts – and his enormous self-confidence, for he claims that at his interview he cut open the stomach of a monkey, sewed it up again, and challenged his competitors to do likewise. In 162 CE, he made his way to Rome, where, aided by friends from Pergamum, he made an immediate reputation by his public anatomical displays. His burgeoning clientele among the rich led, he claims, to near-murderous envy, and to his flight from Rome to Pergamum in 166. He may have traveled further looking for rare herbs and minerals in the Levant, but he was not forgotten. In 168 CE, he was called back to join the emperors on campaign in north Italy, but the sudden death of Lucius Verus caused his co-emperor Marcus Aurelius to return to Rome forthwith, and Galen may not have joined him until summer 169. From then on, he remained attached to the court, rejecting the Emperor's request to return with him on campaign, pleading a veto from Asclepius, but serving as doctor to the young prince Commodus. He served successive emperors well into the next century, although he may have lost (or slipped away from) his post towards the end of the reign of Commodus, "the worst in recorded history" in his view, and he may have returned briefly to Pergamum. Arabic sources place his death at the age of eighty-seven, after seventy years of medical practice.

Galen was immensely wealthy. His stipend as an imperial doctor equaled that of a leading government official, and he had estates in Asia Minor and Campania. The fire of Rome in 192 destroyed much of the gold, silver, plate, and bankers drafts he had deposited temporarily in a repository, along with much of his enormous library and his irreplaceable collections of instruments and recipes. The recently discovered *Avoiding Distress* reveals much about the contents of his library, which included rare works of literature and philosophy as well as medicine.[50] He claims to have made generous gifts to needy fellow practitioners and to have refused to charge fees – although a present of 800 gold pieces from a grateful client will have more than made up for that. Penetrating behind his ebullient rhetoric is difficult, but two things are clear. He obtained a great reputation in Rome, influencing a group of Christians and impressing as a doctor the philosopher Alexander of Aphrodisias, who, however, thought little of his philosophy. But, secondly, Galen cannot be viewed as a typical practitioner, although we

[50] Galen, *Avoiding Distress* 12–37 = Galen, *Psychological Writings*, ed. Peter N. Singer (Cambridge: Cambridge University Press, 2013), pp. 81–9.

know of others, like Statilius Criton, doctor to Trajan, or Statilius Attalus, Galen's contemporary at court, who made an identical passage from the bourgeoisie of Asia Minor to Rome. With great wealth, education, and status, he stands at the apex of a pyramid of healers. The complaint of competitors that he got where he did because of his wealth is not easy to rebut.

Galen was an enormously learned and prolific author, with over 300 titles to his credit, half of which are now lost. But what remains, in Greek and in a variety of translations into Syriac, Arabic, Latin, Hebrew, and Armenian, occupies, in modern editions, roughly two meters of shelf space. They range in date from his early student days (*Medical Experience*) to the end of his long life (*My own Opinions, My own Books*), and some run into several individual books. They cover moral philosophy, lexicography, logic, bloodletting, anatomy, pharmacology, physic, and erudite commentary. Only surgery is missing, although the final books of the *Method of Healing* show his competence, even if they are not the big tract of surgery that he planned to write. Some of this learning, particularly in pharmacology, is not his own, and it is always hard to disentangle what is Galen's from what he inherited from others, such as Rufus, whose names he does not always mention.[51] His habit of adding comments and cross-references over the years to his own copies complicates still further the dating of his ideas. But he frequently insists on his own point of view, even against fellow Hippocratics, and often one can follow his individual train of thought as one comment sparks another, even if the result is neither as coherent nor as complete as he insists when he looks back.

His main achievements fall under three headings, each of which looks back to an earlier hero: philosophy, to Plato; anatomy, to Aristotle; and clinical medicine, to Hippocrates. But although each may be treated separately, the great strength of Galenic medicine depended on their interaction.

This interaction was helped by Galen's command of language and logic. Merciless in argument, impeccable in his deductions, albeit often from dubious premises, and scathing in his rhetoric, he overpowered with words all he disagreed with or who needed to be convinced of the truth of his discoveries. Logic lay at the base of his clinical investigations, drawing conclusions or analyzing causes from observational data. His long treatise, *Demonstration*, is lost in Greek, but the surviving fragments show that this was a major contribution to logic, criticizing and in its sophistication going well beyond Aristotle's theories of demonstration. It was widely read in Arabic translation by medieval scholars in Baghdad, and at least the Arabic translation may one day come to light.[52]

[51] Boudon-Millot, *Galien*, pp. 87–234.

[52] Teun Tielemann, "Methodology," in Hankinson (ed.), *Cambridge Companion*, pp. 49–65; Ben Morrison, "Logic," in Hankinson (ed.), *Cambridge Companion*, pp. 66–115; Riccardo Chiaradonna, "Galen on What is Persuasive (Pithanon) and What Approximates to Truth," in Peter Adamson, Rotraud Hansberger, and James Wilberding (eds.), *Philosophical*

His medical explanations depend on a universe based on Aristotelian physics and on Platonic psychology. Plato's belief in a tripartite soul, particularly as expounded in the *Republic* and *Timaeus*, Galen wrongly attributed to an acquaintance with Hippocrates, drawing from it further proof that Hippocrates had taught that the body depended on three systems, veins, coming from the liver, arteries, from the heart, and nerves, from the brain. Galen also posited a strong interaction between physical and mental states, claiming a particular expertise in identifying stress diseases and, vice versa, at times suggesting a philosophical approach to behavioral problems. His own calm reaction to the disaster of the fire of 192 CE showed, he claimed, the value of a proper family background and a sound philosophical training.[53]

Hippocrates was for him the supreme doctor, familiar with all aspects of medicine, including anatomy, although it was Aristotle and the Alexandrians who most inspired Galen with a passion for the knife. He dissected daily, initially at Rome in public but later in front of a select group of adherents as part of a coherent programme of research. His dissections of the brain and nervous system show a remarkable technique and a sharp eye, discovering new structures, like the recurrent laryngeal nerve, even if he also made many errors, most notably in his belief in holes in the intraventricular saeptum of the heart and in positing a *rete mirabile* in humans. Most of his mistakes were owed to the conditions under which he dissected, and especially to his necessary reliance on animals, mainly rhesus monkeys, but also sheep, pigs, and even horses and cows, where their size made investigation of tiny structures like the hyoid bone possible. He repeated Erasistratus' experiments on bloodflow and the brain, but his rejection of mechanism in favor of a belief in teleology hampered his approach to the vascular system. But the sheer detail of his dissections convinced many that it would be impossible to go beyond what he had written and, moreover, few had the wealth to buy so many animals to dissect. Likewise, his pious belief in a Creator God confirmed to Christians, Muslims, and Jews that Galen had allied religious to philosophical truth.[54]

Hippocrates in his view had laid the foundations of true medicine. Galen adhered strongly to the theory of the four humors, although he more often preferred qualitative explanations for illness. He devoted many hours to writing detailed commentaries on the often obscure Hippocratic treatises, using them as guides to practice. In this he stressed three things: the

Themes in Galen, Supplement to the Bulletin of the Institute of Classical Studies (London: Institute of Classical Studies, London, 2014), pp. 61–88.

53 Singer, *Psychological Writings*.

54 Julius Rocca, *Galen on the Brain* (Leiden: Brill, 2003); Julius Rocca, "Anatomy," in Hankinson (ed.), *Cambridge Companion*, pp. 242–62.

individuality of the patient; acute observation; and prognostic skill. Observational ability his writings display in abundance, whether he is commenting on the phenomena of illness, the habits of animals, or the agricultural landscape through which he traveled. This was the foundation of his belief in prognosis, a skill he claimed almost entirely neglected since Hippocrates. It involved not only telling the future course of an illness, to the amazement of patient and audience, but also understanding the present and past of an illness. It encompassed diagnosis, reached particularly through taking the pulse and examining urine, but also by talking to patients and observing their surroundings, and, when correctly done, it allowed the doctor to establish the cause of the disease and devise a means of curing the patient by removing the cause. Galen favored dietetics over drugs and surgery, but his skill with the knife was also used to let blood safely and on one occasion to remove a suppurating breastbone from a slave. His knowledge of the pathways of the nerves also allowed him to diagnose why a famous public speaker had lost the use of some of his fingers. He had fallen off his chariot onto his back, and it was his spine that needed attention, not, as Galen's competitors insisted, the hand or fingers. The cooperation of doctor and patient was for him an essential element in the healing process, and his medical ethics, with advice on such things as length of hair or tone of voice, were aimed at creating the confidence in the doctor that would lead to a successful cure. The placebo effect may explain many of Galen's cures, but he also deployed appropriately many of the substances recommended by modern herbalists.[55]

Galen's pre-eminence was recognized by his contemporaries, and still more by later generations, who looked back to his achievements in awe and wonder. They were not foolish, but their justifiable desire to reduce his huge oeuvre to manageable proportions eliminated much that defined his superiority – his rhetoric, his remarkable powers of observation, and his experimentalism. He was not entirely the dogmatist bogey-man his renaissance opponents denounced. But his stature is also enhanced by the absence of competitors. Inscriptions and papyri reveal the names, and often activities, of many doctors, lecturing, writing, and receiving the praise of a grateful community. Some are even depicted palpating or taking the pulse. Some doctors held public office, participated in festivals and processions, or simply paid taxes on their lands. Some wrote poetry or philosophy, others told of life on campaign or at the flamboyant court of Cleopatra. Some belonged to dynasties of medical men, others learned their trade as slaves or freedmen. But little or nothing survives to provide a picture of an individual practitioner to rival that of Galen. Rufus and Soranus remain shadows by comparison. That melancholy conclusion

[55] Philip van der Eijk, "Therapeutics," in Hankinson (ed.), *Cambridge Companion*, pp. 283–303; Rebecca Flemming, "Commentary," in Hankinson (ed.), *Cambridge Companion*, pp. 323–54.

helped confirm the superiority of Galen and Galenic medicine, for why else should he survive when others did not? But it should not blind us to the ways that, even before he bestrode the stage, others had made major discoveries and enjoyed great reputations. Nor should his rhetoric be allowed to obscure what is arguably the most important development of this period. His career shows the extent to which by 150 CE, if not much earlier, Greek medicine, Greek medical practitioners, and Greek medical institutions had become Roman.[56]

[56] Samama, *Les Médecins*; Geoffrey E. R. Lloyd, "Galen and his Contemporaries," in Hankinson (ed.), *Cambridge Companion*, pp. 34–48; Vivian Nutton, "The Fortunes of Galen," in Hankinson (ed.), *Cambridge Companion*, pp. 355–90.

18

GREEK MATHEMATICS

Nathan Sidoli

This chapter discusses the tradition of theoretical mathematics that formed a part of Greek literary culture. This was not the only kind of mathematics that existed in the ancient Greek world. There were also traditions of elementary school mathematics, and the subscientific traditions of mathematics that were handed down by various professionals who used mathematics in their work, such as tradesmen, builders, accountants, and astrologers.[1] In fact, in the ancient Mediterranean, these subliterary traditions almost certainly formed the vast majority of all the mathematics studied and practiced, while literary, theoretical mathematics was practiced by only a privileged few.[2] Nevertheless, the elite, literary status of the theoretical mathematicians, along with the brilliance of their work, ensured that much of their project was preserved for posterity, while the predominantly oral, subscientific tradition survives only in scattered, material fragments and a few collections passed down through the manuscript tradition, attributed, probably erroneously, to Heron (ca. mid-first to third century CE).

In the classical period, when the forms of literary mathematical texts were being established, elite, theoretical mathematicians appear to have done as much as they could to separate themselves from professionals who used mathematics, just as they strove to distinguish themselves from other elites, such as sophists and philosophers who did not engage in mathematical activity. Theoretical mathematics was originally not a professional, institutionalized activity. During the Hellenistic period, when this, now more

[1] The term "subscientific" is due to J. Høyrup. See, for example, J. Høyrup, "Sub-scientific Mathematics: Observations on a Pre-Modern Phenomena," *History of Science* 27 (1990), 63–87. The social context of such "subscientific" mathematics is treated by S. Cuomo, *Ancient Mathematics* (Routledge: London, 2001).

[2] M. Asper, "The Two Cultures of Mathematics in Ancient Greece," in E. Robson and J. Stedall (eds.), *The Oxford Handbook of the History of Mathematics* (Oxford: Oxford University Press, 2008), pp. 107–32.

institutionalized, theoretical mathematics was applied to serious problems in natural science and engineering, there were a number of attempts to wed the theoretical and practical traditions of mathematics, but the social context and institutional settings of the two remained distinct. Finally, in the Imperial and Late Ancient periods, although creative mathematics was less practiced, mathematical scholarship was thoroughly institutionalized in the philosophical schools, and mathematics and philosophy were finally united, although for many mathematical scholars of the late period, philosophy was accorded a superior position.

The range of ideas and activities that were designated by the word *mathēmatikē* are not identical to those denoted by our word *mathematics*.[3] From the earliest times, *mathēmatikē* was connected with any branch of learning, but came to denote the mathematical sciences centered around arithmetic, geometry, astronomy, and harmonics. From the time of the Pythagoreans (ca. late sixth to fifth century BCE) to that of Ptolemy (ca. mid-second century CE), astronomy and harmonics were not regarded as applications of mathematics, but as core areas of the enterprise. *Mathēmatikē* eventually denoted those literary disciplines that used mathematical techniques or investigated mathematical objects, whether actual or ideal, and which included fields such as optics, sphere-making, or astrology along with abstract investigations such as the theories of whole numbers or conic sections.

Nevertheless, ancient thinkers had a fairly clear idea of what constituted a mathematical work and who was a mathematician. They described different arrangements of the mathematical sciences and set out various relationships between mathematical fields and other branches of theoretical knowledge. Ancient discussions of mathematicians revolve around a core group of frequently repeated names.[4] We find debates about the legitimacy of certain mathematical arguments in philosophical works, but rarely in the surviving mathematical texts. Mathematicians defined mathematics by the nature of the texts they wrote. From around the middle of the fifth century BCE, mathematical texts were highly structured, involving explicit arguments, often centered around diagrams and letter-names or other technical apparatus, and mathematicians were those people who could produce such texts. They developed particular ways of speaking, or rather of writing, and these further served to reinforce the exclusive tendencies of this small literary group.

[3] See G. E. R. Lloyd, "Methods and Problems in the History of Ancient Science," *Isis* 83 (1992), 564–77, 569–70.

[4] Netz identified 144 mathematicians in ancient sources, many of whom are only known to us by name. See R. Netz, "Classical Mathematicians in the Classical Mediterranean," *Mediterranean Historical Review* 12 (1997), 1–24.

OBSCURE ORIGINS

The origins of Greek mathematics, either in Greek philosophy or as an independent discipline, used to be a favorite topic for historians of Greek mathematics.[5] The difficulty is that we have little certain evidence about the details of these origins. Most of our evidence for this early work comes from writings that were produced centuries later, or through the filter of philosophical writings that were not intended to be of historical value and were irregular in their use of mathematical prose. One of our most important sources for this early period is the so-called catalog of geometers, which appears as a passage in Proclus' *Commentary on Elements Book I*.[6] Parts of this passage are thought to go back to a now lost *History of Geometry* written by Aristotle's student and colleague, Eudemus of Rhodes (ca. late fourth century BCE). Proclus (412–85 CE), however, does not explicitly attribute this passage to Eudemus, in contrast to other borrowings, and it has clearly been modified by other authors over the years. Moreover, this passage does not provide much description of the mathematical activities of the individuals it names and, hence, serves us mostly as a relative chronology. The earlier evidence we have for the origins of Greek mathematics, such as in the writings of Plato and Aristotle, is often vague, not attributing work to individual mathematicians, or presenting incomplete arguments.[7] It should be clear from this that our most expansive sources are far removed in time from the persons and events they recount and must be treated with due caution.

The images of Thales and Pythagoras writing detailed mathematical proofs now seem to be the stuff of legends. When we consider how late our sources for this activity are, the other writings that remain from this early period by their contemporaries, and the tendency of Greek doxographers to produce rationalized histories by associating well-known results with well-known thinkers, it appears increasingly unlikely that the early Presocratics or Pythagoreans produced writings containing deductive mathematics. Although Thales and the early Pythagoreans may well have made various correct assertions about elementary geometry and some sort of argument for why these were true, the proofs attributed to them by Proclus are now

[5] For examples, see A. Szabó, *The Beginnings of Greek Mathematics*, trans. A. M. Ungar (Dordrecht: Reidel, 1978), and W. Knorr, *The Evolution of the Euclidean Elements* (Dordrecht: Reidel, 1975).

[6] G. Friedlein, *Procli Diadochi in Primum Euclidis Elementorum Librum Commenatarii* (Leipzig: Teubner, 1873). For a recent discussion of the catalog of geometers, see L. Zhmud, *The Origin of the History of Science in Classical Antiquity* (Berlin: Walter de Gruyter, 2006), pp. 179–90.

[7] For the mathematics contained in Plato and Aristotle, see R. S. Brumbaugh, *Plato's Mathematical Imagination* (Bloomington, IN: Indiana University Press, 1954) and T. Heath, *Mathematics in Aristotle* (Bristol: Thoemmes Press, 1949). For examples of the difficulties involved in interpreting the mathematics in Plato and Aristotle, see F. Acerbi, "Plato: Parmenides 149a7–c3. A Proof by Complete Induction?," *Archive for History of Exact Sciences* 55 (2000), 57–76, and H. Mendell, "Two Traces of Two-Step Eudoxan Proportion Theory in Aristotle: A Tale of Definitions in Aristotle, with a Moral," *Archive for History of Exact Sciences* 61 (2007), 3–37.

generally thought to have been reconstructed by later ancient authors.[8] Although Pythagoras almost certainly had a profound interest in numbers, there is no early evidence that he or the early Pythagoreans produced deductive mathematics.[9] The earliest Pythagoreans of whom we can develop a clear picture are Philolaus of Croton (ca. turn of the fourth century BCE) and Archytas of Tarentum (ca. mid-fifth to mid-fourth centurries BCE) – but the mathematical work of Philolaus is rather meager, while Archytas is already a contemporary of Plato's and his mathematics does not appear to have been uniquely Pythagorean.[10]

Although much of the early history of Greek mathematics was probably reconstructed by Eudemus and other authors on the basis of vague attributions and the guiding belief that mathematics must develop in a rational way, in the case of Hippocrates of Chios (ca. mid-fifth century BCE), Eudemus seems to have had some written sources. Indeed, Simplicus (ca. early sixth century CE), in his *Commentary on Aristotle's Physics*, tells us that Eudemus regarded Hippocrates as one of the earliest mathematicians.[11] We are told that Hippocrates worked on the problem of doubling the cube and reduced this to the more general problem of constructing two mean proportionals between two given lines,[12] and was the first to write up the principles of geometry in an *Elements*.[13] Most importantly, however, we have a long fragment of his work preserved by way of Eudemus' *History of Geometry*. In this passage, he shows how to square three different types of lunes, in an attempt to use these figures to reduce the problem of squaring a circle to something more manageable.[14] From these writings we learn that at this time there were certain problems that were considered worthy of solution and that Hippocrates approached these using a general strategy of reducing them to problems that were soluble by constructions made on lettered diagrams. From the details of his arguments, we see that the types of

[8] Proclus, in his *Commentary to Elements Book I*, assigns *Elements* I 15, 26 to Thales and *Elements* I 32 and 42 to the Pythagoreans. See Friedlein, *Procli Diadochi*.

[9] W. Burkert argued that there is no early evidence for mathematical activity by Pythagoras himself and that there was nothing uniquely Pythagorean about the mathematics practiced by later members of the school. See W. Burkert, *Lore and Science in Ancient Pythagoreanism* (Cambridge, MA: Harvard University Press, 1972). A case that the early Pythagoreans worked in deductive mathematics was remade by L. Zhmud, *Wissenschaft, Philosophie und Religion im frühen Pythagoreismus* (Berlin: Bücher, 1997), but has not gained much acceptance.

[10] For Philolaus and Archytas, see C. Huffman, *Philolaus of Croton* (Cambridge: Cambridge University Press, 1993), and C. Huffman, *Archytas of Tarentum* (Cambridge: Cambridge University Press, 2005).

[11] H. Diels, *Simplicii in Aristotelis Physicorum libros quattor priores commentaria* (Berlin: Remer, 1882), p. 69.

[12] For example, Eutocius, *Commentary on Archimedes' Sphere and Cylinder*. As for Eratosthenes, see J. L. Heiberg, *Archimedis opera omnia cum commentariis Eutocii* (Leipzig: Teubner, 1910–15), vol. 3, 88.

[13] This comes from Proclus in *Commentary to Elements Book I*. See Friedlein, *Procli Diadochi*, p. 66.

[14] W. Knorr, *The Ancient Tradition of Geometric Problems* (Boston, MA: Birkhäuser, 1989), pp. 29–39, and R. Netz, "Eudemus of Rhodes, Hippocrates of Chios and the Earliest Form of a Greek Mathematical Text,: *Centaurus* 46 (2004), 243–86.

constructions allowed and the starting points of the argument were still taken at the geometer's discretion.

Another important development in the fifth century was the discovery of incommensurability – that is, the realization that there are cases where two given magnitudes have no common measure.[15] For example, the combination of Pythagorean interests in number theory and various geometric studies came together in deductive arguments that the diagonal of a square is incommensurable with its side.[16] This, in turn, led to considerable interest in incommensurability and excited the activity of some of the best mathematicians of the fifth and fourth centuries: Theodorus (ca. late fifth century BCE); Eudoxus (ca. mid-fourth century BCE); and Theaetetus (ca. early fourth century to 369 BCE). What we think we know about the work of Theodorus is the product of repeated reconstructions by modern scholars based on a few tantalizing hints found mostly in Plato's writings.[17] In the case of Eudoxus and Theaetetus, however, some of their work is considered to have formed the substantial basis of *Elements* V and X, respectively, and hence to form the foundations of ratio theory and to provide the most complete surviving study of incommensurability.

In the fourth century, there were a number of important mathematicians who were traditionally associated with Plato's Academy. Although it is now clear that the image of Plato as an organizer of mathematical activity is more a product of the scholars of the early Academy than a reflection of reality,[18] Plato's writings are full of mathematical references and claims about the benefits of a mathematical education. Nevertheless, there is no reason to believe that Plato had more influence on his mathematical colleagues than they had on him. On the contrary, it seems clear that Plato used the ideas that he learned from Theodorus, Archytas, and Theaetetus to develop his own philosophy.

Whatever the case, this period of Athenian cultural influence acted as catalyst for significant mathematical work. Solutions to the problem of duplicating the cube, in Hippocrates' reduction to the problem of finding two mean proportionals, were put forward by Archytas, Eudoxus, and Menaechmus (ca. mid-fourth century BCE).[19] Archytas produced theorems in number theory that may have formed the basis of *Elements* VIII and the Euclidean *Section of the Canon*.[20] A theory of conic sections was developed

[15] Knorr dates this discovery to 430–410 BCE. See Knorr, *Evolution*, p. 40.
[16] Knorr, *Evolution*, pp. 21–8.
[17] Ibid., pp. 62–130.
[18] L. Zhmud, "Plato as 'Architect of Science'," *Phronesis* 43 (1998), 211–44.
[19] Knorr, *Ancient Tradition*, pp. 50–76.
[20] Knorr, *Evolution*, pp. 211–25. Note, however, that the attribution to Archytas of *Elements* VIII and *Section of the Canon* is rather tenuous. See C. Huffman, *Archytas*, pp. 451–70.

by Eudoxus, Menaechmus, and others and applied to the solution of various problems. It is difficult to appraise how general or complete this theory was.

Eudoxus was one of the most brilliant mathematicians of this period. He developed a ratio theory that was closely related to the ratio theory we now read in *Elements* V, but which may have relied on a two-stage argument, showing first the case of commensurable magnitudes and then that of incommensurable magnitudes.[21] Although this work has generally been read as an attempt to make a general theory of ratio, it is not a complete foundation for contemporary or later mathematical practice and hence can also be interpreted as a collection of theorems useful for geometry.[22] Eudoxus also wrote a number of theorems concerning the mensuration of objects that were praised by Archimedes and which probably involved double indirect arguments. In astronomy, he put forward the two-sphere model of the cosmos with a spherical earth in the center of a spherical firmament, and a model of homocentric spheres to account for some set of celestial phenomena; unfortunately the details of these works are a matter of speculation.[23]

By the time Alexander began his conquests, Greek mathematics had undergone considerable development. Of the three great problems of antiquity – the quadrature of a circle, the trisection of an angle, and the duplication of a cube – the first two had been clearly articulated and addressed and the duplication of the cube had been solved. Although the details of their content are matters of reconstruction and speculation, general theories of number, ratio, and incommensurable magnitudes had been developed, and texts treating the elements of geometry were in circulation. We do not know the precise relationship between these early works and Euclid's *Elements*, but there is no direct indication in the sources that early geometers were interested in restricting the set of constructions that could be carried out to those that are abstractions from a straightedge and compass.

GEOMETRY IN THE HELLENISTIC PERIOD

Around the beginning of the Hellenistic period there were a number of projects to consolidate and reformulate the considerable body of

[21] W. Knorr, "Archimedes and the Pre-Euclidean Proportion Theory," *Archive for History of Exact Sciences* 28 (1978), 183–244, F. Acerbi, "Drowning by Multiples," *Archive for History of Exact Sciences* 57 (2003), 218–24, and Mendell, "Two Traces of Two-Step Eudoxan Proportion Theory."

[22] K. Saito, "Phantom Theories of Pre-Eudoxean Proportion," *Science in Context* 16 (2003), 331–47.

[23] There is disagreement about how Eudoxus' model functioned, and even about what phenomena it was supposed to model. See, for examples, B. Goldstein and A. Bowen, "A New View of Greek Astronomy," *Isis* 74 (1983), 330–40, H. Mendell, "The Trouble with Eudoxus," in P. Suppes, J. V. Moravcsik, and H. Mendell (eds.), *Ancient and Medieval Traditions in the Exact Sciences* (Stanford, CA: CLSL Publications, 2000), pp. 59–138, and I. Yavetz, "A New Role for the Hippopede of Eudoxus," *Archive for History of Science* 56 (2001), 69–93.

mathematical knowledge that Greek scholars had produced. At the early Lyceum, Eudemus of Rhodes wrote his histories of geometry, arithmetic, and astronomy.[24] Although it is not clear how many written sources he had for the earliest periods, his work became a major source for the later estimation of these fields. Around the turn of the century, or shortly after, Euclid (ca. early third century BCE), the most widely read mathematician of the ancient period, undertook a project of giving a solid foundation to, and a clear articulation of, nearly all branches of the exact sciences.

We know nothing of Euclid's life. It is often assumed that he worked at Alexandria, due to the circulation of a legend that he told King Ptolemy that there is no royal road to geometrical knowledge,[25] and to Pappus' statement that Apollonius worked with his students in Alexandria.[26] Claims that he worked at the Museum, or the Library, are simply embellishments of these two hints. What we do know is that he produced a considerable body of mathematical work, the influence of which increased throughout the Hellenistic period and which had become canonical by the Imperial period.

Euclid worked widely in nearly every area of mathematics and the exact sciences. Most of these works were concerned with geometry or the application of geometry to natural science. Although Euclid will always be known as the author of the *Elements*, he also wrote works on conic sections, solid loci, and porisms, and other specialized works related to geometrical analysis. He was regarded by Pappus (ca. early fourth century CE) as one of the three primary authors of the field of geometrical analysis, along with Aristaeus (ca. mid-fourth to mid-third century BCE) and Apollonius (ca. late third century BCE).[27] In the exact sciences, he wrote works on spherical astronomy, mechanics, optics, catoptrics, and harmonic theory, although the authenticity of the last three has been questioned.[28] Because of the fame of the *Elements*, and because his more advanced mathematical works have been lost, modern scholars have often seen Euclid as a mere compiler and textbook writer, not a productive mathematician. This was not, however, the assessment of ancient mathematicians, such as Apollonius and Pappus. Nevertheless, it is clear that Euclid had a great interest in codifying mathematical knowledge and setting it on a secure foundation.

We see this initially in his division of geometry into different domains based on the types of techniques that can be used and, more explicitly, in his

[24] For a full study of Eudemus' work, see Zhmud, *The Origin of the History of Science*.

[25] There are, in fact, two versions of this legend, both from late authors: one involves Euclid and King Ptolemy, while the other involves Menaechmus and Alexander the Great. See Proclus, *Commentary on Elements Book I*, chap. 4, and Stobaeus, *Anthology* II 31.115.

[26] Pappus, *Collection* VII 35. See F. Hultsch, *Pappi Alexandrini Collectionis* (Berlin: Weidmann, 1876) and A. Jones, *Pappus of Alexandria, Book 7 of the Collection* (New York: Springer, 1986).

[27] Pappus, *Collection* VII 1. See Jones, *Book 7 of the Collection*, p. 83.

[28] These questions of authenticity are highly subjective, often relying on dubious arguments such as claims that an author who produced the *Elements*, which has so many good arguments in it, could not also have produced the *Section of the Canon*, which contains some sloppy arguments.

arrangement of the *Elements*. The *Elements* begins with a series of definitions, postulates, and common notions, some of which may not have been authentic,[29] but most of which almost certainly were. In particular, the postulates, which deal with construction, right angles, and parallelism, are logically required by the development of the book. The line and circle construction techniques and the parallel postulate, along with the logical structure developed on the basis of these, may well have been Euclid's original contribution to the foundations of geometry. It is clear from what Aristotle says that the theory of parallel lines was still plagued with logical issues when he was writing,[30] and we find no mention of the construction postulates in authors before Euclid and no interest in a restriction to these constructions in authors who lived around his time, such as Autolycus, Aristarchus (both ca. early third century BCE), or Archimedes (280s–212 BCE). Hence, the *Elements* can be read as a treatise which develops mathematics along the Aristotelian principle of showing as much as possible on a limited set of starting points.

The early books of the *Elements*, I–VI, treat plane geometry; however, this includes *Elements* V on theory of ratios between magnitudes, where magnitude is understood as a hyperonym, or the abstraction of a feature, of geometric objects – the length of a line, the area of a figure, the volume of a solid, and so forth. The next three books, VII–IX, treat number theory, and *Elements* X deals with incommensurability, which arises as a topic when we try to apply the theory of numbers to magnitudes, such as lines and areas. The final three books develop a theory of solid geometry.

It has often been remarked that *Elements* V and X do not fit very well into this general plan, and hence must have been taken essentially unaltered from previous works. It is also possible, however, to read these books as being in their proper place in the overall deductive structure.[31] If we read the text as having a single architecture, Euclid's strategy appears to have been to introduce an idea or theory as late as possible, so as to show how much can be done without it. Once it became necessary to introduce new concepts and methods, however, Euclid does not treat them as purely instrumental, but presents them in the broader context of an articulated theory. This strategy can be exemplified by the first book. The structure of *Elements* I shows that congruency of triangles can be shown independently of the parallel postulate while equality of areas cannot.[32]

[29] For a discussion of the authenticity of the early definitions, see L. Russo, "The Definitions of Fundamental Geometric Entities Contained in Book I of Euclid's Elements," *Archive for History of Exact Sciences* 52 (1998), 195–219.

[30] Heath, *Mathematics in Aristotle*, pp. 27–30.

[31] For a discussion of the deductive structure of the *Elements* see I. Mueller, *Philosophy of Mathematics and Deductive Structure in Euclid's Elements* (Cambridge, MA: MIT Press, 1981). Mueller, however, does not read the text as having a single deductive structure and alters the ordering of the subject matter.

[32] The theory of congruency does, however, contain another unstated axiom in the construction of *Elements* I 16. This is the assumption that a line can always be extended so as to be longer than any given length, which does not hold on a surface of positive curvature, such as a sphere.

The so-called Pythagorean theorem, *Elements* I 47, is a culmination of the book, but we should not understand it as *the* goal. The book has various goals, such as the production of constructive methods for geometric problem-solving, the articulation of a theory of congruency that is only loosely related to constructive geometry, the foundation of a theory of parallelism, and its development into a theory of area, which results in *Elements* I 47 and in the theory of the application of areas found in *Elements* II.

Another major area of Euclid's mathematical endeavor was in what came to be known as the "field of analysis." Although most of this work has been lost,[33] it formed an important foundation for advanced research in mathematics and contributed to Euclid's high reputation as a mathematician in antiquity. Along with the *Data*,[34] he produced a *Conics*, which was later superseded by that of Apollonius; a short work on the *Division of Figures*;[35] a treatise on loci that form surfaces; and a treatise called *Porisms*, treating problem-like propositions that make assertions about what objects or relations will be given if certain conditions are stated about given objects.[36] While the *Data* was probably written more as a repository of useful results than as a treatise to be read from beginning to end, the *Solid Loci* and *Porisms* may have been written to teach students useful habits of thought in the analytical approach to problems. Whatever the case, by the Imperial period, these works were taken by Pappus, and other teachers of advanced mathematics, as part of a canon for the study of analysis.

In contrast to Euclid, Archimedes is the ancient mathematician that we know the most about.[37] From his own writings we learn that his father was an astronomer,[38] that he was associated with the court of Hieron II of Syracuse, and that he kept in regular contact with his mathematical colleagues working in Alexandria. Because of his iconic status in antiquity, many legends circulated about Archimedes, most of which are probably apocryphal.[39] Many of these tales concern his death, and many of them are likely to have been exaggerated as well. Nevertheless, it seems clear that he was an old man when he was killed by Roman solders during the sack of Syracuse under Marcus Claudius Marcellus, in 212 BCE.[40] It is also probable

33 Only the *Data* survives in Greek, along with a fragment of the *Division of Figures*, in Arabic.

34 See H. Menge (ed.), *Euclidis Data cum commentario Marini et scholiis antiquis* (Leipzig: Teubner, 1896) (*Euclidis Opera omnia*, vol. 6).

35 J. Hogendijk, "The Arabic version of Euclid's *On Divisions*," in M. Folkerts and J. P. Hogendijk (eds.), *Vestigia Mathematica* (Amsterdam: Rodopi, 1993), pp. 143–62.

36 For a discussion of the lost treatises on analysis by Euclid and others, see Jones, *Book 7 of the Collection*, pp. 510–619.

37 The standard work on his life and work is E. J. Dijksterhuis, *Archimedes*, trans C. Dikshoorn (Princeton, NJ: Princeton University Press, 1987).

38 This is based on a plausible conjecture for a meaningless passage in our received manuscripts. See F. Blass, "Der Vater des Archimedes," *Astronomische Nachrichten* 104 (1883), 255.

39 Archimedes was the archetypal Greek mathematician for ancient authors. See M. Jaeger, *Archimedes and the Roman Imagination* (Ann Arbor, MI: University of Michigan Press, 2008).

40 Plutarch, *Life of Marcellus*, 14–19, and Polybius, *Histories*, 8.5–8.8.

that the siege of the city was prolonged due to engines of war that Archimedes built, although none of these are likely to have been burning mirrors. Moreover, from the prefaces that he wrote to his works, we are able to get a sense of his personality. He comes off as having been justly proud of his abilities and aware of the fact that there were very few who could really appreciate what he was doing. He also, however, had a playful side and sometimes teased the Alexandrian mathematicians by sending them false propositions and encouraging them to find the proofs.[41] This playfulness also extended to his own work as we see from his analysis of an ancient game, the *Stomachion*, and his work the *Sand Reckoner*, in which he shows that his system of large numbers, which cannot be represented in the normal Greek number system, can handle computing the number of grains of sand that it would take to fill even Aristarchus' absurdly large universe, in which the earth is assumed to orbit around the sun.[42]

The relationship between Archimedes and Euclid is difficult to untangle. Archimedes may have been Euclid's younger contemporary, but since Euclid's dates are only vaguely known, it is also possible that Euclid died some years before Archimedes was born. Any direct influence from Euclid's work on Archimedes is also difficult to detect. On the one hand, Archimedes appears to have been unimpressed by Euclid's foundational approach, showed no interest in a restriction to line and circle constructions, and occasionally proved things using different techniques from those we find in the *Elements*. On the other hand, he does not repeat material we find in Euclid's works and it may be that, like a number of other great mathematicians, he came to appreciate Euclid's project more as he got older and was increasingly influenced by the *Elements* throughout the course of his long career.[43] Proclus' claim that Archimedes mentions Euclid used to be explained away by pointing out that a reference to *Elements* I 2 in *Sphere and Cylinder* I 2 is an obvious interpolation; however, in the so-called Archimedes Palimpsest there is another reference to Euclid elsewhere in *Sphere and Cylinder* that may be authentic.[44]

Archimedes' works can be divided into three types: geometrical; mechanical; and computational. In all three of these, however, we can see certain features that are characteristic of Archimedes' personal style. They tend to be

[41] Heiberg, *Archimedes*, vol. 2, 2–4.

[42] B. E. Wall, "The Historiography of Aristarchos of Samos," *Studies in the History and Philosophy of Science* 6 (1975), 201–88, and R. Netz, "The Goal of Archimedes' *Sand-Reckoner*," *Apeiron* 36 (2003), 251–90.

[43] W. Knorr used the different ways in which Archimedes employs Euclid's ratio theory to provide a dating of his works. See W. Knorr, "Archimedes and the Elements: Proposal for a Revised Chronological Ordering of the Archimedean Corpus," *Archive for History of Exact Sciences* 19 (1978), 211–90. See, also, B. Vitrac, "A propos de la chronologie des oeuvres d'Archimède," in J.-Y. Guillaumin (ed.), *Mathématiques dans l'Antiquité* (Saint-Étienne: L'Université Saint-Étienne, 1992), pp. 59–93.

[44] Friedlein, *Procli Diadochi*, pp. 68; R. Netz, W. Noel, N. Wilson, and N. Tchernetska (eds.), *The Archimedes Palimpsest*, 2 vols. (Cambridge: Cambridge University Press, 2011), vol. 2, 276–7.

short, never more than two books. They cover distinct problems, or areas, and start with axioms and construction techniques suitable to the task at hand, with little interest in reducing these to more elementary concepts or methods. They show an abiding interest in mensuration, often in the form of a numerical comparison between the properties of a well-known object and those of a less tractable one. On the whole, Archimedes wrote advanced mathematical texts for mathematicians and, with the notable exception of his *Method*, seemed to have had little interest in questions of pedagogy or foundations.

Archimedes wrote no elementary geometric treatises. He wrote a text *On Spiral Lines*, which defined the curves by moving points, set out some of their properties, based on arithmetic progressions and *neusis* constructions involving setting a given length between given objects, and concluded by finding a number of significant areas related to the curves. In *On Conoids and Spheroids*, he opens with a series of definitions and a discussion of the problems to be addressed, proves a number of theorems on arithmetic series and the areas of ellipses, and then enters the main body of the work, in which he treats the volumes of various conics of revolution and their sections. One of his most striking works is *On the Sphere and Cylinder*. The first book is a systematic treatment of the relationships between various objects – like a circle and polygons that inscribe and circumscribe it, a cone and pyramids that inscribe and circumscribe it, a sphere and the series of sections of a cone that inscribe and circumscribe it, and so on – which leads to finding key relationships between a sphere and a cone, or cylinders, which inscribe or circumscribe it. The second book then uses this material to solve a number of difficult problems and to show theorems dealing with the volume and area of segments of a sphere.

Archimedes' surviving works in mechanics have to do with statics and the equilibrium of floating bodies. In *On the Equilibria of Planar Figures*, he demonstrates the principle of the balance, which appears to have played a major role in his research. In the so-called *Method*, he explains to his correspondent Eratosthenes (ca. mid-third to early second century BCE) how he used the idea of a virtual balance to derive many of his important mensuration results, found in works like *Quadrature of the Parabola* and *Conoids and Spheroids*, using the notion of suspending what we could call infinitesimals so as to maintain equilibrium on a balance.[45] Using as an example a hoof-shaped object that he had not treated previously,[46] in *Method* 12–15, Archimedes shows how one can first investigate the solid heuristically using the virtual balance, then develop a proof strategy using indivisibles, and finally write a fully rigorous proof using a double indirect

[45] E. Hayashi and K. Saito, *Tenbin no Majutsushi: Arukimedesu no Sūgaku* (Tokyo: Kyoritsu Shuppan, 2011).
[46] A section of a right cylinder formed by a plane passing through the diameter of the base.

argument. Although the *Method* was not known before 1906 and appears to have exerted no influence on the development of mathematics, it is an invaluable glimpse into the working habits of antiquity's greatest mathematician.

Probably the most widely read of Archimedes' computational works was the *Measurement of the Circle*. Although our current version of the text appears to be an abridged and highly edited epitome of Archimedes' orginal work, it provides us with at least part of his general approach to the classical problem of squaring the circle.[47] It begins by showing that the area of a circle is equal to a triangle whose base is the circumference of the circle and whose height is the radius of the circle, and it concludes with a general, although slow, iterative method for approximating the value of π by inscribing and circumscribing a circle with a polygon. In *Measurement of the Circle*, this process ends with the 96-gon; however, Heron, in his *Measurements*, states that Archimedes produced an even more precise set of bounds, but there are sufficient difficulties with the numbers in the received manuscripts to have generated a good deal of discussion but little agreement.[48]

Archimedes, unlike Euclid, neither wrote for beginners nor made grand efforts towards systematization. Instead, he wrote monographs addressing specific sets of problems or areas of theory that were susceptible to mensuration but were difficult enough to demonstrate his considerable abilities. When the necessity for a rigorous presentation required that he include trivial material, we can sense his boredom and haste, but when he ventures into uncharted waters he is careful to present a thorough proof, occasionally giving two arguments for the same result.[49]

The most successful mathematician of the next generation was Apollonius; however, he appears to have followed more closely in the tradition of Euclid than that of Archimedes. The only indication that we have of Apollonius continuing an Archimedean project is Pappus' description in *Collection* II of Apollonius' system of large numbers. On the whole, however, Pappus' statement that Apollonius studied with the students of Euclid in Alexandria seems to fit well with his mathematical style.[50] His works are both systematic and comprehensive. He had an interest in laying the foundation of the areas in which he worked and in providing elementary texts for more advanced mathematics, such as the theory of conic sections

[47] For the history of the text, see W. Knorr, *Textual Studies in Ancient and Medieval Geometry* (Boston, MA: Birkhäuser, 1989), pp. 805–16.

[48] There are as many reconstructions for these values and their derivation as there are scholars who have studied the matter. W. Knorr provides a summary of previous work, followed by his own reconstruction. See W. Knorr, "Archimedes and the Measurement of the Circle: A New Interpretation," *Archive for History of Exact Sciences* 15 (1975), 115–40.

[49] Examples of the former are *Sphere and Cylinder* I 28–32 and *Spheroids and Conoids* 11, and of the latter *Sphere and Cylinder* II 8.

[50] It is also possible, however, that the idea that Apollonius studied with Euclid's students was simply an inference by Pappus, or someone else, based on the similarity of their mathematical style.

and the field of analysis. Like Euclid, he divided his texts according to the constructive approaches that were used: line and circles; conic sections; and general curves.[51]

Apollonius' most significant work was his *Conics*, the first four books of which provided "a training in the elements" of conic sections,[52] and the latter four exploring more advanced topics that were of use in geometric analysis.[53] The final book of the treatise, in which Apollonius showed how his theory could be used to solve interesting problems, has been lost since ancient times. The entire book was clearly written to be of use in geometrical analysis: in the introductions to the individual books, Apollonius explains how the material he covers will serve the student of analysis, and Pappus includes the *Conics* as the last work in his treatment of the field of analysis. The elementary books were apparently based on Euclid's work in conic theory, although Apollonius redefined the conic sections, derived the principle properties (*symptōmata*) by a construction from any initial cone, gave a more general treatment, and furnished a number of theorems that were not known to Euclid. The more advanced books are said to be an addition to the work of various predecessors, but again Apollonius points out how he clarified, simplified, and extended their contributions.

The treatise begins by showing the analogies that exist between various types of conic sections such as triangles, circles, parabolas, ellipses, and hyperbolas,[54] and how the fundamental properties of the three final conic sections can be derived from relationships between certain straight lines related to the original cone. After developing the basic theory of tangents, *Conics* I ends by providing constructions of the conic sections given certain straight lines that involve finding the cone, of which the sought conic is a section. *Conics* II then treats diameters and asymptotes, and ends with a series of problems involving tangents. Book III provides a sort of metrical theory of conic sections and their tangents that shows the invariance of certain areas constructed with parallels to the tangents, and which Apollonius regarded as essential for a complete solution to the three and four line locus problems. The final books deal with topics like the intersections of various conic sections, minimum and maximum lines, equal and similar sections, and the diameters of conic sections and the figures that contain them. In each case, we see that the material is related to certain types

[51] See Pappus, *Collection* IV 36, for a discussion of the division into three classes of problems, or domains of geometry.

[52] J. L. Heiberg, *Apollonii Pergaei quae graece exstant cum commentariis antiquis*, 2 vols. (Leipzig: Teubner, 1891–3), vol. 1, 2–4.

[53] For a discussion of the contents of the *Conics*, see H. G. Zeuthen, *Die Lehre von den Kegelshnitten in Altertum* (Copenhagen: Andr. Fred. Höst & Sohn, 1886) and M. Fried and S. Unguru, *Apollonius of Perga's* Conics (Leiden: Brill, 2001).

[54] Apollonius used the term "hyperbola" to refer to just one branch of the modern curve and referred to a pair of branches as "opposite sections." Nevertheless, he realized there was an important relationship between the two and in many ways treated them under a unified approach.

of problems in the analytic corpus, and Apollonius is careful to point out how these books will be useful to various aspects of geometrical analysis. The goal of the treatise, then, was to furnish theorems useful to analysis in such a way that the theory of conic sections would be laid on a solid foundation, analogous to that of elementary geometry, and so that the topics which were raised would be handled with a certain completeness.

Apollonius' tendency to treat things exhaustively is also seen in the other area to which he made extensive contributions, the field of analysis. Besides the *Conics*, Apollonius wrote six of the treatises that Pappus mentions as belonging to the "field of analysis." Of these, *Cutting off a Ratio* survives in an Arabic translation,[55] and the others can be plausibly reconstructed based on Pappus' description, at least as far as the general structure and the methods used.[56] In contrast to Archimedes' short works, the topics covered in these treatises do not seem to be inherently interesting. For example, *Cutting off a Ratio* exhaustively solves the problem of producing a line through a given point cutting two given lines such that the segments cut off between the intersections and two given points on the lines have a given ratio.[57] The treatment of this mundane problem is then carried out in twenty-one cases for the arrangements of the original given objects (*dispositions*), and eighty-seven cases for the arrangements of the line that solves the problem (*occurrences*). The entire treatment is analytical, giving analyzed propositions for both problems and theorems, and stating the total number of possible solutions and specifying their limits. Although it is possible that some of this material could have been of use in studying conic sections, recalling that Apollonius claims that the *Conics* itself was to be of use in analysis, it seems likely that his short works were meant to be training texts in analytical methods. By reading these works, one could develop a strategy for formulating an analytical approach, as well as considerable experience in understanding the structure of analytical arguments.

When we look at the scope of his work, it appears that Apollonius viewed himself as Euclid's successor. He produced a body of texts that could be studied as a course in analysis, dividing up the subject matter of these treatises according to the types of constructions involved and crowned by what he rightly believed would become the elements of conic theory. Nevertheless, Apollonius, naturally, brought his own personal interests to his work. In contrast to Archimedes, he did not spend much effort on questions of mensuration but focused on relative placement and arrangement, treating at considerable length topics such as the number and location of the intersections of curved lines, the disposition of maximal and minimal

[55] R. Rashed and H. Bellosta, *Apollonius de Perge, La section des droits selon des rapports* (Berlin: Walter de Gruyter, 2010).
[56] For summaries of the contents of these lost treatises Jones, *Book 7 of the Collection*, pp. 510–619.
[57] For an overview of *Cutting off a Ratio*, see K. Saito and N. Sidoli, "The Function of Diorism in Ancient Greek Analysis," *Historia Mathematica* 37 (2010), 595–608.

lines, the placement and arrangement of tangents, and so forth. Moreover, he took Euclid's inclination to completeness to new heights and treated a number of topics so exhaustively as to strike many modern readers as tedious.

The Hellenistic period was the most active for researches in pure geometry. As well as the three authors we have surveyed, a number of other important mathematicians were at work during this period, most of whom are now known to us only by name. Nevertheless, in the works of these others which do survive in some form, such as Diocles' *Burning Mirrors*, in Arabic,[58] or Hypsicles' treatment of a dodecahedron and the icosahedron inscribed in the same sphere, modified to make an *Elements* XIV,[59] we find explicit mention of the work of Euclid, Archimedes, and Apollonius and a clear indication of respect. Despite everything that we have lost, it seems clear, from the regard in which they were held by their contemporaries and successors, that these three mathematicians were the most active geometers of the Hellenistic period and that we are able to make a fair appraisal of their efforts.

ARITHMETICS AND ALGEBRAIC THINKING

There has been much debate as to the role that algebra and algebraic thinking have played in Greek mathematics. Although few scholars still argue for interpreting the Greek theory of the application of areas as "geometric algebra,"[60] there is still considerable evidence for algebraic modes of thought in arithmetical problem solving and the theories devoted to this activity. Nevertheless, depending on how we define our terms, it is still possible to argue that Greek activity falls short of algebra. If, on the one hand, we regard algebra as an explicit study of equations and their methods of solution as reduced to the arithmetic operations, then it is possible to argue that Greek work in this area was proto-algebraic. If, on the other hand, we regard algebra as the use of an explicit unknown, the application of various stated operations to equations, and the exposition, through examples, of methods that have various applications, then Greek mathematicians did, indeed, do algebra. Whatever the case, for this style of mathematics we can use the term *arithmetics*, which is not far from the Greek usage.

[58] G. J. Toomer, *Diocles, On Burning Mirrors* (New York: Springer, 1976) and R. Rashed, *Les catoptriciens grecs* (Paris: Les Belles Lettres, 2002).

[59] B. Vitrac and A. Djebbar, "Le Livre XIV des Élements d'Euclid : versions greques et arabe (première partie, seconde partie)," *SCIAMVS* 12 (2011), 29–158, 13 (2012), 3–158.

[60] The subject of "geometric algebra" was at the core of a debate on the historiography of Greek mathematics that erupted in the 1970s, following a paper by S. Unguru. For a collection of these papers see, S. Unguru (ed.), "Methodological Issues in the Historiography of Greek Mathematics," in J. Christianidis (ed.), *Classics in the History of Greek Mathematics* (Dordrecht: Kluwer, 2004), pp. 383–461.

Reflecting the various subscientific traditions of practical, and recreational, mathematics, some examples of problems in practical arithmetics, of the kind that would have been taught in schools, have survived on fragments of papyri and excerpted in compilations, especially in the Heronian corpus.[61] In the case of elementary arithmetics, which are clearly the products of an oral tradition, no justification of the method of solution is given and no operational, or algebraic, reasoning is explicitly invoked. Understanding and reapplying the methods for which these sources are evidence would have required verbal explanations provided by a teacher. Although most of the problems in these traditions are of the first degree, there are a number of second-degree problems as well that appear to have been solved by the application of a set of identities that must have formed part of the oral instruction.[62] Learning to solve these sorts of problems was probably part of the education of professionals such as builders and accountants, and may have been part of the general education of the literate. Because second-degree equations were not used for practical applications before the early modern period, their presence in our early sources indicates that, even in this practical tradition, the development of general problem-solving skills was a goal of mathematical practice and education.

Concurrent with this practical tradition, Greek mathematicians produced an advanced, theoretical form of arithmetics that went well beyond the practical needs of schoolteachers and engineers. Our knowledge of higher Greek arithmetics is due almost entirely to the *Arithmetics* of Diophantus. Although there were probably other works in this tradition, and Diophantus himself refers to another of his own works, the *Porisms*, none of these have survived.[63] Hence, our only knowledge of this tradition comes through the rather idiosyncratic work of Diophantus, who, like the Hellenistic geometers, was more interested in solving problems than in giving a general exposition of the methods he employed.

All that we may say with certainty about Diophantus is that he was associated with Alexandria and lived sometime between the middle of the second century BCE and the middle of the fourth century CE. These dates derive from the fact that he mentions Hypsicles (ca. mid-second to early first century BCE) in his *On Polygonal Numbers*, and is, in turn, mentioned by Theon of Alexandria (ca. late fourth century CE).[64]

[61] For a discussion of the evidence from papyri for the practical tradition of Greek mathematics, which shows evidence of its connection with Mesopotamian and Egyptian sources, see J. Friberg, *Unexpected Links Between Egyptian and Babylonian Mathematics* (Singapore: World Scientific, 2005), pp. 105–268.

[62] For texts and translations of second-degree problems in elementary arithmetics from the Greek tradition, see J. Sesiano, "An Early Form of Greek Algebra," *Centaurus* 40 (1998), 276–302.

[63] J. Christianidis raises the possibility of a lost work called *Elements of Arithmetics*, but the evidence is not certain. See. J. Christianidis, "Ἀριθμητικὴ Στοιχείωσις: un traité perdu de Diophante d'Alexandrie?," *Historia Mathematica* 18 (1991), 239–46.

[64] P. Tannery, *Diophanti Alexandrini opera omnia*, 2 vols. (Leibzig: Teubner, 1893), vol. 1, 470; vol. 2, 35. See also F. Acerbi, *Diofanto, De Polygonis numeris* (Pisa: Fabrizio Serra Editore, 2011).

Our knowledge of Diophantus' most famous work, the *Arithmetics*, is based on both the Greek and Arabic traditions.[65] The Arabic version, based on a single manuscript, consists of four books that are not found in the Greek version. Moreover, there are a number of differences of presentation between the Greek and Arabic books. While the Greek text employs symbolic abbreviations that are explained in the text, the Arabic version is fully rhetorical, conforming to the practice in earlier medieval algebraic texts in Arabic. Furthermore, the Arabic text contains far more elementary details, going through the resolution of simple equations, giving full computations and verifying that the numbers so found solve the original problem. Hence, it has been suggested that the Arabic books are derived from an otherwise lost *Commentary to the Arithmetics* produced by Hypatia (ca. late fourth to early fifth century CE).[66]

For the historian of mathematics, the *Arithmetics* is a difficult text because it contains considerable mathematics but few general discussions of method. Hence, it is susceptible to a broad range of interpretations. Nevertheless, the discovery of the Arabic books has made it clear that Diophantus teaches his methodology in much the same way as Apollonius in *Cutting off a Ratio*, by repeated exposure to the application of specific methods. The *Arithmetics* provides a general framework for approaching problems and then sets out many different types of problems that can be handled by these methods.[67] The treatise is structured as a series of problems, where *problem* has the sense that it generally does in Greek mathematics. The focus of the treatment is more on the production of certain numbers that meet the conditions of the problem and less on the resolution of the equation that finds these numbers, which is often alluded to only briefly. Most of the text of a problem in the *Arithmetics* is devoted to setting out the problem in the terms of the "arithmetical theory," while the solution of the equation that results from this is relatively short and often referred to with a cryptic reference to the operations involved, such as "let the common wanting [terms] be added and

The testimonials that are used to date Diophantus more precisely, to the middle of the third century, date from the eleventh century. They can be, and have been, called into question. For example, see W. Knorr, "Arthimêtikê stoicheiôsis: On Diophantus and Hero of Alexandria," *Historia Mathematica* 20 (1933), 180–92, who also argues that Diophantus wrote a work on the elements of arithmetics.

[65] The Greek editions are Tannery, *Diophantus* and A. Allard, *Diophante d'Alexandrie, Les Arithmetiques* (Tourpes: Fonds national de la recherche scientifique, 1980). The Arabic editions are J. Sesiano, *Books IV to VII of Diophantus'* Arithemetica (New York: Springer, 1982), and R. Rashed, *Diophante, Les Arithmétiques* (Paris: Belles Lettres, 1984).

[66] This commentary is attributed to Hypatia in the *Suda*; see Tannery, *Diophantus*, vol. 2, 36. It should be noted that this attribution can be questioned.

[67] For general accounts of the methods used by Diophantus, see T. L. Heath, *Diophantus of Alexandria* (Cambridge: Cambridge University Press, 1964), pp. 54–98, Sesiano, *Books IV to VII of Diophantus' Arithemtica*, pp. 6–7, A. Bernard and J. Christianidis, "A New Analytical Framework for the Understanding of Diophantus' *Arithmetica* I–III," *Archive for History of Exact Sciences* 66 (2012), 1–69, and J. Christianidis and J. Oaks, "Practicing Algebra in Late Antiquity: The Problem-Solving of Diophantus of Alexandria," *Historia Mathematica*, 40 (2013), 127–63.

like [terms taken] from like."[68] This constructive aspect of the project is expressed by the refrain that concludes many of the propositions, in which Diophantus points out that the produced numbers "make the problem" (*poiousi to problēma*). As in geometric problems, the goal is to produce a specific object.

There are sometimes necessary conditions that must be stipulated to make a rational solution possible. The only times Diophantus explicitly refers to such a condition he uses the term *prosdiorismos*, but this may be because in these particular cases he is actually introducing a further condition, so that the word should be understood to mean "further specification." In a number of cases he uses *diorizesthai*, a verbalization of the standard *diorismos*, for initial specifications.[69] Diophantus is only concerned to give one solution, despite the fact that he is aware that there are often more possible solutions.

In his introduction to the work, Diophantus explains the basics of his method of applying symbolic abbreviations and the use of an unknown number. He gives a symbol for squares, δ^{υ}, cubes, κ^{υ}, and units, $\overset{\circ}{\mu}$, and states that a number that is some multitude of units is "called an unknown (*alogos*) number, and its sign is ς."[70] There is also a sign introducing wanting terms, ⋔, which always follows the extant terms.[71] With these symbols, what we call a polynomial can be written as a series of signs and numerals. For example, ς$\overline{\varepsilon}\overset{\circ}{\mu}\,\overline{\alpha}$ ⋔ $\delta^{\upsilon}\overline{\kappa\beta}$ can be anachronistically transcribed as $5x + 1 - 22x^2$. Diophantus, however, did not have the concept of a polynomial as made up of a number of terms related by operations, but rather expressed a collection of various numbers of different kinds (*eidos*) of things, some of which might be in deficit, where certain stated operations could be carried out on equations involving these sets of things.[72] Perhaps we should think of ς$\overline{\varepsilon}\overset{\circ}{\mu}\,\overline{\alpha}$ ⋔ $\delta^{\upsilon}\overline{\kappa\beta}$ as expressing something like "$5x$, 1, less $22x^2$."

Using these symbols, in *Arithmetics* I, as a means of demonstrating the utility of his new methods, Diophantus treats the sort of simple problems whose solutions would have been taught in many schools and would have been well known to professionals who used arithmetic methods. He shows how to use the conditions of the problem to produce an equation involving an unknown number, which can then be solved to produce the sought numbers, where all of the numbers involved are assumed to be rational. In *Arithmetics* II and III, Diophantus begins to treat indeterminate problems that are handled by finding an equation involving the unknown number,

[68] Tannery, *Diophantus*, p. 90. See J. Christianidis, "The Way of Diophantus: Some Clarifications on Diophantus' Method of Solution," *Historia Mathematica* 34 (2007), 289–305.
[69] Acerbi discusses these issues on pp. 10–11 of F. Acerbi, "The Meaning of πλασματικόν in Diophantus' Arithmetica," *Archive for History of Exact Sciences* 63 (2009), 5–31.
[70] Allard, *Diophante*, p. 375.
[71] Wanting terms are mathematically equivalent to negative terms; however, ancient and medieval mathematicians did not seem to think of them in terms of operations.
[72] Christianidis and Oaks, "Practicing Algebra in Late Antiquity."

which must be set equal to a square. A number of the problems solved in *Arithmetics* II, namely 8–11 and 19, compose a toolbox used extensively in the rest of the treatise. In the Arabic books, IV–VII, Diophantus continues to show how the methods set out in *Arithmetics* II can be applied to more difficult problems, now involving higher powers of the unknown. The introduction to *Arithmetics* IV introduces a third operation that allows us to reduce an equation containing two powers of the unknown to one with a power equal to a number, although this operation has already been used a few times in the previous books.[73] This allows the treatment of equations with higher powers of the unknown. In the last three Greek books, the problems become yet more difficult, often involving the construction of an auxiliary problem because the solution to the original problem requires special conditions to be met in order to make a rational solution possible. The final Greek book treats problems involving the determination of metric properties of right triangles, given various conditions.

Greek arithmetics exhibits a number of features that we regard as essential to algebra, such as the use of an unknown and the application of algebraic operations. Nevertheless, a number of other important features of algebra are not present. For example, the solutions and types of equations, as such, do not seem to have been a subject of direct study. Hence, arithmetics appears to have formed a group of problem-solving methods that were used as techniques for the production of sought numbers but which did not focus on the equation as a subject of study.

COMBINATORICS

Although it used to be believed that the Greeks did no substantial work in combinatorics, it is now clear that this assessment was simply due to a loss of the primary sources. The evidence that Greek mathematicians worked in combinatorics is slight, either due to the vagaries of our transmission or to a relatively narrow range of their activity. Although we find a number of recursive arguments and proofs in various mathematical works, and some clear combinatorial statements in later authors, such as Pappus, it is not clear that any works devoted to combinatorial mathematics were written. What seems much more likely is that combinatorial methods were developed in the context of technical works on logic and then disappeared with the loss of these texts.[74]

The most certain testimony comes from Plutarch, who tells us that all "the arithmeticians" and particularly Hipparchus (ca. mid-second century

[73] Sesiano, *Books IV to VII of Diophantus' Arithemtica*, p. 284.

[74] It has also been argued that Archimedes did work in combinatorics, but the actual evidence for this is rather slight. See R. Netz, F. Acerbi, and N. Wilson, "Towards a Reconstruction of Archimedes' *Stomachion*," *SCIAMVS* 5 (2004), 67–99, and G. Morelli, "Lo Stomachion di Archimede nelle testimonianze antiche," *Bollettino di Storia delle Scienze Matematiche* 29 (2009), 181–206.

BCE) contradict the claim made by Chrysippus that the number of conjunctions produced through ten assertibles is greater than a million.[75] In fact, Hipparchus calculated that the number of conjoined assertables for affirmation is 103, 049 while that for negation is 310, 954.[76] It has been shown that these numbers are the tenth Schröder number and half the sum of the tenth and the eleventh Schröder numbers, which are the correct solutions to a well-defined problem involving bracketing ten assertions, and their negations.[77] Since the calculation of these numbers could almost certainly not have been carried out by brute force, it is clear that Hipparchus must have had at his disposal a body of combinatorial techniques that he could draw on in formulating and solving this problem. Moreover, a survey of the ancient literature on logic reveals a concern with combinatorial thinking, which goes back at least as far as Aristotle.[78] It seems likely that combinatorial methods were developed in the technical study of logic, which naturally gave rise to many combinatorial assertions and problems.

THE EXACT SCIENCES

The Greek exact sciences were produced by people who regarded themselves, and were regarded by their contemporaries, as mathematicians. There was no special institutional setting for people who did optics or astronomy. What divided optics from geometry, or observational astronomy from spherics, was the different traditions of texts in which these fields were transmitted. Nevertheless, ancient thinkers divided up the disciplinary space of the mathematical sciences in various ways, which probably reflected individual interests and tastes as much as institutional or educational realities. Plato famously justified the Pythagorean division of the mathematical sciences into arithmetic, geometry, music, and astronomy – what would later be known as the quadrivium – but reading the work of his contemporary Archytas, for example, makes it clear that an evaluation of the contents and methods of these early sciences must not be made on the basis of later categories.[79] Aristotle placed the mathematical sciences as one branch of the theoretical sciences, between theology and physics;[80] and within the mathematical sciences he asserted various subordinate

[75] Plutarch, *On the Contradictions of the Stoics*, 1047c–e.

[76] The manuscripts actually have 310,952 for the second number, but this appears to be an error.

[77] R. P. Stanley, "Hipparchus, Plutarch, Schröder, and Hough," *The American Mathematical Monthly* 104 (1997), 344–350, and L. Habsieger, M. Kazarian, and S. Lando, "On the Second Number of Plutarch," *The American Mathematical Monthly* 105 (1998), 446.

[78] The ancient evidence on combinatorics has been surveyed by F. Acerbi, who gives a full account of Hipparchus' approach. See F. Acerbi, "On the Shoulders of Hipparchus: A Reappraisal of Ancient Greek Combinatorics," *Archive for History of Exact Sciences* 57 (2003), 465–502.

[79] Plato, *Republic*, VII 520a–532c, and Archytas, frags. 1 and 3.

[80] For example, Aristotle, *Metaphysics*, 6.1, 1026a and 11.7, 1064a–b.

relationships, such as optics to geometry, and harmonics to arithmetic.[81] Geminus (ca. first century BCE), in some general work on mathematics, provides an extensive classification of the mathematical sciences that reflects the variety of texts that we still possess. His first division was into pure and applied mathematics, of which the applied branches form what we call the exact sciences. The branches of pure mathematics were arithmetic and geometry, while those of applied mathematics were mechanics, astronomy, optics, geodesy, harmonics (*kanonikē*), and calculation (*logistikē*).[82] Ptolemy, who considered himself a mathematician, seems to agree with Aristotle's division of the theoretical sciences, but he took his point of departure by asserting that only mathematics is capable of producing knowledge and placing it, in this sense, above the other two.[83] Moreover, in the *Harmonics*, he makes it clear that he regards mathematics as the study of beautiful things, of which the highest branches are astronomy and harmonics.[84] In this way, geometry and arithmetic are reduced to the position of "indisputable instruments."[85] As these examples make clear, there was no generally accepted classification of the mathematical sciences and the organization and importance of the different fields was highly influenced by the classifier's own interests.

An important area of mathematics that was necessary in the exact sciences was the application of numerical methods to problems of mensuration. Although these techniques were developed by a number of mathematicians in the Hellenistic and Imperial periods, most of the surviving texts of this type are preserved under the authorship of Heron. While the Heronian texts also preserve many problems in the subscientific traditions that would have formed part of the education of professionals, such as engineers and accountants who used mathematics in their work, one of Heron's primary goals appears to have been to produce academic mechanics and theories of mensuration following the pattern of some of the more established exact sciences such as optics or astronomy.[86] Heron blends the methods of the Hellenistic geometers with the interests and approaches of mechanics and other practical fields, and, unsurprisingly, he mentions Archimedes more than any other single author. In Heron's work, we find a number of practices that typify the Greek exact sciences. He uses the structures of mathematical propositions, such as theorems, problems, analysis, and synthesis, although their content and actual meaning is altered to accommodate the new subject

[81] R. D. McKirahan, "Aristotle's Subordinate Sciences," *British Journal for the History of Science* 11 (1978), 197–220.

[82] Geminus' classification is quoted by Proclus. See Friedlein, *Procli Diadochi*, pp. 38–42. See also J. Evans and J. L. Berggren, *Geminos' Introduction to the Phenomena* (Princeton, NJ: Princeton University Press, 2006), pp. 43–8.

[83] Ptolemy, *Almagest* I 1.

[84] Ptolemy, *Harmonics* III 3.

[85] I. Düring, *Die Harmonielehre des Klaudios Ptolemaios* (Gothenburg: Elanders, 1930) p. 94.

[86] K. Tybjerg, "Hero of Alexandria's Mechanical Geometry," *Aperion* 37 (2004), 29–56.

matter. He models objects in the physical world using geometric diagrams and then focuses his discussion on these diagrams.[87] He blends metrical, computational approaches with geometric, constructive ones, and uses arguments involving *givens* to provide a theoretical justification for computations.[88] Although Heron is one of the first authors we have who uses these techniques, and although his field was mechanics and the mathematics of engineering and architecture, the fact that similar techniques are found in the astronomical writings of Ptolemy indicates that they were probably general methods of the exact sciences developed in the middle or end of the Hellenistic period.

In terms of the development of Greek mathematics, the most significant exact science was astronomy. For the purposes of handling problems that arose in astronomy, Greek mathematicians developed a number of mathematical techniques that are only found in astronomical texts but became an essential part of the canonical education of mathematicians in the late ancient and medieval periods.

For the purposes of ancient astronomy, timekeeping, and cosmology, Greek mathematicians developed a branch of applied geometry known as *spherics*. Spherics was the development of a theoretical geometry of the sphere and its application to problems of spherical astronomy, that is, the study of the motion of the fixed stars and the sun, regarded as located at some point on the ecliptic. Some works in spherics were probably produced in the classical period, by Eudoxus and others; however, as usual, the first texts in this field that survive are from the beginning of the Hellenistic period. We have works in pure spherical astronomy by Autolycus and Euclid, but it is clear from the way these works are presented that there was also a body of knowledge of spherical geometry on which they could draw.[89]

Towards the end of the Hellenistic period, Theodosius produced a new edition of elementary spherical geometry, his *Spherics*, which was so successful that previous versions of this material only maintained historical value and were eventually lost.[90] Theodosius' *Spherics* is in three books, the first of which is purely geometrical and the second two of which deal with topics applicable to spherical geometry, but still expressed in an almost purely geometrical idiom. The first book treats the properties of lesser circles and great circles of a sphere that are analogous with the properties of the chords and diameters of a circle in *Elements* III. The second book explores those properties of lesser circles and great circles of a sphere that are analogous with

[87] See, for example, Heron, *Dioptra* 35.

[88] See, for example, Heron, *Measurements* I 7 and 8; and *Measurements* I 10, or *Measurements* III.

[89] J. Mogenet, *Autolycus de Pitane* (Louvain: Bibliothèque de l'Université, 1950) and Euclid, *Phaenomena et Scripta musica*, in Menge (ed.), *Euclidis Opera omnia*, vol. 5 (1916).

[90] See J. L. Heiberg, *Theodosius Tripolites Sphaerica* (Berlin: Weidmann, 1927) and C. Czinczeheim, "Édition, traduction et commentaire des Sphériques de Théodose" (PhD thesis, Paris IV, 2000).

those of circles and lines in *Elements* III, which leads to theory of tangency, and theorems dealing with the relationships between great circles and sets of parallel lesser circles. Although still expressed in almost purely geometrical terms, the book ends with a number of theorems of largely astronomical interest, having to do with circles that can model the horizon, the equator, and the always visible, and always invisible, circles. The third book deals with what we would call the transformation of coordinates, or the projection of points of one great circle onto another, and concludes with theorems that can be interpreted as concerning the rising- and setting-times of arcs of the ecliptic, again without naming these objects explicitly, as Euclid had done, over 200 years earlier, in his *Phenomena*.[91]

Around the end of the first century CE, Menelaus took a new approach to spherical geometry that focused on the geometry of great circles on a sphere and was orientated towards Hellenistic developments in trigonometry, enabling him to produce the theoretical basis for a more elegant, metrical spherical astronomy. Only fragments of the Greek text of Menelaus' *Spherics* survive,[92] but we have a number of Arabic, Latin, and Hebrew versions. Our knowledge of this work is somewhat tentative because only late, heavily modified, versions of the text have been edited or studied in depth; nevertheless, we may sketch out the general approach and trajectory of the treatise.[93] Book I develops a theory of the congruency of spherical triangles modeled on the first part of *Elements* I, but including the congruency of pairs of triangles having three equal angles.[94] This is then followed by a run of theorems that involve inequalities that can be asserted in a given spherical triangle in which the sum of two sides is equal to, greater than, or less than a semicircle. Book II, then, develops a theory of bundles of lesser circles having some angle with a great circle, acting as an analogy with the theory of parallels in *Elements* I. The versatility of this approach is then demonstrated with three theorems on the intersections of great circles with sets of parallel lesser circles, which elegantly do the work of six long theorems in Theodosius' *Spherics* III. Book III begins with the theorem passed down under Menelaus' name, but which may have been already well-known in his time,[95] which gives a compound ratio between the chords that subtend various parts of a convex quadrilateral made up of great circles. This is

[91] For an overview of Theodosius' approach see J. L. Berggren, "The Relation of Greek Spherics to Early Greek Astronomy," in A. Bowen (ed.), *Science and Philosophy in Classical Greece* (New York: Garland, 1991), pp. 227–48. The text of the *Phenomena* is in Menge (ed.), *Euclidis Opera omnia*, vol. 5. There is an English translation by J. L. Berggren and R. S. D. Thomas: *Euclid's Phaenomena: A Translation and Study of a Hellenistic Treatise in Spherical Astronomy* (New York: Garland, 1996).

[92] The Greek fragments are collected on pages 22–5 of A. Björnbo, "Studien über Menelaos' *Sphärik*," *Abhandlungen zur Geschichte der mathematischen Wissenshaften* 14 (1902).

[93] For Ibn ʻIrāq's version, see M. Krause, *Die Spärik von Menelaos aus Alexandrien in der Verbesserung von Abū Naṣr Manṣūr b. ʻAlī b. ʻIrāq* (Berlin: Weidmann, 1936).

[94] Not even the book divisions are consistent in the various versions of the text; see Krause, *Die Spärik von Menelaos*, pp. 8–9. The following numbers are those in Krause's edition.

[95] N. Sidoli, "The Sector Theorem Attributed to Menelaus," *SCIAMVS* 7 (2006), 43–79.

then combined with Menelaus' theory of spherical triangles and his theory of bundles of parallel lesser circles to provide tools for developing a fully metrical spherical astronomy.[96] In this way, Menelaus combined the Hellenistic trigonometric methods with his new theory of the spherical triangle to produce a new approach to spherical trigonometry.

A major area of mathematical development in astronomical texts was trigonometry, which in antiquity always literally involved the mensuration of elements of a triangle. We may talk of Greek trigonometry in three main stages.[97] The first stage, which is proto-trigonometric, is attested in the astronomical writings of Aristarchus and Archimedes, *On the Sizes and Distances of the Sun and the Moon* and the *Sand Reckoner*, respectively. In these texts, the authors implicitly rely on a pair of ratio inequalities that hold between a ratio of corresponding sides and a ratio of corresponding angles of a pair of triangles. Both Aristarchus and Archimedes seem to assume these inequalities as well-known lemmas, but we have a number of proofs of their validity in later authors.[98] Using these inequalities, Greek mathematicians were able to derive fairly precise upper and lower bounds on sought angles or lengths. There were, however, a number of difficulties involved in this method. The computations were cumbersome, and the more lengthy the computation, the more accuracy would be lost in regard to the term to be bounded. Furthermore, these inequalities are only sufficiently precise for small angles, and if large angles were involved they would produce substantially inaccurate results. Although these methods were sufficient for the laborious calculation of a few chosen values it is difficult to imagine that an accurate, predictive astronomy could have been established using such techniques.

The next stage in the development of Greek trigonometry, although not well documented in our sources, appears to have been directly spurred by the desire to produce a predictive astronomy based on geometric models. During the Hellenistic period, Greek astronomers came into possession of Babylonian sources that showed them that it was possible to produce an accurate, predictive astronomy, and provided them with observation reports and numerical parameters to put this project into effect.[99] By at least the time of Hipparchus, in the second century BCE, the goal was to produce a numerically predictive astronomy on the basis of geometric models. In order to meet this goal, Greek

[96] R. Nadal, A. Taha, and P. Pinel, "Le contenu astronomique des *Sphériques* de Ménélaos," *Archive of History of Exact Sciences* 58 (2004), 381–436.

[97] For an overview of the history of Greek trigonometry, see G. Van Brummelen, *The Mathematics of the Heavens and the Earth: The Early History of Trigonometry* (Princeton, NJ: Princeton University Press, 2009), pp. 20–90.

[98] W. Knorr, "Ancient Versions of Two Trigonometric Lemmas," *Classical Quarterly* 35 (1985), 361–91.

[99] A. Jones, "The Adaptation of Babylonian Methods in Greek Numerical Astronomy," *Isis* 82 (1991), 441–53, and A. Jones, "On Babylonian Astronomy and its Greek Metamorphoses," in F. J. Ragep and S. Ragep (eds.), *Tradition, Transmission, Transformation* (Leiden: Brill, 1996) pp. 139–55.

mathematicians developed tabular trigonometry, which consisted of a group of theorems about right triangles and tables that allowed a length to be determined as a function of a given angle, and conversely. Tabular trigonometry allowed them to use observed angles and period relations to calculate the numerical parameters of assumed geometric models. Although it is not certain that Hipparchus was the first to produce a chord table, there is evidence that Hipparchus tabulated the chords subtending angles at 71/2° intervals, from 0° to 180°, for a circle with $d = 6{,}875$.[100] Using such chord tables, mathematicians, such as Hipparchus, Diodorus (ca. mid-first century BCE), and Menelaus, were able to solve computational problems in plane and spherical trigonometry by treating them with forms that we would write using the sine function, since $\mathrm{Crd}(2\alpha) = 2\, sin\, \alpha$.

The final stage of Greek trigonometry is found in the writings of Claudius Ptolemy. In his *Almagest* I 10–13, Ptolemy sets out the mathematics necessary for the trigonometric methods he will use. He begins by showing how a chord table of 1/2° intervals, on a circle with $d = 120$, could be calculated using a number of geometrically derived formulas and an approximation of the chord of 1/2° using a lemma similar to that at the basis of the proto-trigonometric tradition. This serves as a justification for his more precise chord table, although Ptolemy does not actually say that he used these methods to derive it.[101] This more precise chord table was the computational tool at the foundation of Ptolemy's plane and spherical trigonometry, as found in works such as the *Almagest*, the *Planisphere*, and the *Analemma*.[102] For spherical trigonometry, Ptolemy restricted himself to the theorem based on a convex quadrilateral of great circle arcs, known as the Menelaus theorem, avoiding any use of the more powerful spherical-trigonometric methods that Menelaus had developed.

A final mathematical tool of the exact sciences should be mentioned: numerical tables. Tables presumably entered Greek mathematical practice along with predictive astronomy from Babylonian sources during the Hellenistic period. Although mathematicians such as Hipparchus, Diodorus, and Menelaus must have used tables, again, our earliest texts that contain tables are by Ptolemy, particularly his *Almagest, Analemma*, and *Handy Tables*. Despite the fact that tables were used long before the function

100 The argument for the details of Hipparchus' chord table is given in G. Toomer, "The Chord Table of Hipparchus and the Early History of Greek Trigonometry," *Centaurus* 18 (1973), 6–28, and D. Duke, "Hipparchus' Eclipse Trios and Early Trigonometry," *Archive for History of Exact Science* 56 (2005), 427–33.

101 For an argument that Ptolemy actually used interpolation methods in constructing his chord table see G. Van Brummelen, "Mathematical Tables in Ptolemy's Almagest" (PhD thesis, Simon Fraser University, 1993), pp. 46–71.

102 J. L. Heiberg (ed.), *Claudii Ptolemaei opera omnia*, 2 vols. (Leipzig: Teubner, 1898–1903), N. Sidoli and J. L. Berggren, "The Arabic Version of Ptolemy's *Planisphere*," *SCIAMVS* 8 (2007), 37–139, and D. R. Edwards, "Ptolemy's Περί ἀναλήμματος" (PhD thesis, Brown University, 1984).

concept became explicit, they exhibit relationships that are similar to certain modern conceptions of a function.[103] For example, tables were treated as general relations between members of two, or more, sets of numbers. Moreover, the algorithms that describe how to use the tables make it clear that they describe a computational rule that maps a single member of the domain to a single member of the codomain. These proto-functions were used in a number of ways to handle the metrical aspects of geometrical objects and moving components of a geometrical model.

In Ptolemy's writings, tables are sets of numerical values that correspond to lengths and arcs in the geometric models from which they are derived. At least in principle, they are produced by direct derivation from geometric objects with assumed numeric values. We can understand the tables themselves as a numerical representation of the underlying model, which is geometric. The tables are then used, either by Ptolemy or by the reader, to provide an evaluation of specific numerical values that represent the underlying model. In the *Almagest*, mathematical tables are a component of Ptolemy's goal of producing a deductively organized description of the cosmos, presented in an essentially single argument.[104] Indeed, Ptolemy makes a number of explicit assertions that the structure of the tables in the *Almagest* should exhibit both the true nature of the phenomena in question and have a suitable correspondence with the mathematical models.[105] For Ptolemy, a table, like a mathematical theorem, is both a presentation of acquired mathematical knowledge and a tool for producing new mathematical results.

EXPOSITORS AND COMMENTATORS

Our knowledge of the substantial texts of Greek mathematics comes through the filter of the scholarship of the mathematicians of late antiquity, most of whom were associated with schools of philosophy and regarded mathematics as an important part of a broader cultural and educational project. We have seen that the texts of the earlier periods were edited and commented upon by these mathematical scholars, and this process acted as an informal process of selection, in so far as texts which did not receive attention had a dramatically reduced chance of being passed down.

These late-ancient scholars were primarily responsible for creating the image of theoretical mathematics that was transmitted to the various cultures around the Mediterranean in the medieval and early modern periods.

103 O. Pedersen, "Logistics and the Theory of Functions," *Archives Internationales d'Histoire des Sciences* 24 (1974), 29–50.
104 N. Sidoli, "Mathematical Tables in Ptolemy's Almagest," *Historia Mathematica* 41 (2014), 13–37.
105 See, for example, Heiberg (ed.), Claudii Ptolemaei opera omnia, vol. 1, 208, 251.

Through their teaching and scholarship, they established various canons of the great works of the past, arranged courses of study through select topics, reinforced a sound and lasting architecture by shoring up arguments and making justifications explicit, and, finally, secured their place in this tradition by intermingling their work with that of their predecessors and situating the whole project in contemporary modes of philosophic discourse.

One of the most impressive of these scholars was Pappus of Alexandria, who was a competent mathematician and a gifted teacher, who made important strides to associate mathematics with areas of interest in philosophy by constantly arguing for the relevance of mathematics to other aspects of intellectual life. Pappus worked in many areas of the exact sciences, wrote commentaries on canonical works, such as the *Elements* and the *Almagest*,[106] and produced a series of short studies that were later gathered together into the *Mathematical Collection*.[107] It is clear from Pappus' writing, that he was part of an extended community of mathematicians and students who had regard for his work and interest in his teaching.

Pappus' *Collection*, although incomplete, is indispensable for our understanding of both the early history of Greek mathematics and the main trends of mathematical thought in late antiquity.[108] Book II, which is fragmentary, discusses a system of large numbers attributed to Apollonius. Books III and IV present a number of topics in advanced geometry framed in philosophical discussions that show, among other things, how mathematical argument can be of relevance to philosophical issues.[109] Here we find discussions of the difference between theorems and problems; the division of geometry into linear, solid, and curvilinear methods; treatments of mechanical and geometrical curves; examples of problem solving through analysis and synthesis; discussions of geometrical paradoxes; and examples of solutions to three classical problems: squaring a circle; duplicating a cube; and trisecting an angle. Pappus' style is to situate his own work in a wealth of material drawn from historically famous mathematicians, to mix geometrical and mechanical approaches, and to combine numerical examples with pure geometry. Book V deals with isoperimetric and isovolumetric figures, again situating the discussion in the context of past work, for example, that of

[106] His commentary on *Elements* X survives in an Arabic translation; see W. Thomson and G. Junge, *The Commentary of Pappus on Book X of Euclid's Elements* (Cambridge, MA: Harvard University Press, 1930). What survives of his commentary on the *Almagest* was edited by A. Rome (ed.), *Commentaires de Pappus et Théon d'Alexandrie sur l'Almageste*, 3 vols. (Rome: Biblioteca Apostolica Vaticana, 1931–43), vol. 1: *Pappus, Commentaire sur les livres 5 et 6* (1931).

[107] Hultsch, *Pappi Alexandrini Collectionis*, Jones, *Book 7 of the Collection*, and H. Sefrin-Weis, *Pappus of Alexandria, Book 4 of the Collection* (London: Springer, 2010).

[108] For a study of Pappus in his intellectual context, see S. Cuomo, *Pappus of Alexandria and the Mathematics of Late Antiquity* (Cambridge: Cambridge University Press, 2000).

[109] A. Bernard, "Sophistic Aspects of Pappus' *Collection*," *Archive for History of Exact Sciences* 57 (2003), 93–150.

Archimedes, Zenodorus (ca. third to second century BCE), and Theodosius.[110] The next two books were clearly produced as part of Pappus' teaching activities, and they are mainly made up of lemmas to canonical works. Book VI deals with what Pappus calls "the field of astronomy," which came to be known as the *Little Astronomy*.[111] It organizes and discusses the treatises that students should master before studying Ptolemy's *Almagest*.[112] Book VII covers "the field of analysis," which consisted of a group of works by authors such as Aristaeus, Euclid, and Apollonius.[113] It begins with a general account of analysis, discusses the overall content of each work, and provides numerous lemmas for individual propositions, many of which give justifications for common practices of the Hellenistic geometers, such as operations on ratios and the use of compound ratios. Book VIII treats theoretical mechanics in much the same vein as Heron, whom Pappus refers to a number of times. It models various machines geometrically and sets out the mathematical theory of certain practical constructions.[114]

The other mathematical scholars of the late ancient period were also involved in teaching and expounding the classics, and hence mostly worked through the medium of commentaries. Theon of Alexandria, in the fourth century, edited works by Euclid and wrote commentaries to Ptolemy's *Almagest* and *Handy Tables*.[115] Hypatia, his daughter, collaborated with her father on various projects and wrote commentaries to Apollonius and Diophantus.[116] In the following century, Proclus of Athens wrote a commentary on the *Elements*. Eutocius of Ascalon, in the sixth century, edited works by Archimedes and Apollonius, and wrote commentaries to them.

This work was a continuation of a tradition of commentating and editing that began in the Imperial period. The scholars of this later period paid particular attention to issues of logical completeness, formal structure, and

110 Another important text for this material is the anonymous introduction to Ptolemy's *Almagest*. See pp. 92–196 of F. Acerbi, N. Vinel, and B. Vitrac, "Les *Prolégomènes à l'Almageste*, Introduction générale – Parties I–III," *SCIAMVS* 11 (2010), 53–210.

111 Hultsch, *Pappi Alexandrini Collectionis*, p. 474.

112 For Pappus' treatment of Theodosius' *Spherics*, see M. Malpangotto, "Sul commento di Pappo d'Alessandria alle 'Sferiche' di Teodosio," *Bollettino di Storia delle Scienze Matematiche* 23 (2003), 121–48.

113 Jones, *Book 7 of the Collection*, p. 83.

114 The Arabic version of Book VIII, which has not yet been thoroughly studied, contains a number of topics not found in the Greek version. For an example, see D. E. P Jackson, "Towards a Resolution of the Problem of τὰ ἑνὶ διαστήματι γραφόμενα in Pappus' Collection Book VIII," *The Classical Quarterly* 30 (1980), 523–33.

115 Rome (ed.), *Commentaires de Pappus et Théon d'Alexandrie sur l'Almageste*, J. Mogenet and A. Tihon, *Le "Grand Commentaire" de Théon d'Alexandrie aux Tables Faciles de Ptolémée, livre I* (Rome: BAV, 1985), A. Tihon, *Le "Grand Commentaire" de Théon d'Alexandrie aux Tables Faciles de Ptolémée, livres II, III*, 2 vols. (Rome: BAV, 1991–9), and A. Tihon, *Le "Petit Commentaire" de Théon d'Alexandrie aux Tables Faciles de Ptolémée* (Rome: BAV, 1979).

116 For Hypatia's life and work, see M. Dzielska, *Hypatia of Alexandria* (Cambridge, MA: Harvard University Press, 1995).

readability. They produced fuller texts with more explicit arguments; wrote auxiliary lemmas; introduced internal references to other parts of the canon; restructured the treatises and individual elements of the text; added introductions and conclusions; advocated explicit classifications; rewrote theories from new perspectives; and summarized long works for the purposes of study.[117]

All of this was part of a broad trend, begun in the Imperial period by authors such as Geminus, Heron, and Ptolemy, to incorporate the mathematical sciences into the philosophical tradition.[118] Although in the Classical and early Hellenistic periods, philosophers showed interest in mathematical approaches, there is little indication that mathematicians had a similar regard for philosophy. The mathematicians of the late ancient period, however, were concerned that mathematics be part of an education in philosophy and rhetoric.[119] Their texts show a combination of modes of thought from the traditions of pure mathematics with those from the various exact sciences, and a mixture of philosophical concerns with mathematical issues. Their project, situated as it was in the philosophical schools, argued both explicitly and implicitly for the value of mathematics to philosophy. The final stages of Greek mathematical practice furnished the image of Greek mathematics that remains with us to this day: that of a fully explicit, interconnected, literary product.

CONCLUSION

Theoretical Greek mathematics underwent considerable change and development. What began as a leisure activity for elite scholars was then applied in solving interesting problems in the natural sciences, and it became of practical use to a wide range of scholars and professionals. The resulting combination of the theoretical and practical traditions was then institutionalized in courses in mathematical sciences at the philosophical schools. Throughout this long transition, however, Greek mathematics, as a type of intellectual activity, was primarily defined by the cohesion of certain traditions of practices and texts, not by the social position, or institutional setting, of the practitioners themselves.

[117] R. Netz, "Deuteronomic Texts: Late Antiquity and the History of Mathematics," *Revue d'histoire des mathématiques* 4 (1998), 261–88.

[118] For a discussion of the philosophy of Ptolemy, see L. Taub, *Ptolemy's Universe* (Chicago, IL: Open Court, 1993), and J. Feke and A. Jones, "Ptolemy," in L. P. Gerson (ed.), *The Cambridge History of Philosophy in Late Antiquity* (Cambridge: Cambridge University Press, 2010), pp. 197–209.

[119] A. Bernard, "Comment definer la nature des textes mathématiques de l'antiquité grecque tardive? Proposition de réforme de la notion de 'textes deutéronomiques'," *Revue d'histoire des mathématiques* 9 (2003), 131–73.

19

GRECO-ROMAN ASTRONOMY AND ASTROLOGY

Alexander Jones

Astronomy constituted one of the more long-standing and complex scientific traditions of the Greco-Roman world, and its scope is thus difficult to define exactly. In this chapter, we will be concerned with systematic descriptions, observations, explanations, and predictions relating to the visible heavenly bodies (fixed stars, Sun, Moon, and planets) and their real or supposed connections with terrestrial phenomena including human lives and society; the organization of the chapter will be primarily thematic rather than chronological. Even under the Roman Empire astronomy was a Greek science par excellence, designated by the interchangeable names *astronomia* ("management of stars") and *astrologia* ("study of stars").[1] Both terms were applicable from the outset both to speculations and investigations of the heavenly bodies in their own right and to systematic predictions of mundane phenomena based on the heavenly bodies, such as astral weather forecasting and astrology in the proper sense.

THE EVIDENCE

Texts are our most important sources of information about Greek and Greco-Roman astronomy. They may be divided, in the first instance, into texts that were preserved through a tradition of manuscript copying from antiquity into the Middle Ages and sometimes beyond, either preserving their original language (usually Greek) or translated into another language (usually Arabic or Latin), and texts written on ancient artifacts. Those that have come down to us through the medieval manuscript traditions, which we may call transmitted texts for short, are mostly writings that individual,

[1] Wolfgang Hübner, *Die Begriffe "Astrologie" und "Astronomie" in der Antike: Wortgeschichte und Wissenschaftssystematik. Mit einer Hypothese zum Terminus "Quadrivium"* (Stuttgart: Franz Steiner Verlag, 1990).

named authors produced for a specialist or wide readership; their survival depended not only on accidental circumstances but also on the deliberate choices that later scribes and book owners made with respect to which works among the currently available literature to copy or collect. They thus represent an elite among the range of astronomical texts that subsisted in their day according to some criterion: the most authoritative, or comprehensive, or practically useful, or elementary. The other category, which we can loosely call archaeologically recovered texts, were favored by chance and by local conditions (both natural and social) and so are more representative of what was *typical* in their times and places. They are generally very fragmentarily preserved.

The corpus of transmitted astronomical texts is modest in bulk: comparable to the transmitted corpus of Greek mathematics, vastly smaller than the transmitted corpus of Greek medicine. Like the medical corpus, however, it is dominated by a single author of the second century CE: Galen's slightly older contemporary, Ptolemy. We have what appears to be close to the whole of Ptolemy's astronomical writings, mostly in a good condition of textual preservation in Greek, supplemented by parts extant in Arabic and Latin translations of at least decent fidelity. These are works of an advanced, technical character. There is very little else from earlier periods approaching this level, most notably Aristarchus of Samos' *On Sizes and Distances of Sun and Moon* (first half of the third century BCE) and an assortment of deductive mathematical treatments of topics in spherical astronomy by authors ranging chronologically from Autolycus (late fourth century BCE) to Menelaus of Alexandria (ca. 100 CE). The period following Ptolemy is chiefly represented by didactic commentaries on Ptolemy's works.

We also have a few elementary presentations of astronomical topics directed to a nonspecialist readership, including Aratus' poem *Phenomena* (early third century BCE), Archimedes' *Sand Reckoner* (mid- to late third century BCE), and Geminus' *Introduction to the Phenomena* (mid-first century BCE). The popularity of Aratus' poem gave rise to a copious body of commentaries, among them Hipparchus' *Exposition of the Phenomena of Aratus and Eudoxus* (mid- to late second century BCE), the only one of Hipparchus' many astronomical works that has survived.

Discussions of astronomy and references to astronomers occur in many transmitted texts of other genres, including works by historians, biographers, medical and mechanical writers, and encyclopedists. Of particular importance are passages in Plato, Aristotle, and various Platonist philosophers ranging from the unknown author of the *Epinomis* ascribed to Plato, through Theon of Smyrna (early second century CE), and on to Proclus (fifth century CE), Simplicius (sixth century CE), and other Neoplatonists of late antiquity. Additionally, authors from later periods, especially Byzantine and Islamic writers, sometimes report information about Greek astronomy, presumably derived from sources that have since become lost. Greek origins may also sometimes be inferred for methods and concepts in texts from the astronomical traditions of India.

Turning to the archaeologically recovered texts, a first category comprises inscriptions. We have only a few public astronomical inscriptions on stone, dating from the second and first centuries BCE, but they are all valuable for the historian. Time-keeping instruments, such as sundials, portable sundials, and components of water clocks, sometimes bear inscriptions, typically brief labels and dedications. The texts inscribed on the mechanical Antikythera Mechanism (probably early first century BCE) are extensive and very important for our understanding of Hellenistic planetary and eclipse theory.[2]

Remains of Greek astronomical texts are much more numerous among the papyrus manuscripts from Ptolemaic and Roman Egypt.[3] Those of Ptolemaic date are few, but include the only substantially complete text extant in this medium, the so-called *Ars Eudoxi* or "Eudoxus Papyrus" (PPar. 1 = PLouvre N 2329+2388, ca. 200 BCE).[4] From the Roman period we have between one and two hundred fragments of astronomical tables and instructions for the use of tables, most of which were made for the sake of astrological practice, and a very small number of fragments of theoretical works.

There are also late Ptolemaic and Roman astronomical papyri written in the Egyptian language, primarily in demotic script, which provide evidence for circulation of knowledge between Greek-speaking and Egyptian-speaking astronomers. Babylonian astronomical tablets of the last several centuries BCE are significant for the historian of Greek astronomy primarily as witnesses to a tradition that influenced the Greek science, but the presence of Greek elements has occasionally been conjectured in cuneiform tablets too.

The principal physical remains of Greco-Roman astronomy, other than inscriptions and manuscripts, are sundials and other time-keeping devices.[5] Over five hundred sundials are known, many of them fragmentary, crudely executed, or of repetitive types, though a minority are of considerable individual interest beyond their collective significance for the diffusion of this variety of astronomical technology. Water clocks, though driven by mechanical means, displayed time according to an astronomically defined

2 *The Inscriptions of the Antikythera Mechanism* (special issue), *Almagest*, 7.1 (2016).

3 For an inventory of Greek and Egyptian astronomical papyri published up to 1999 see Alexander Jones, *Astronomical Papyri from Oxyrhynchus*, 2 vols. in 1 (Philadelphia, PA: American Philosophical Society, 1999), vol. 1, 301–7, in addition to the many Greek papyri published in that work.

4 Friedrich Blass, "Eudoxi ars astronomica qualis in charta Aegyptiaca superest," Kiel, Universitätsprogramm Sommersemester 1887. Reprinted in *Zeitschrift für Papyrologie und Epigraphik* 115 (1997), 79–101.

5 For an ongoing database of sundials see the Topoi Ancient Sundials website, http://repository.edition-topoi.org/collection/BSDP.

framework, represented for us by surviving fragments of two display dials. The single most remarkable astronomical *object* from Greco-Roman antiquity, however, is the already-mentioned Antikythera Mechanism (National Archeological Museum, Athens, X 15087), an incompletely preserved but largely reconstructable gearwork device that simulated simultaneous chronological cycles and astronomical phenomena through pointers revolving around dials.[6] The only other extant example of such a device is a simple and rather crude gearwork system built into a portable sundial made around 500 CE (London Science Museum inv. 1983–1393).

The evidence for Greco-Roman astrology deserves to be considered separately from that for astronomy. Our primary source of information is again transmitted astrological texts, and these include books with named authors, among them Ptolemy. But in the surviving literature of this field Ptolemy is a comparatively *early* author. Our only transmitted astrological works composed before Ptolemy's *Tetrabiblos* are Manilius' *Astronomica* (early first century CE), a Latin didactic poem popularizing the science for a general readership, and the poem of Dorotheus of Sidon (mid-first century CE), meant for practitioners, which we have only in an Arabic translation of a lost Pahlavī translation. Other treatises with identified authors from after Ptolemy include those of Vettius Valens (late second century CE), Firmicus Maternus (first half of the fourth century CE, in Latin), and Hephaestion of Thebes (ca. 400 CE). Besides such more or less coherent works, however, the Byzantine manuscript tradition presents us with a vast quantity of anonymous astrological texts and texts that bear dubious or pseudepigraphic attributions, for the most part difficult to date or trace to original sources, and hence constituting a great challenge to the historian of the subject. The majority of this indiscriminate material is still either unpublished or available in unsatisfactory editions such as those in the appendices of the *Catalogus Codicum Astrologorum Graecorum* (1898–1953).

Many papyri and ostraca from the first century BCE on contain astrological texts. Much of this material is similar in character to the transmitted astrological literature, consisting of didactic and reference texts. They are an underutilized resource, and numerous identified fragments still lack even an edition. This is even truer of the many astrological papyri in demotic Egyptian, which have the potential to illuminate the role of Egypt in the formation of Greek astrology as well as later relations between the two linguistic groups in Egypt.

Personal horoscopes constitute a special and important category of astrological papyri and ostraca.[7] Unlike the horoscopes intermittently embedded

[6] Alexander Jones, *A Portable Cosmos* (New York: Oxford University Press, 2017).

[7] For an inventory of ancient horoscopes, including both original documents and transmitted texts, see Stephan Heilen, *Hadriani Genitura: Die astrologischen Fragmente des Antigonos von Nikaia*, 2 vols. (Berlin: de Gruyter, 2015), vol. 1, 204–333.

in the transmitted literature, which were chosen if not actually fabricated to illustrate particular theoretical or methodological points, these are the unmediated horoscope documents pertaining to real people, and thus they inform us in various ways about the chronological and social patterns of astrology's popularity in Egypt. A small number of archaeologically recovered horoscopic documents are also known from outside Egypt in the form of graffiti, inscriptions, and inscribed gems and jewellery.

One example of a public astrological inscription exists, the so-called horoscope frieze of Antiochus I of Commagene (first century BCE) at Nemrud Dag, a royal monument displaying visually what must have been understood to be an astrologically significant configuration of the heavenly bodies, though it is not a complete horoscope. Other inscribed objects relating to astrology include zodiac boards used by astrologers to display astrological configurations to their clients and peg-board inscriptions that allowed one to track significant time cycles including the astrological seven-day week.[8]

Lastly, images relating to astronomy and astrology in ancient visual art are witnesses to the extent to which these sciences were in the public eye, and also sometimes to religious or political appropriation of this imagery. The most common entities to be portrayed were celestial spheres, provided with the principal circles in spherical astronomy, such as the equator, tropics, and ecliptic, or with constellation figures, or with both; sundials; zodiacs; and figures of individual constellations, especially those of the zodiac.

FILLING IN THE GAPS

The remains of Greco-Roman astral science, as of practically all aspects of Greco-Roman civilization, are a tiny fraction of what once existed, and are very unevely distributed. Consider the items of direct evidence according to chronological period, that is, those that are archaeologically recovered and that actually date from the period as well as the transmitted texts that were originally written during the period. Relative to other intervals, the second century CE is extremely well served across the board, not just through Ptolemy's surviving writings, but also through the papyri and such astronomically or astrologically engaged authors as Theon of Smyrna, Galen, and Vettius Valens. From the Hellenistic period, especially the last two centuries BCE, a variety of texts and artifacts relate to the public face of astronomy,

[8] James Evans, "Images of Time and Cosmic Connection," in Alexander Jones (ed.), *Time and Cosmos in Greco-Roman Antiquity* (Princeton, NJ: Princeton University Press, 2016), pp. 143–69; Daryn Lehoux, *Astronomy, Weather, and Calendars in the Ancient World: Parapegmata and Related Texts in Classical and Near Eastern Societies* (Cambridge: Cambridge University Press, 2007).

occasionally also offering glimpses of its more technical aspects. For late antiquity (fourth century CE and after), our direct sources give a good picture of how mathematical astronomy was taught and of the prominent place occupied by astronomical topics in the teachings and polemics of the Neoplatonist schools. We have nothing from before the Hellenistic period except some treatments of astronomical topics by nonexpert authors, and little from the first and third centuries CE other than papyrus fragments of astronomical tables, horoscopes, and astrological texts.

The patchiness of documentation can be compared to the situation of the astral sciences in other ancient cultures. For Mesopotamia, the evidence is almost exclusively textual and archaeologically recovered, and is dense for celestial divination and related observational practices in the Neo-Assyrian Empire (roughly the first half of the seventh century BCE), observational practices and period-based prediction in Babylon (from the eighth century BCE through the first century CE), and mathematical predictive astronomy and astrology in Babylon and Uruk (from the late fifth through the first century BCE).[9] In each of these bodies of texts, it is easier to discern diachronic continuities than developments. For China, transmitted texts, especially the "standard histories," are the survivors of a narrower and more centralized process of selection and editing than their Greco-Roman counterparts, while the number of archaeologically recovered texts pertaining to astronomy, excavated from tombs, is still small though increasing.[10] Given these materials, historians of Mesopotamian and early Chinese science have tended to work more on interpreting and contextualizing the sources within their own periods than on seeking a story of origins and evolution.

A distinctive feature of Greco-Roman writing, however, is the frequency with which authors make what we may loosely call historical statements. These range from testimonia about specific individuals and texts to broader narratives of past developments. To take a few instances pertaining to individuals, all our information concerning Meton of Athens (second half of the fifth century BCE) consists of testimonia in an assortment of source texts, perhaps none of which predates the second century BCE; the heliocentric hypothesis of Aristarchus of Samos (first half of the third century BCE) is known only by reports, of which the earliest and most important is in Archimedes' *Sand Reckoner*; and for Hipparchus' researches on solar and lunar theory (mid-second century BCE) we depend chiefly on passages in Ptolemy's *Almagest*. Examples of narratives are Geminus' account of the development of Greek calendar cycles (*Introduction to the Phenomena* 8) and Ptolemy's of the search for accurate period relations for the Moon's motions (*Almagest* 4.2).

9 See chapter 4 in the present volume by John M. Steele.
10 See chapter 30 in the present volume by Christopher Cullen.

In addition to exploiting testimonia and other historical statements, historians of Greco-Roman astronomy employ conjectural reconstructions to fill in the gaps in our knowledge between and beyond our sources. Reconstruction of what probably happened, or (as is often claimed) what *must* have happened, when we lack direct evidence or credible testimony, is common practice in all branches of the historiography of antiquity; but the history of astronomy lends itself to special kinds of reconstruction that take as their basis quantitative or otherwise technical content in the sources. One of these is the hypothesis that the occurrence of the same element – say a numerical parameter, theoretical structure, or computational method – in two disparate contexts implies a transmission of this knowledge from one context to the other, or from an antecedent context to both. Another is the hypothesis that if a conjectural route of calculation or scientific derivation reproduces an attested element, this is how the element was in fact obtained. Both these kinds of more or less abstract reconstruction may be pursued in more human terms, for example by conjecturing who must have been responsible for a transmission or what must have motivated an astronomer to make a calculation in a certain way.

Words like "must" and "doubtless" in historical writing are signals that a claim is not self-evidently true; and obviously one cannot naively rely on testimony and reconstructions as having the same factual solidity as the extant texts and artifacts. Any author can inadvertently or deliberately distort or falsify information or skew its interpretation through omissions and emphases; alternative derivations can often be hypothesized for numerical parameters and other elements. Some historians have argued that no ancient author's historical statements should be adduced as evidence unless they meet strict criteria, for example that the author quotes or paraphrases *extant* sources accurately, or that the statements in question are concordant with immediate evidence from the period to which they refer; and it is sometimes denied that reconstructions can ever function as evidence.[11] The logical consequence of adhering to highly severe principles of evidence is that large areas of the history of Greco-Roman astronomy become off limits, for example the great part of Hipparchus' work on solar and lunar theory. If, on the other hand, while excluding most indirect evidence from consideration, one equates the state of knowledge and practices in a period such as the second century BCE with what the direct, contemporary sources attest to, one risks missing ideas and methods that, though important, were not widely diffused at the time. In short, too much credulity leads to exaggerating the sophistication of astronomy at any period and assigning to developments dates that are too early, whereas too much skepticism dates them too late, though the risks are asymmetrical since the direct evidence for

[11] These positions have been advocated with notable rigor by Alan C. Bowen and Bernard Goldstein, together and separately, in many publications since 1988.

each period provides a limiting minimum. This chapter attempts a moderately skeptical approach, based on a flexible – and admittedly subjective – case-by-case appraisal of the reliability of sources and the cogency of reconstructions.

CALENDRICS AND CHRONOLOGY

The Greeks had numerous regional calendars, having in common a basic lunisolar structure of twelve named lunar months occurring in a fixed order, with occasional consecutive repetitions of a month ("intercalary months") to keep the calendar year roughly in a fixed relation to the natural seasons, thus ensuring that religious festivals tied to the calendar would not stray too far from appropriate stages of the agricultural year. While the operation of such a calendar must have originally involved some observational element in deciding both when the new months began and when an intercalary month was needed, nothing is known about who made the observations or how.

The responsibility for managing the calendar must have rested with different authorities from one city to another, and other considerations sometimes took precedence over astronomical phenomena. In Athens, for example, inscriptions often record two dates for the same event, one *kat' archonta* ("according to the archon," an annually appointed magistrate), the other *kata theon* ("according to the god"), where the *kata theon* date apparently was closely aligned with the Moon's phases while the *kat' archonta* date, the official one, typically lagged behind by several days. In fact a month was frequently prolonged beyond its natural duration by inserting intercalary days bearing the same name (in one documented instance as many as eight in a row), perhaps, for example, to prevent a festival from coinciding with a military operation, and later in the year days would be skipped over to compensate.

Notwithstanding such caprices, regularities arose in the operation of Greek calendars that reflect competent astronomical input. For several centuries, starting in the mid-fourth century BCE, Athenian calendar years having intercalary months almost invariably followed a repeating nineteen-year cycle that spread them as evenly as possible while keeping the year's beginning at the first new Moon following the summer solstice. This pattern could have arisen in two ways: either the date of the summer solstice was reliably observed through this whole interval, or a fixed 19-year cycle of intercalations was adopted that also effectively determined the solstice date for each year. Inscriptional evidence from the second century BCE indicates that the calendars of several other cities followed the same pattern, so that these calendars were locked in synchronization with the Athenian calendar.[12]

Later sources (including Diodorus, *Bibliotheca Historica* 12.36, IMilet inv. 84 + inv. 1604, and pseudo-Theophrastus, *De Signis* 4) credit the institution of a nineteen-year cycle (*enneakaidekaetēris*) to Meton of Athens, with the inauguration of this cycle dated to the thirteenth of the Athenian month Skirophorion in the year when Apseudes was Archon Eponymos (i.e. 433/432 BCE). As some sources make explicit, this was understood as the date of the summer solstice in the final month of Apseudes' archon year. According to a scholion to Aristophanes' *Birds* (997), Meton erected a *hêliotropion* on the Pnyx in Athens in that year, likely meaning an instrument for determining the solstice date, while Ptolemy (*Almagest* 3.1) presents the solstice date – expressed as an Egyptian calendar date equivalent to June 27, 432 BCE – as an observation made by "those around" Meton and Euctemon. The historical reality behind these disparate testimonia is beyond certain recovery; but it is noteworthy that the assumptions that the interval between successive summer solstices – the tropical year – is 12 7/19 lunar months (as is implied by a nineteen-year lunisolar calendar cycle) and that the solstice fell upon the thirteenth day of the last month of a particular year suffices to determine all solstice dates and the pattern of intercalations in the nineteen-year cycle. The cycle followed by the Athenian calendar from about 375 BCE through 128 BCE or later appears to be the one derivable from Meton's solstice date.

Since the Babylonian calendar's intercalations had been regulated by a nineteen-year cycle since the beginning of the fifth century BCE, it is tempting to suppose that Meton somehow learned of it directly or indirectly from Near Eastern sources. On the other hand such a short periodicity could have been noticeable from two or three decades of reasonably competent solstice observations, and the circumstance that it was also known in early Chinese astronomy should discourage us from ruling out independent discovery in Greece.

The little information we have about Greek methods of deciding which months should be full (thirty days) and which hollow (twenty-nine days) suggests that astronomy played no role beyond providing the ratio of full to hollow months. Geminus, *Introduction to the Phenomena* 8, says that a common practice was simple alternation of full and hollow months – which however would have eventually needed recalibration through extra full months since the mean synodic month (the month of the Moon's phases) is greater than 29.5 days. Both Geminus and the inscriptions of the Antikythera Mechanism delineate patterns of evenly distributed "omitted"

[12] John D. Morgan, "The Calendar and the Chronology of Athens" (abstract), *American Journal of Archaeology* 100 (1996), 395; Alexander Jones, "The Miletos Inscription on Calendar Cycles: IMilet Inv. 84 + Inv. 1604," *Zeitschrift für Papyrologie und Epigraphik* 198 (2016), 113–27.

(*exairesimoi*) days, that is, day numbers skipped over in lunar months that are all nominally thirty days long, through a nineteen-year cycle on the assumption that nineteen calendar years comprise 6940 days. Whether these day-skipping schemes were employed by any Greek city is doubtful.

A grouping of four nineteen-year cycles into a seventy-six-year cycle comprising 27759 days is associated with Callippus of Cyzicus (fourth century BCE); this was merely a reconciliation of the nineteen-year cycle, understood as an equation of solar years and lunar months, with a 365 1/4 day year, and thus reflected no new astronomical knowledge. On the other hand, the proposal by Hipparchus (second century BCE), known from Ptolemy, *Almagest* 3.1, to cluster four seventy-six-year cycles into a 304-year cycle comprising 111035 days was intended to accommodate Hipparchus' estimate of the tropical year as about 1/300 day less than 365 1/4 days. The inscription IMilet inv. 84 + inv. 1604 (ca. 100 BCE) refers to the seventy-six-year cycle, possibly reflecting an incorporation of this cycle in an adjustment of the calendar of Miletus.

For certain specifically astronomical applications, Greek astronomers employed artificial calendars and chronological frameworks that had greater regularity than civil calendars and thus facilitated calculations involving time intervals. The Callippic Periods, likely instituted by Callippus himself and continuing in use as late as the first century CE, were serially numbered seventy-six-year cycles of an idealized version of the Athenian calendar, such that Period 1 began with the new Moon that almost immediately followed the summer solstice of 330 BCE. More short-lived was the "Calendar of Dionysius," employed in several anonymous third century BCE observations of the planets reported subsequently in Ptolemy's *Almagest*. The Dionysian years were solar, numbered serially such that year 1 began with the summer solstice of 285 BCE, and divided into twelve months named after the sign of the zodiac traversed by the Sun during each month. The Egyptian calendar in its pre-Roman form with constant years of 365 days was the most enduring "astronomical" calendar, often in conjunction with a count of years from an era such as the Era Nabonassar (year 1 beginning in 747 BCE, employed in Ptolemy's *Almagest*) or the Era Philip (year 1 beginning in 324 BCE, employed in Ptolemy's *Handy Tables*).

STARS, CONSTELLATIONS, AND THE NATURAL YEAR

The *Phenomena* of Eudoxus of Cnidos (early fourth century BCE) – a work surviving only in extensive quotations by Hipparchus, but whose contents were substantially versified in the extant poem of the same title by Aratus of Soli (early third century BCE) – described in terms of images and relative placements a system of forty-eight constellations. Allowing for minor variations and additions, this became established as the conventional system, portrayed on celestial globes (such as the globe held aloft by the Farnese

Atlas, a second-century CE Roman copy of a Hellenistic statue, now in the National Archeological Museum of Naples), and eventually cataloged star-by-star with numerical coordinates and magnitudes in Ptolemy's *Almagest* (Books 7–8). Before Eudoxus, the history of the Greek constellations is poorly documented. One would be hard pressed to find a dozen constellations or named stars in the surviving literature from before the fourth century. The stellar phenomena that the parapegma tradition – to be discussed presently – ascribes to Euctemon (late fifth century BCE) pertain to fifteen stars and constellations. More than half of Eudoxus' forty-eight constellations have no known earlier mention.

Among Eudoxus' constellations, the twelve that constituted the standard Greek zodiac appear for the first time, with the exception of Scorpius which had already been among Euctemon's parapegmatic repertoire. (Pliny the Elder, *Natural History* 2.31, asserts that the astronomer Cleostratus of Tenedos introduced the zodiacal constellations, but the accuracy of his rather obscurely expressed claim is doubtful and Cleostratus' date highly uncertain.) The zodiac originated in Babylonian astronomy, though several of its constituent constellations acquired different names and different iconographies in the Greek version. Our imperfect knowledge of the broader system of Babylonian constellations makes it difficult to assess the extent of its influence on the Eudoxian system.

Most early citations of stars and constellations use the dates of their first and last risings and settings, along with solstices and equinoxes, as a means of marking stages of the natural year (as distinct from the lunisolar calendar year, whose seasonal alignment was variable), for example the appropriate seasons for agricultural activities in Hesiod's *Works and Days* (ca. 700 BCE) or meteorological conditions related to patterns of illness in a local community in the Hippocratic *Epidemics* (fifth and fourth centuries BCE). Only a few easily identifiable asterisms were in common use as date markers, chiefly Arcturus, the Pleiades, and Sirius. The genre of astronomical text known as *parapēgma* ("beside-pegging") extended this idea to a denser set of annual dates determined by a larger number of asterisms.

The many *parapēgmas* surviving complete or in fragments exhibit some variation in their contents, but central features of the tradition may be illustrated from two closely similar examples: the so-called "Geminus *parapēgma*" which is preserved through the medieval manuscript tradition as a kind of appendix to Geminus' *Introduction to the Phenomena*, and the second-century BCE inscription from Miletus IMilet inv. 456A + 456C + 456D + 456N.[13] Both texts presume a 365-day solar year and subdivide it into twelve "zodiacal months" corresponding to the intervals of around thirty days during which the Sun traverses each of the twelve signs of

[13] Daryn Lehoux, "The Parapegma Fragments from Miletus," *Zeitschrift für Papyrologie und Epigraphik* 152 (2005), 125–40.

the zodiac – not the zodiacal constellations as described by Eudoxus, but equal divisions of the zodiacal circle, another originally Babylonian concept. In the Geminus *parapēgma* individual days are numbered serially within the zodiacal months, whereas in the inscription each day of the zodiacal month is represented by a drilled hole that was evidently meant to hold a movable peg that indicated which was the current day, thus accounting for the etymology of *parapēgma*.

In the Miletus *parapēgma* the holes corresponding to some days have no text inscribed next to them; similarly, the Geminus *parapēgma* simply omits some day numbers. When there is text, it consists of statements relating to risings and settings of stars or to weather associated with that day, such as "eagle rises in the evening according to Euctemon" or "for Eudoxus, rain." The attributions are to individual Greek astronomers including Euctemon, Meton, Eudoxus, and Callippus, or to non-Greeks such as "Kallaneus of the Indians" and "the Egyptians," and more than one authority may be cited for the same phenomenon, either on the same day or on distinct but nearby days. Other *parapēgmas*, such as IMilet inv. 456B and the one inscribed on the Antikythera Mechanism, have no authorities or redundant statements.

The earliest surviving *parapēgma*, the papyrus PHib. 1.27, dates from around 300 BCE, and although it uses the Egyptian calendar as its chronological framework and incorporates information such as Egyptian religious festivals that are atypical of the *parapēgma* tradition, it bears signs of having been adapted from a source that followed the zodiacal year framework. The Euctemon data, which presumably are from the second half of the fifth century BCE and functioned, as it were, as the foundations of the tradition, probably antedated Greek knowledge of the zodiac; in fact the earliest bodies of parapegmatic data may have been expressed differently from the extant documents, perhaps as lists of intervals of days separating phenomena.

The *parapēgma* tradition illustrates the principle that an abundance of evidence does not guarantee answers to our most basic questions. We do not know the extent to which either the astronomical or the meteorological statements in the *parapēgmas* were derived from direct observations or from some kind of theoretical processing of empirical data, except in the special case of Ptolemy's *Phaseis*, which explicitly confronts dates of weather phenomena ostensibly observed by past authorities with dates of stellar risings and settings that Ptolemy has computed by applying a mathematical model of stellar visibility to his own star catalog. The paths of transmission and adaptation leading from the originators of the data to the documents that we possess remain obscure. Most crucially, we do not know what the correlation of stellar and solar phenomena and meteorological phenomena meant for the authors of the *parapēgmas*. In his chapter on the subject (*Introduction to the Phenomena* 17), Geminus asserts that the belief was prevalent in his time that the stellar phenomena were the physical causes of weather changes, but he maintains (on dubious evidence, it must be said) that the original compilers of *parapēgmas* recorded the stellar phenomena merely to provide a chronological framework for observed annual patterns of weather that was independent of the local calendars. The format of the extant *parapēgmas* rendered actual observation of risings and settings of constellations superfluous for weather forecasting, because the text structure itself established an easily counted cycle of days that simultaneously predicted both the stellar phenomena and the weather.

GEOMETRICAL AND PHYSICAL MODELING

The calendrical and astrometeorological astronomy described in the foregoing sections seems originally to have been independent of efforts to explain astronomical phenomena and to draw inferences about the structure of the cosmos. These efforts first appear in philosophical contexts. Thus two preserved lines of the poem of Parmenides (early fifth century BCE) probably reflect the assumption that the Moon receives its light from the Sun, and in the following generation Anaximander and Empedocles evidently applied this assumption to explaining the Moon's phases and eclipses.[14] Corollaries – not necessarily realized immediately – include the sphericity of the Moon (from the appearance of the phases) and of the Earth (from the outline of the Earth's shadow on the eclipsed Moon) as well as that the Sun is much more distant than the Moon from the Earth (because half-Moon phase occurs when the Moon's elongation from the Sun is close to a right angle).

The origins of the geocentric "Two-Sphere Model," a cosmological framework consisting of a stationary, spherical Earth concentric with a vast celestial sphere that carries the visible heavenly bodies while performing a daily revolution on a fixed axis, are more obscure. By the early fourth century the Two-Sphere Model was well established as the normal underpinning of more elaborated cosmologies that sought to explain not just the daily risings and settings of all the stars but also the slower independent movements of the "wanderers" (*planētes*), i.e. the Sun, Moon, and the five planets known in antiquity. The "Myth of Er" in Plato's *Republic* delineates a system of relative distances of the heavenly bodies from the Earth, based on their apparent rates of revolution – fixed stars outermost, then Saturn, Jupiter, Mars, Mercury, Venus, Sun, and Moon – while expressing it in such riddling terms that it must have been familiar to his intended readers (616c–617b). In Plato's later *Timaeus*, the celestial equator representing the daily revolution of the heavens and the ecliptic representing the path of the Sun, Moon, and planets are

[14] Daniel Graham, *Science Before Socrates: Parmenides, Anaxagoras, and the New Astronomy* (Oxford: Oxford University Press, 2013), pp. 87–96.

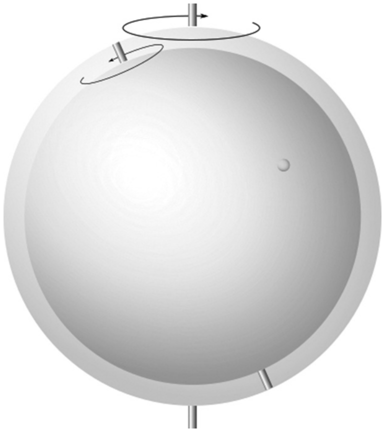

Figure 19.1. Homocentric spheres. The outer shell revolves around the celestial poles with the rate of the daily rotation of the stars around the Earth; the inner shell revolves more slowly around the poles of the ecliptic, which have fixed positions on the sphere of the stars. This minimal model would suffice for a heavenly body revolving along the ecliptic with constant speed relative to the stars.

designated figuratively as the circles "of the same" and "of the different" (36c–d).

Aristotle (*Metaphysics* 1073b17–1074a5) and Simplicius (*Commentary on Aristotle De Caelo*)[15] are our principal informants on the earliest known sophisticated geometrical models of Greek astronomy, the so-called homocentric spheres of Eudoxus. As a younger contemporary of Eudoxus, Aristotle's witness is invaluable notwithstanding the brevity and nontechnical character of his descriptions; on the other hand Simplicius wrote nearly a millennium after Eudoxus and by his own account depended on intermediate sources, and his historical reliability both in points of detail and wholesale is highly controversial. The fundamental principle of Eudoxus' models is not in doubt: a complex apparent motion of a heavenly body arises as the aggregate of the uniform revolutions of a succession of concentric spherical surfaces, such that the poles of the sphere bearing the visible body – presumably on the equator of its revolution – are fixed on a second revolving sphere, those of the second sphere are fixed on a third revolving sphere, and (in the case of the planets) those of the third sphere are fixed on a fourth. For each of the seven bodies, the final sphere was equated with the sphere bearing the fixed stars in their daily revolution, while the penultimate sphere's poles were those of the ecliptic (Figure 19.1).

Aristotle says practically nothing about the phenomena that Eudoxus' models were meant to account for, while Simplicius' remarks on this topic

[15] J. L. Heiberg (ed.), *Simplicii in Aristotelis De Caelo* (Commentaria in Aristotelem Graeca 7; Berlin: consilio et auctoritate Academiae Litterarum Regiae Borussicae, 1894) pp. 491–8.

may be influenced by hindsight. But even without invoking Simplicius, we can deduce from Aristotle's outline of the models themselves that Eudoxus assumed that both the Sun and the Moon appear to travel among the fixed stars along circles inclined to the ecliptic and with shifting nodes, and that the motions of the planets relative to the stars are compounded of a uniform longitudinal motion and an oscillatory one (with the planet's synodic period according to Simplicius) involving both periodic latitudinal (north–south) deviations from the ecliptic and periodic variations in longitudinal speed, which under suitable conditions can entail retrogradations, the intervals when planets are observed to reverse their normal eastward direction of movement relative to the stars. Aristotle converts Eudoxus' models into a single physical and quasi-mechanical organism by hypothesizing "unwinding" spheres that successively cancel out the spins of each model's spheres so that the poles of the outermost sphere belonging to Jupiter are fixed on the innermost unwinding sphere of Saturn and so forth down to the Moon.

Callippus (mid-fourth century BCE) is reported by Aristotle as having modified some of Eudoxus' models by hypothesizing additional spheres. In the case of the Sun and Moon, Simplicius claims that the intention was to account for solar anomaly (as manifested in inequality of the intervals between solstices and equinoxes) and lunar anomaly. For a century or more following Callippus, reliable information about Greek astronomical modeling is almost completely lacking, and so we are ignorant of how and when it came about that homocentric modeling was supplanted by modeling based on combinations of uniform circular motions that are *not* all assumed to be concentric. The two fundamental models of this type are the simple eccenter and the simple epicycle. In a simple eccentric model (Figure 19.2), a heavenly body travels with uniform speed along a circular orbit (the *eccenter*) that encloses the Earth and whose center is either at a fixed displacement from the Earth's center or revolves uniformly along a circle concentric with the Earth. In a simple epicyclic model (Figure 19.3), the heavenly body travels with uniform speed along a circular path (the *epicycle*) that does not enclose the Earth but whose center revolves along a circular orbit (the *deferent*) concentric with the Earth. For any specific epicyclic model there exists an eccenter model that generates the identical path through space for the visible body in identical times, and vice versa.

According to Ptolemy, *Almagest* 12.1, the geometer Apollonius of Perge (late third or early second century BCE), in company with other unnamed mathematicians, proved a theorem determining where a heavenly body must be located with respect to an epicyclic or eccentric model at one of its stationary points, such that the quantities on which this location depends are the ratio of the radii of the two circles in the model and the ratio of the periodicities of the two circular motions. This is the earliest evidence for epicyclic or eccentric models being treated mathematically, and it might be interpreted as indicating an interest in establishing that such models could

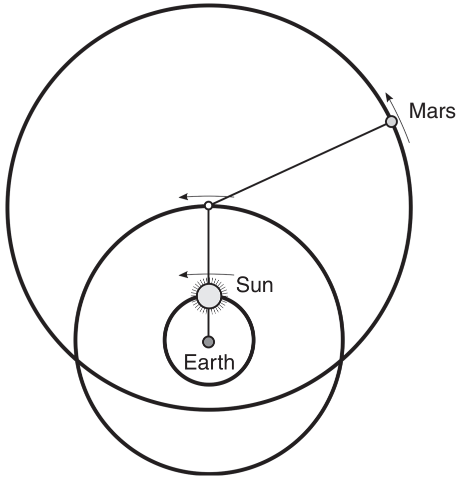

Figure 19.2. A simple eccentric model for Mars. The center of the eccenter is always aligned with the Sun.

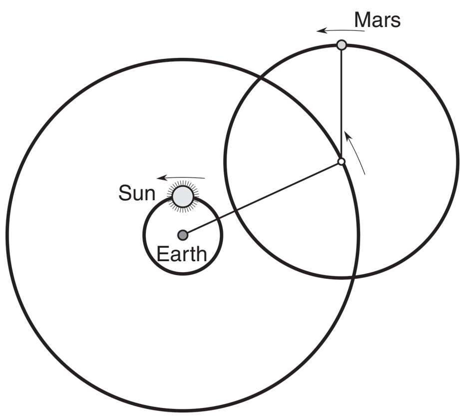

Figure 19.3. A simple epicyclic model for Mars. The direction from the center of the epicycle to the planet is always parallel to the direction from the Earth to the Sun.

indeed account for the basic phenomena of the planets' synodic cycles, without necessarily involving numerical calculation of the sizes and durations of each planet's retrogradations. Hipparchus – again, our principal informant is Ptolemy, *Almagest* 3.4 and 4.11 – carried out numerical deductions of the ratios of radii in epicyclic and eccentric models for the Sun and Moon from specific dated observations or at least from ostensibly empirical data. On the other hand, Ptolemy reports elsewhere (*Almagest* 9.2) that Hipparchus demonstrated that observations of the planets were inconsistent with models such as the simple eccenters or epicycles that would require the planets to have invariable synodic cycles.

With some justification, Ptolemy presents his own models for the planets in *Almagest* 9–11 as the successful response to Hipparchus' critique of the models current in his time. The planetary models (as well as Ptolemy's lunar model) employ not only a combination of the two model types, by making the deferent bearing the epicycle an eccenter, but also a definition of uniform circular motion that is more flexible than simply equating it with constant speed along an orbit or constant angular velocity as seen from the circle's center: a revolution is now considered uniform so long as it is seen as constant angular velocity from *some* defined point (the *equant*) inside the circle which need not be the center. In his models for Venus, Mars, Jupiter, and Saturn, the equant is the point such that the deferent's center is the midpoint between the equant and the center of the Earth, but there are also equants defined differently in his models for Mercury and the Moon.

The eccenter-with-epicycle model is very likely to have been developed at least as early as the first century CE; at any rate Pliny the Elder, *Natural History* 2.63–4, speaks of the planets as having *apsides* (in this context perhaps meaning "wheel-rims") that appear to be eccentric deferents. Ptolemy himself has often been credited with introducing the equant, but it is not clear that he makes any such claim in the *Almagest*, and certain algorithms for computing planets' positions in medieval Sanskrit astronomical treatises have been interpreted as derivations from Greek equant models antedating Ptolemy.[16]

Through the entire documented history of Greek astronomical modeling, we find philosophers and philosophically conscious writers contrasting an "astronomical" or "mathematical" approach, concerned with spatial and quantitative aspects of celestial motions and bodies, with a "physical" approach appealing to presumed causes residing in the essences or natures of the bodies, for example their sphericity, their eternal unchangeability, and their innate tendencies to move only in particular ways. One reason that Eudoxus' homocentric spheres were congenial to Aristotle was that they allowed for a physical cosmology in which an inner sphere composed of the always irregularly changing four elements earth, water, air, and fire, whose natural tendency to motion is towards the Earth's center, is enclosed in an outer spherical shell composed of the unchanging fifth element ether, whose natural tendency is to move circularly without changing its distance from the center.

Geminus, allegedly retailing Poesidonius' views in a passage from a lost work quoted by Simplicius (*Commentary on Aristotle Physics*),[17] stresses how the astronomical approach to explaining phenomena relies on hypotheses,

[16] Dennis Duke, "The Equant in India: The Mathematical Basis of Ancient Indian Planetary Models," *Archive for History of Exact Sciences* 59 (2005), 563–76.

[17] H. Diels (ed.), *Simplicii in Aristotelis Physica* (Commentaria in Aristotelem Graeca 10; Berlin: consilio et auctoritate Academiae Litterarum Regiae Borussicae, 1895), pp. 291–2.

such as epicycles or eccenters, and on principles taken from physics, such as the simplicity and uniformity of the heavenly bodies' motions. In the absence of physical considerations – such as the "common sense" knowledge that the Earth that we stand on is stationary – an astronomer would have no criterion for preferring one or the other of two models that generate the same appearances. The notion that at some level one cannot discriminate between optically equivalent models seems to underly Aristarchus' hypothesizing, in a lost "book of hypotheses" as reported by Archimedes (*Sand Reckoner* 4–5), that "the fixed stars and the Sun remain unmoved, while the Earth revolves around the Sun along the circumference of a circle, the Sun being situated in the middle of the course." For Archimedes, Aristarchus' heliocentric hypothesis was worth mentioning because it would entail a cosmos vastly larger than Greek geocentric cosmologies typically assumed; otherwise it left scarce traces in the surviving literature.

For Theon of Smyrna (*Mathematics Useful for Reading Plato*),[18] physics provides grounds for preferring epicycles over eccenters because an epicycle can have a physical meaning as the equator of a revolving sphere of etherial matter embedded in a revolving spherical shell of ether concentric with the Earth, whereas from the astronomers' geometrical standpoint the two types of model are entirely interchangeable. Ptolemy, however, overturns the conventional privileging of physics over mathematical astronomy by declaring that mathematics is superior to both physics and theology because its objects are uniquely knowable and subject to regularity (*Almagest* 1.1). The physical, three-dimensional embodiments of his models offered in his *Planetary Hypotheses*, which are similar in principle to Theon's but much more complex, are offered as plausibilities, not certainties.

COSMIC DIMENSIONS

The fact that half-Moon phase occurs a quarter of the way through a lunar month, when the Moon's elongation from the Sun is roughly a right angle, implies that the Sun is many times further than the Moon from the Earth. Since the apparent disks of the two luminaries are about the same size, the ratio of their actual diameters is approximately the same as the ratio of their distances. Aristarchus, *On Sizes and Distances* prop. 7, calculated that the ratio of the Sun's distance to the Moon's was between 20:1 and 18:1, assuming an elongation of 87° at half-Moon phase. If Archimedes' statement (*Sand Reckoner* 9) that Eudoxus asserted that the Sun's diameter is nine times the Moon's is correct, it is likely that Eudoxus too estimated this ratio using some consideration about the phases. Otherwise, the homocentric sphere

[18] E. Hiller (ed.), *Theonis Smyrnaei philosophi platonici expositio rerum mathematicarum ad legendum Platonem utilium* (Leipzig: Teubner, 1878) pp. 181–8.

models, since they involve no eccentricity, could be imagined as tightly packed as one pleases, though Aristotle does indicate that Eudoxus posited *some* order of distances for the heavenly bodies, which Callippus retained. The usual ordering principle in Greek astronomy was that the fixed stars are the furthest heavenly bodies from the Earth, and that the order of the remaining bodies from outermost to innermost is from slowest to fastest in terms of their sidereal periods of revolution, i.e. starting with Saturn, Jupiter, and Mars, and ending with the Moon. This criterion does not decide an order for the Sun, Venus, and Mercury, and in fact several arrangements of these three are attested.

At a qualitative level, it appears to have been broadly accepted by the third century BCE that, with the exception of the Moon, the heavenly bodies were sufficiently far from the Earth to justify the principle that "the Earth has the relation of a point and center" to the spherical shells carrying them. In the initial hypotheses of his *On Sizes and Distances*, Aristarchus even asserts this with respect to the sphere of the Moon, though it would follow from Aristarchus' data and his deductions from them that the Moon's distance is only on the order of fifteen to twenty Earth-radii. Certain empirical considerations, for example variations in magnitudes of a solar eclipse observed at different locations near the same meridian, or comparison of the apparent disk sizes of the Moon and the Earth's shadow at a lunar eclipse, could provide at least a crude measure of lunar parallax and hence lunar distance.

Aristarchus, Hipparchus (in his lost *On Sizes and Distances*), and Ptolemy (*Canobic Inscription* and *Almagest* 5.13–16) all invoked geometrically defined relationships between the Earth, Moon, and Sun that at least theoretically established quantitative relationships between the distances of the Sun and Moon and the Earth's radius. Hipparchus used two methods, one making the assumption that the Sun's distance was very large (specifically that the maximum solar parallax was 7′ so that the distance was 490 Earth-radii), the other that it was effectively infinite, and obtained estimates of the Moon's mean distance as respectively 66 2/3 and 77 Earth-radii.[19] In the *Canobic Inscription* Ptolemy gives the lunar and solar distances as respectively 64 and 729 Earth-radii, a numerologically appealing pair of numbers, and it seems probable that here the lunar distance was the given datum, though its origin is obscure. In the *Almagest* Ptolemy ostensibly derives his mean lunar distance at syzygy, 59 Earth-radii, from a parallax obtained by comparison of an observed and computed position of the Moon, and from this he obtains the solar distance as 1210 Earth-radii. In other words, Greek astronomers were able to estimate the lunar distance fairly successfully

[19] Gerald J. Toomer, "Hipparchus on the Distances of the Sun and Moon," *Archive for History of Exact Sciences* 14 (1974), 126–42.

(the correct value is about 60) whereas their solar distances were highly unstable and more than an order of magnitude too small.

A scheme of cosmic distances ascribed – how reliably it is impossible to say – by the Christian author Hippolytus (ca. 200 CE) to Archimedes illustrates one approach to the problem of assigning distances to all the heavenly bodies making up a complete cosmic system.[20] The distances – expressed in stades, the conventional Greek unit for terrestrial distances, not Earth-radii – are sums of progressively increasing integer multiples of two constants: 10^7 stades and 272,065 stades, which is apparently an approximation of $\pi\sqrt{(3/4\times10^{10})}$. The rationale for these constants is not obvious, but underlying the integer multiples of the second constant is an allegedly Pythagorean scheme of intervals of cosmic musical pitches in steps of tones and semitones. Ptolemy's *Canobic Inscription* ends with a different arithmetical scheme of cosmic pitches that also seems to have had some intended relation to the cosmic distances that he believed in when he composed the inscription.

In his much later *Planetary Hypotheses*, Ptolemy deduced the series of cosmic distances according to a fundamentally different, physical rationale. If each planetary model in the *Almagest* is understood to be a geometrical idealization of a system of etherial bodies that together make up a spherical shell concentric with the Earth, the *Almagest* provides the ratio of radii of the inner and outer surfaces of each shell; Ptolemy further assumes in the *Planetary Hypotheses* that the shells neither overlap nor are separated by gaps. The absolute dimensions of the shells for the Moon and Sun follow from the *Almagest*'s data, and Ptolemy finds that the interval between them is just sufficient for the shells of Mercury and Venus. Continuing the series of nested shells outward in the conventional order Mars, Jupiter, Saturn, and fixed stars, Ptolemy arrives at a scale of cosmic dimensions such that the radius of the inner surface of the etherial part of the cosmos, i.e. the minimum distance of the Moon from the Earth, is 33 Earth-radii while that of the outer surface, effectively the maximum distance of Saturn since the stellar sphere requires no significant thickness, is 19,865 Earth-radii.

OBSERVATIONS AND MEASUREMENTS

Among the varieties of empirical data associated with Greco-Roman astronomy, dated observations gained a special status as foundations for quantitative deductions. In contrast, however, to the thousands of dated observation records preserved in archaeologically recovered cuneiform tablets from Babylon, fewer than 200 Greco-Roman dated observations survive, chiefly preserved through the medieval manuscript tradition. Ptolemy's *Almagest* is

[20] Catherine Osborne, "Archimedes on the Dimensions of the Cosmos," *Isis* 74 (1983), 234–42.

our source for almost all such observations up to his own time, and in that work they appear singly or in small groups, embedded in Ptolemy's deductive arguments. Ptolemy's sources for his predecessors' observation reports seem to have included collections of such reports as well as earlier theoretical works in which reports were incorporated.

With the exception of the Meton-Euctemon summer solstice report from 432 BCE, which was remembered outside the technical astronomical literature as the calibration date for a calendrical cycle, the earliest known dated observations were made during the early Hellenistic period. We know of several disparate observational programs during the third and early second centuries BCE. An observation of the summer solstice of 280 BCE by "those around" Aristarchus of Samos (Ptolemy, *Almagest* 3.1) was probably again calendrical in purpose. On the other hand, from the 290s through the 270s BCE Timocharis, an astronomer working at Alexandria, recorded dated observations of lunar eclipses (*Almagest* 3.1) and of passages of the Moon, Venus, and likely other planets near fixed stars (*Almagest* 7.3 and 10.4), with what immediate application is not known. Partly overlapping Timocharis' period of activity, another astronomer or group of astronomers of unknown identity but probably working in Egypt recorded passages of planets (Mercury, Mars, and Jupiter) by fixed stars from the 270s through the 240s BCE (*Almagest* 9.7, 9.10, 10.9, 11.3, and *POxy. astr.* 4133). Archimedes (mid- to late third century BCE) observed dates and times of solstices in several years (*Almagest* 3.1), possibly to measure the length of the year. Lastly, we have four anonymous records of lunar eclipses observed in Alexandria from 201 through 174 BCE (*Almagest* 4.11 and 6.5).

The rise of Greek observational programs, especially in Ptolemaic Egypt, in the third century might have taken some inspiration from a deeper acquaintance with Babylonian observational practices than was reflected in such vague assertions concerning the antiquity of Near Eastern observations as we find in earlier authors (e.g. Pseudo-Plato, *Epinomis* 987a, and Aristotle, *De Caelo* 292a7). The lunar and planetary passages by fixed stars have a loose resemblance to the Normal Star passages in the Babylonian Astronomical Diaries, but metrology and other details are different. In the latest of the Alexandrian eclipse records, from 174 BCE, the magnitude is given in eclipse digits, a Babylonian unit, but the timings for all the eclipses are in Greco-Roman seasonal hours.

Ptolemy's discussions of aspects of the theoretical work of Hipparchus (fl. 162–127 BCE) sometimes enable us to identify how Hipparchus used his own observations, though even here one cannot always presume that he made the observations with these applications, sometimes many years later, in mind. He is also the earliest Greek astronomer known to have used observation reports of his predecessors, both Greek and Babylonian; he had access, in fact, to translations of lunar eclipse records from Babylon dating as far back as the eighth century BCE. Hipparchus' own dated observations include equinoxes (autumnal for many years from 162 to 158 and from 147 to 143 BCE, vernal for most years from 146 to 128 BCE, *Almagest* 3.1), summer solstices (158 and 135 BCE, and certainly several other years, *Almagest* 3.1 and *PFouad* inv. 267A), lunar eclipses (146, 141, and 135 BCE, *Almagest* 3.1 and 6.5), and lunar elongations from the Sun (128–127 BCE, *Almagest* 5.3 and 5.5). The place where he observed, when known, was Rhodes. He applied solstice, equinox, and eclipse observations – his own, predecessors', and at least one equinox observed by an unknown contemporary in Alexandria – to investigate solar and lunar periodicities and lunar anomaly.

A gap of more than two centuries separates Hipparchus' latest observation reports from the next surviving ones, three lunar passages of fixed stars observed in the 90s CE by Agrippa (an otherwise unknown astronomer working in Bithynia) and Menelaus of Alexandria (working in Rome). Of exceptional interest is a 104 CE observation of Jupiter's position relative to fixed stars coinciding with its opposition to the Sun, made by the unknown author of an astronomical treatise, a fragment of which survives as *POxy. astr.* 4133. The context suggests that it was used for a determination of Jupiter's periodicities. The last known observer before Ptolemy is a certain Theon "the mathematician," who was an acquaintance of Ptolemy's and passed on to him reports of stellar passages of Venus and Mercury (127–32 CE).

More than thirty dated observation reports in the *Almagest* are ascribed by Ptolemy to himself. Ranging from 127 through 141 and including representatives of nearly every variety of observation known from the Greco-Roman tradition, they are nevertheless a somewhat perilous witness to the observational practices of his time because many of them appear to have been adjusted or even fabricated to yield theoretical results that Ptolemy must have obtained by other routes than he sets out in the *Almagest*. Thus Ptolemy's reports of equinoxes and solstices (*Almagest* 3.1, 3.4, and 3.7) have an ostensible precision of one equinoctial hour, which is difficult to reconcile with his stated preference for basing such determinations on meridian instruments, that is, measurements made only at noon.

The few dated observations from the time after Ptolemy appear to have been made in a context of teaching and verifying Ptolemy's theories and tables. Theon of Alexandria reports what he claims are observed times of the stages of a solar eclipse in 364 (which would be the only scientific solar eclipse record from the Greco-Roman tradition), but the times, since they exactly match those that he computes using Ptolemy's tables, are suspect. Finally, the manuscript tradition of the *Almagest* preserves seven observations of the Moon and planets made by the Neoplatonist Heliodorus, his brother Ammonius, and their uncle Gregorius over the interval 475–510 CE; for two of these observations, discrepancies with computed positions are noted.

Numerical parameters frequently appear in the ancient technical literature that ostensibly originated in measurements without giving rise to dated observation records. These include such quantities as geographical latitudes, the obliquity of the ecliptic, maximum celestial latitudes for the Moon and planets, and angular apparent disk sizes of the Sun and Moon. Details of the method of derivation are often lacking; when they can be deduced, superficially precise parameters may turn out to reflect rather crude empirical data, as is the case with Ptolemy's latitude 30° 58′ for his own location, Alexandria (*Almagest* 5.12), which was calculated to a spurious precision from an equinoctial noon shadow ratio of 3:5.

Instruments employed in astronomical observations and measurements may be classed into those that determined directions to celestial objects directly (by sighting) or indirectly (through shadows), and time-measuring devices. No instrument specifically dedicated to scientific observation is extant, but descriptions feature in the astronomical literature, most notably the *Almagest*. The most complex instrument described by Ptolemy (*Almagest* 5.1), the observational armillary, comprised seven nested bronze rings and could be used to determine both the celestial longitude and latitude of a heavenly body and the time. Many of the transmitted observational reports, however – including some of Ptolemy's own – could have been performed with no instruments at all.

ASTRONOMICAL PREDICTION

The notion that the occurrences of certain celestial phenomena can be predicted is at least as old in Greek culture as Herodotus' notorious statement (*Histories* 1.74) that Thales of Miletus "forecast to the Ionians" that a sudden turning of day into night – usually interpreted as a solar eclipse – would take place in a specific year. In reality early predictive methods relied on simple periodicities and were certainly far below the level of sophistication required, say, to forecast that a solar eclipse would be seen in a particular locality. *Parapēgmas* could be used to predict dates of solstices, equinoxes, and stellar visibility phenomena as occurring at intervals of one solar year, while calendrical cycles could be understood as schemes of prediction for the Moon's phases. Thucydides (2.28) was aware that solar eclipses only took place at a certain stage of the lunar month, though Greek knowledge of eclipse periods likely came much later. The eclipse prediction dials of the Antikythera Mechanism (second or first century BCE) operate on the assumption that syzygies at which lunar or solar eclipses are spaced at intervals of six or five months through a repeating 223-month "Saros" cycle, with the times of true syzygy supposedly shifting later by eight hours every cycle.

More complex, numerical methods of prediction likely first appeared in the Greek world during the second century BCE, though direct and indirect evidence for them is scarce before the late first century BCE. Astronomical tables and instructional texts in Greek (and sometimes demotic) papyri attest to two methodologies: one based on arithmetical sequences and algorithms that can be traced in large part to Babylonian astronomy, the other incorporating trigonometrical functions derived from geometrical models of uniform circular motion. The fact that both methodologies worked with the Babylonian metrology of degrees and sexagesimal place-value representation for fractions (with a zero symbol employed more consistently than its Babylonian counterpart) strongly suggests that a large-scale transmission of Babylonian mathematical astronomy, probably connected with the circulation of astrological practices, was the stimulus for the rapid growth of Greek numerical prediction that culminated in the trigonometrically based tables of Ptolemy's *Almagest* and *Handy Tables*.

Most Greek arithmetically-based methods resembled their Babylonian antecedents in that their primary output consisted of discrete predictions of the dates and other circumstances of a series of successive events of the same kind, for example first morning appearances of Jupiter. Computed positions of heavenly bodies at arbitrary dates, required for horoscopic astrology as well as for astronomical research, were obtained by bridging the intervals of time and position between the discrete events by means of arithmetical sequences; though these interpolation techniques too were first developed in Babylonia, they appear greatly generalized and standardized in the papyri. By contrast, the trigonometrically based methods begin with a concept of continuously flowing time that translates first into the individual uniform circular motions of the various components in a geometrical model, and secondly, through trigonometry, into the apparent motion observable from the Earth. In Ptolemy's tables, such discrete events as planets' first and last visibilities and stations are found within a series of computed positions by applying certain threshold tests; for example a planet is predicted as being visible at some time of night if its elongation from the Sun exceeds a certain tabulated amount dependent on the observer's terrestrial latitude and the planet's celestial longitude. The tables attain their highest level of sophistication in the treatment of eclipses, where account is taken of factors absent from Babylonian eclipse theory (e.g. variation in the apparent disk size of the Moon and the Earth's shadow dependent on the Moon's distance from the Earth, and parallax).

GENERAL ASTROLOGY

In a loose, nontechnical sense, astrological beliefs can be traced among the Greeks as early as Hesiod (ca. 700 BCE), whose *Works and Days* attributes (perhaps metaphorically) a parching power to Sirius (587–8), and concludes with a passage delineating the character of specific days of the lunar month with respect to various activities (765–821). Besides such acceptance of periodically repeating astral causes or signs of mundane conditions as affecting people, one can point to instances when *irregular* celestial phenomena were so interpreted. Thus in Thucydides' famous narrative of the Athenian campaign against Syracuse (7.50), the general Nicias is said to have delayed the Athenian withdrawal by twenty-seven days at the advice of diviners (*manteis*) following the lunar eclipse of August 28, 413 BCE. It is impossible to say, however, whether the diviners had a systematic approach translating aspects of the eclipse into a specific forecast bearing on the military situation in Sicily, comparable to Mesopotamian astral omens.

Texts providing interpretations of celestial events, detail by detail, as predictors of future conditions for countries, peoples, and rulers are preserved in demotic and Greek papyri from Roman Egypt (first through third centuries CE) as well as in transmitted astrological texts such as the treatise of Hephaestion of Thebes (ca. 300 CE). Much of this material was inspired ultimately by Mesopotamian omen lore and astrology, though in the form it takes in the extant texts it reflects an Egyptian environmental and geopolitical perspective. The astral events treated as ominous are chiefly eclipses and the configuration of the heavenly bodies at the morning rising of Sirius, which occurs about the time of the onset of the Nile's annual flooding. While certain Egyptian eclipse omens point to a pre-Hellenistic transmission from Mesopotamia to Egypt around 500 BCE, the prominence of the zodiac in the Greco-Egyptian tradition, as well as various details in the astrological forecasts, point to composition in the mid-Hellenistic period, roughly the second century BCE. Egyptian temples, in accordance with their role as centers of preservation and development of intellectual traditions, were likely the principal milieu of this variety of astrology, but we have scarcely any indications of actual practitioners or clients.

In his *Tetrabiblos*, Ptolemy explains astrological prediction as a science based on knowledge of cause-and-effect relations between the eternally unchanging and ever revolving heavenly bodies, which radiate powers of physical change reducible to the fundamental oppositions hot–cold and wet–dry, and the sublunary world in which we live. Predictions of large-scale changes operating on the scale of regions and peoples constitute "general" astrology as distinct from the kind of astrology that concerns individuals; the key significant events are eclipses, synodic phenomena of the planets, and cardinal stages of the year. Since causes originating in the heavens must compete with many other contingent causes, astrological prediction in Ptolemy's view is inherently uncertain and inexact (in contrast to the mathematical certitude that he attributes to the astronomy of the *Almagest*), but it is most reliable when dealing with the large scale.

PERSONAL ASTROLOGY

Greco-Roman personal astrology was founded, probably about the second century BCE, on a combination of two elements, one of Babylonian origin, the other of Egyptian. From Babylonia came the idea that the configuration of heavenly bodies in the zodiac on an individual's birthday, together with other astronomical phenomena falling on or near that date, provide a basis for predicting the character and life of the individual. Babylonian horoscope tablets recording such combinations of data, derived sometimes from observational records and sometimes from mathematical predictions, are attested from the late fifth century through the first century BCE. From Egypt came the idea – an adaptation of the older systems of "hour-watcher" decan constellations – that the points of the zodiac crossing the horizon and meridian at any moment had special bearing on that moment.[21] Thus a Greek horoscope consisted at a minimum of the longitudes of the Sun, Moon, and five planets in the zodiac at the time of a person's birth, together with the longitude of the ascendant point (*horoskopos*, literally "hour-watcher"). All the data in a Greek horoscope were computed; the enormous popularity of horoscopic astrology under the Roman Empire thus was a prime motivation for the development and propagation of predictive astronomical tables. Among the personal horoscopes preserved in papyri, the great majority contain only the minimum data, i.e. the positions of the seven heavenly bodies and the ascendant, and limit precision to entire zodiacal signs. A few, however, express longitude with precision to degrees and minutes, and supply additional data, both astronomical (e.g. latitudes of the Moon and planets, and data on nearby fixed stars) and astrological (e.g. subdivisions of the zodiacal signs under the astrological lordship of heavenly bodies).

A variety of expressions existed for the practitioner of astrology, including *genethlialogos* ("interpreter of nativities"), *chaldaios* ("Chaldean," commemorating the supposed Babylonian origins of the art), *mathematikos* ("mathematician" or more generally, "one who has knowledge"), and *astrologos* ("astronomer"). Notwithstanding the ambiguity of the last two words, astrologers were astronomers only in a superficial sense – they made no observations and had little or no interest in astronomical theory – and beyond the specialized skill of performing arithmetic with base-sixty fractions, they needed no mathematics. In Egypt, temples such as those of Medinet Madi and Tebtunis in the Fayum were loci of astrological practice, but most astrologers were

[21] Dorian G. Greenbaum and Micah T. Ross, "The Role of Egypt in the Development of the Horoscope," in L. Bareš, F. Coppens, and K. Smoláriková (eds.), *Egypt in Transition: Social and Religious Development of Egypt in the First Millennium* BCE (Prague: Czech Institute of Egyptology, 2010), pp. 146–82.

probably independent operators who learned their profession by apprenticeship.[22]

The process of personal horoscopy began with a client providing the astrologer with the date, time, and (if not local) place of a nativity: his or her own, or a child's. (The times were generally to the precision of whole hours, and must often have been rough guesses.) Using astronomical almanacs or numerical tables, the astrologer would then calculate the astronomical configuration in effect at the nativity. This configuration could be displayed to the client by laying out colored or engraved stones on a zodiac board, or written down on papyrus for the client to take away. The interpretation was evidently delivered orally, since the horoscopic documents practically never say anything about what the astronomical information meant for the individual. The abundant transmitted Greek and Latin astrological literature largely consists of precepts such as the following, matching combinations of data from a horoscope with outcomes for the owner of the horoscope:

> If Venus and Mars together take the house-rulership of action, they produce dyers, perfumers, tinsmiths, leadsmiths, goldsmiths, silversmiths, farmers, dancers in armor, makers of drugs, and physicians who treat by means of drugs. (Ptolemy, *Tetrabiblos* 4.4)

> Venus in trine aspect with the Moon in a nocturnal nativity and in feminine zodiacal signs produces agreeable and happy people, some of them also lead throngs and are deemed worthy of purple vestments and gold adornments according to the magnitude of the nativity; they become philosophers and musicians and learned men and people within the friendship of kings. (Vettius Valens, *Anthologiae* 2.17)

The astrologer would need to select and adapt from his corpus of precepts to obtain statements appropriate for the client's circumstances.

A different approach to "astrology for the masses," not personalized like horoscopy but doing away with the need for a specialized interpreter, is represented by a variety of table called *ephēmeris*. Many fragments of ephemerides survive on papyri, and literary references cast light on their purpose and users. An ephemeris was organized according to a civil calendar (in the extant examples, the Egyptian or the Roman calendar), tabulating for each day of each month of a particular year the computed longitudes of the Moon and, usually, also the Sun and the five planets. Following comparatively simple rules, the possessor of an ephemeris could identify significant configurations of the Moon relative to the other bodies ("aspects"), from which is derived an assessment of the day as auspicious or inauspicious for important actions. This was a particular application of a division of astrology

[22] Alexander Jones, "The Place of Astronomy in Roman Egypt," *Apeiron* 27.4 (1994): *The Sciences in Greco-Roman Society*, ed. Timothy D. Barnes, pp. 25–51.

called *catarchic*, which was concerned with predicting how the outcomes of actions are determined by the astral configuration in effect when they are initiated.

Astrology's claim to scientific validity rested less directly on empirical and deductive argument than on authority, especially by appeal to certain foundational texts, now lost except for fragments and testimonia, by sages of old, in particular two legendary Egyptians, King Nechepsos and his associate Petosiris, whose knowledge came by some kind of revelation. Although astrology ostensibly operated within a more or less physical conception of how the world works, the literature tends to be vague about its relation to astronomical models or to recognized physical processes of change. Ptolemy is exceptional in his concerted efforts to rationalize and reform the received tradition by grounding it in a version of the Aristotelian five-element cosmology; in antiquity, however, his influence was less among astrologers as an astrological theorist than as the composer of the most reputable and convenient astronomical tables.

20

GREEK AND GRECO-ROMAN GEOGRAPHY

Klaus Geus

The Greek term "geography" was coined rather late, probably by Eratosthenes, the famous polymath and librarian of Alexandria, in the late third century BCE. It superseded the older expression *gês periodos*, which means "traveling along the edges of the Earth," i.e. "description of the known world (according to a map)" or, in some contexts, simply "map." It seems that *gês periodos* was the title of nearly all the works with geographical content before Eratosthenes (e.g. those of Anaximander, Eudoxus, or Dicaearchus). That it took until the heyday of Hellenistic times to find a more suitable term for regional and topographical studies attests to the fact that the Greeks and Romans did not consider "geography" a discipline sui generis. To be sure, ancient authors never tired of emphasizing that spatial knowledge was important for generals, politicians, merchants, and even for philosophers and the like, but geography was never studied for its own sake; geographical knowledge always served another practical or ancillary purpose.

We are not informed why Eratosthenes felt compelled to use *geôgraphia* for his work. He may have thought that the term *gês periodos* was adequate only for the description of the coasts and borders of the world, but not for the description of its entirety, or he may have been inspired by a word play of Herodotus (4.36) – but his aim in coining the new term is clear: *geography* should mean also, and primarily, cartography. A geographer was a scholar who "dared" (Agath. 1.1 [*Geographi graeci minores*, ed. Müller (henceforth "GGM"), II 471]; Eustathius *Commentary on Dionysius Periegetes* [GGM II 208]; cf. Strabo 1.1.1) to draw a map of all parts of the known world. In this vein the word geography was used throughout antiquity, e.g. by Ptolemy in his famous work *Geôgraphikê Hyphêgêsis* (*Introduction to geography*). For the ancients, a geographer was basically a cartographer.

It is consistent with this line of thinking that Eratosthenes considered as the first geographers not Homer or Hesiod but Anaximander and Hecataeus, who had indeed drawn maps. Of course, this was a break with tradition, which commonly named Homer the "proteus heuretês" ("first inventor") of every discipline (cf., e.g., the lengthy discussion in chapters 1 and 2 of Strabo's first book). But Eratosthenes demonstrated that most of Homer's geographical and topographical information was false, or at least questionable, combatting the view of Homer as the founder of geography with the acerbic statements "that poets only want to entertain and not to teach" (Strabo 1.1.10; cf. 1.2.3) and "that one will find where Odysseus wandered when you find the cobbler who sewed up the hide of winds" (Strabo 1.2.5; cf. Eustathius 1645 [*ad Homeri Odysseam* 10.19]).

In spite of that, much interesting and reliable spatial information permeates Homer's works, which were composed in Asia Minor in the second half of the eighth century BCE. At that time the Greek colonization had started from Miletus, and Homer may have even heard about lands and peoples in the western Mediterranean or along the Atlantic coasts. But this geographical knowledge was superimposed by layers and elements of mythology such as the Elysian Fields or the Grove of Persephone, located at the fringes of the known world (*Od.* 4.563; 10.508–12). Scanning through the *Iliad* and *Odyssey*, we can extract Homer's worldview as follows: the "limitless Earth" is a circular plane, surrounded by the Okeanos, the "soft flowing" and "deep" river (*Iliad* 14.245 f.; 18.607; 20.7; *Odyssey* 11.157; cf. Strabo 1.1.7). The eternal sky rests like an arch on columns, supported by the giant Atlas in the west, and encompasses the lands and seas of the Earth. At its center rises Mount Olympos, where the Greek gods dwell (cf. *Odyssey* 1.52 ff.).

Homer's worldview was accepted by his younger contemporary Hesiod (ca. 700 BCE) with only a few modifications. Like Homer, Hesiod made the Earth encircled by Okeanos, which he vividly called "back-flowing" and "all-circling" (*Theogony* 242, 959, 983). But for him, the flat Earth has roots and is no longer limitless. Above and beneath the disk two hemispheres converge at the horizon and form a self-contained universe, the *kosmos* (Hesiod *Theogony* 728, 807; cf. *Works and Days* 19). Later poets adhere to this basic concept. Like Homer and Hesiod, they deemed the Earth to be a flat disk. But according to Pindar and the Attic tragedians, the *omphalos*, the "navel" or center of the world, was no longer Mount Olympus, but Delphi, or, more exactly, the Delphic oracle (cf. Pindar, *Pythian* 4.74; fr. 54 = Strabo 9.3.6; Pausanias 10.10.3; Bacchylides 4.4; Aeschylus *Eumenides* 40).

Later, the so-called Presocratics, in their search for causes of certain phenomena, had a more "rational" or, one may even say, more "scientific" approach to explaining the world. Modern scholars distinguish two branches of Presocratics, the "Ionians" and the "Italians." Such a distinction is justified not only by the regional distribution of their protagonists but also in regards to the content of their works: these groups disagreed fundamentally in their cosmological and geographical concepts.

The earliest Presocratics hailed from Ionia in Asia Minor. As already mentioned, the Greek colonization started from here. Its metropolis Miletus was an international center of the exchange of goods and

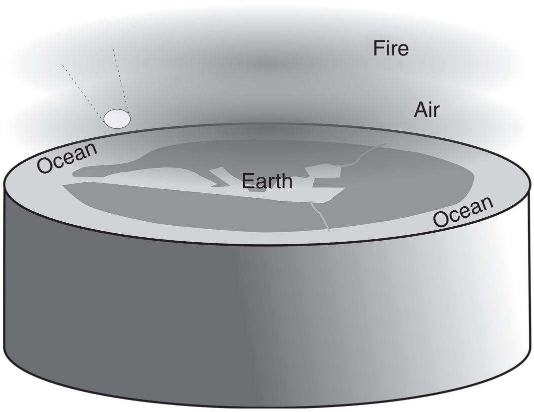

Figure 20.1. The cosmology of Anaximander.

information. Hence, it was not by accident that the first thinkers who tried to put together a coherent picture of the world were based there. Yet, since their knowledge of the world was still rather limited in geographical scope, the Ionians used geometrical models and arguments such as symmetry and analogy, as they had already done in their cosmological theories, in order to move into yet unknown, uncharted territory.

Anaximander (first half of the sixth century BCE), allegedly the pupil of Thales, thought of the Earth as a drum, floating in the middle of the *kosmos*. The height of this cylinder is one-third of its diameter. All peoples live on the top surface of this drum, which is surrounded by the world river Okeanos (see esp. *VS* 12 A 10 and 11; fig. 20.1).

Ancient sources agree that Anaximander was the first who "dared" to draw a map of the world (Agathemerus 1.1 = *VS* 12 A 6). There is no reason to doubt this since Anaximander also worked in the adjacent field of astronomy and is cited as the inventor of the gnomon and the globe, two instruments also essential for geographers. His basic conception for the map is in accordance with that of Homer. Anaximander seemingly made an effort to integrate even the parts of the Earth into his map that were unknown or inaccessible to the Greeks at his time. Despite the fact that Anaximander's map is more about geometry than geography, his idea of using numbers and ratios for the construction of the world map was a key innovation in ancient geographical thought.

Anaximander's disciple Anaximenes (ca. 585–525 BCE) also described the Earth as a flat disk hovering in the air. Sun and Moon float like leaves around it. Interestingly, Anaximenes explained the alternation between day and night as a movement of the heavenly bodies which disappear behind a huge mountain range in the north, called the Rhipaean Mountains, at regular intervals (*VS* 13 A 7; cf. A 14).

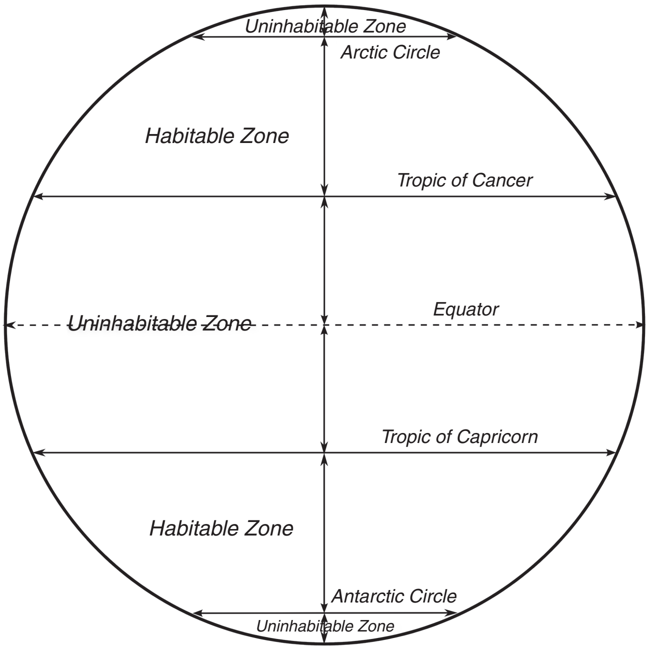

Figure 20.2. The system of five geographical zones. The equation of the zones' boundaries with the tropic and arctic circles is probably later than Parmenides.

The cosmology of Anaximander and Anaximenes exhibits the same basic features that we have found in the archaic poets. They were so common that the idea of a flat Earth was ascribed to other thinkers from Asia Minor as well, such as Thales, Anaxagoras, Democritus, Hecataeus, Herodotus, and Ephorus. One may even call the flat-Earth theory a dogma of Ionian thought.

In contrast, philosophers of Magna Graecia, such as the Pythagoreans and the Eleatic philosophers, took a different stance in cosmology and geography. They advanced a new theory: that of the sphericity of the Earth. Nevertheless, we find some differences between the various "Italian" schools. The Pythagoreans contended that the Earth, a "counter-Earth" (*antichthon*), the sun, the moon, and the five known planets rotate around a central fire. In this, the Pythagoreans were probably more influenced by speculation about ideal forms and numbers than by actual experience.

Parmenides, the founder of the Eleatic school, remained true to his basic philosophical conviction that all being is "all at once, one and continuous." He proposed that the *kosmos* must have the most perfect form, the sphere (*VS* 28 B1, B8; Diogenes Laertius 8.48 = *VS* 28 A44; cf. Diogenes Laertius 9.21). Later doxographers ascribe to Parmenides also other astronomical and geographical ideas (cf., e.g., Diogenes Laertius 9.23). He divided the Earth into "belts" (called *klimata*) running around the surface of the Earth (Figure 20.2). According to Parmenides there are five belts or zones, three uninhabitable – two cold ones around the poles and a hot one around the equator – and two habitable ones in between. Parmenides' contention that there are not one but two habitable zones on the Earth was revolutionary: other regions exist on the surface of the Earth where mankind can live. This idea fathered a new term: *oikoumenê*, i.e. "inhabited" or "known world," as opposed to the "whole Earth" (*gê*).

Parmenides' model of the world remained popular until the end of antiquity. Although subsequent voyages of Greek mariners extended beyond the equator, establishing that the equatorial belt was in fact inhabited, the notion of an uninhabitable zone persisted as late as Ptolemy, who asserts in his *Almagest* that no one from his part of the world had reached the equator so that the question of habitability was open. The authority of Aristotle and the poems of Aratus and Vergil helped to canonize the notion of the hot, impassable zone at the equator. Theory, not autopsy, won the day.

The idea of the spherical Earth advanced by the Pythagoreans and the Eleatics had, at first, only minor repercussions on the Ionian Presocratics in the east. Hecataeus of Miletus (ca. 500 BCE), whom Eratosthenes considered to be the first geographer, integrated the known geographical data into his work. He also attached a map to his text. But Hecataeus' map was an improvement over Anaximander's older map only in a limited sense. Like all Ionians Hecataeus considered the Earth to be a disk surrounded by the Okeanos. And like Anaximander, Hecataeus used geometrical figures in order to structure his map; for example, he drafted Sicily as a triangle (Figure 20.3).

Despite the enormous growth of geographical information at the end of the sixth century BCE, the basic geographical conception of the world had not changed much. The *oikoumenê* was still considered an "island" in the Okeanos. Hecataeus' contribution consisted mainly in updating Anaximander's map and in adding some new data.

We have only a few reports about maps in the period immediately after Hecataeus. We can surely assume that expanded and modified versions were made following the Ionian model. The Ionians, like the Italian Presocratics, distinguished the inhabited from the uninhabitable world, inscribing an oblong rectangle in a circle. This part of the world was also called – just like the inhabited zones in Parmenides' model – *oikoumenê*, "inhabited (world)." But the arguments that the Earth is a sphere and not a disk could no longer be ignored by the end of the classical period. By the time of Plato and Aristotle at the latest, such a view was prevalent within the scientific community (see, e.g., Plin. *NH* 2.64.160).

The geographical conception of the historian Ephorus (ca. 405–330 BCE) also gained currency. The famous geographer Strabo (ca. 64 BCE–23 CE) and the Christian monk Cosmas Indicopleustes (sixth century CE) both reported how the historian Ephorus made use of a rectangular map (Strabo 1.2.28; Cosmas *Topographia* 2.148). In the fourth book of his *Histories* Ephorus

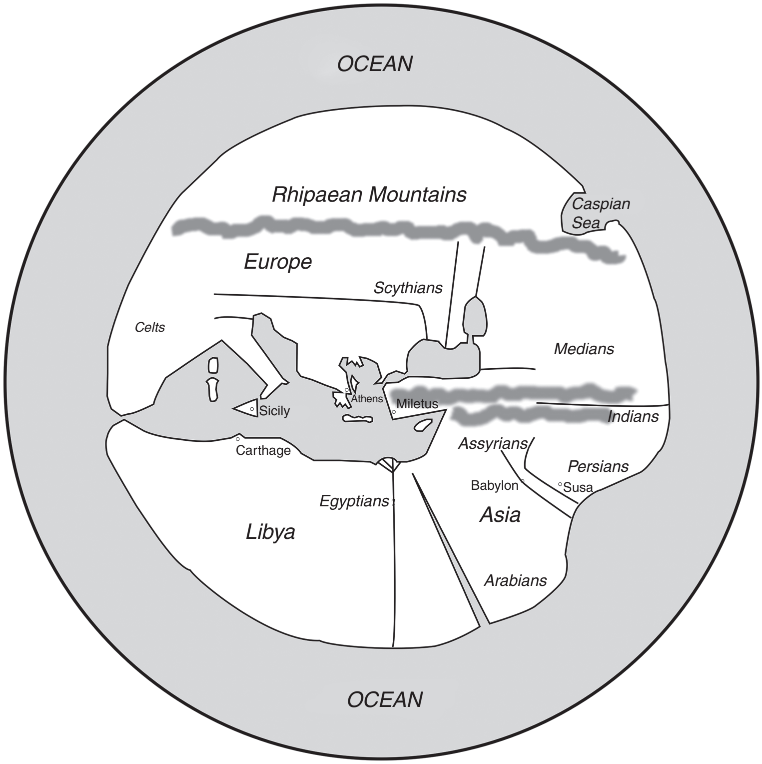

Figure 20.3. Modern reconstruction of the map of Hecataeus.

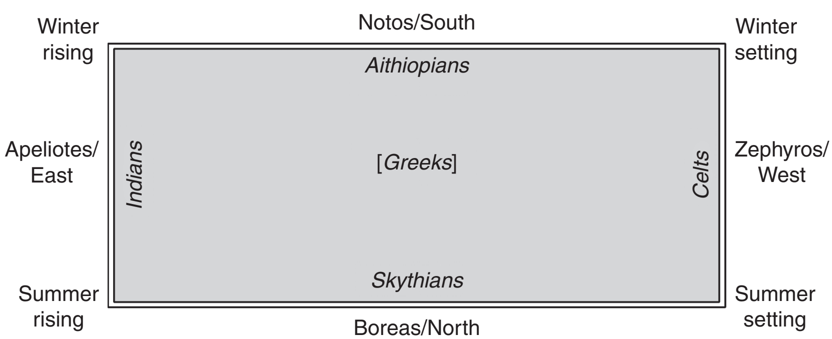

Figure 20.4. Schematic reconstruction of the world map of Ephorus.

sketched out the *oikoumenê* in the traditional form of a rectangle (where the south is on top, see Figure 20.4).

Ephorus equated each side and angle of the rectangle with cardinal directions, expressed as winds and astronomical points, namely where the sun rises and sets on the days of the summer and winter solstice. Each side is also associated with peoples who are supposed to inhabit regions of great extent. While the Skythians are in the north and the Celts in the west, the Aithiopians are situated along the south and the Indians along the east. The northern and southern lines are considered to be longer than the eastern and western, which underlines the fact that the horizon, marked by certain fixed points, is defined with respect to a Greek observer. Thus, Ephorus' model was based on the old concept that the Greeks are located at the center of the *oikoumenê*.

Furthermore, an influence of the ethnographical tradition on geography is discernible here. In a view of the Earth like that of Ephorus, "normality," in other words "order" and "civilization," dominate at its center, that is, in Greece, whereas towards the edges the environmental conditions are more extreme and give rise to simplistically characterized peoples. Blissful barbarians, such as the Hyperboreans, Skythians, and Aithiopians, live in the north and south. Fantastic lands stretch out far to the west and east: the garden of the Hesperides and Erytheia lie in the west, India in the east, and Amazonia in the north-east. Ephorus seems to owe much to this ethnological model, when he claims that the *oikoumenê* is bounded by certain peoples. Such a worldview could even influence the first Greeks who visited these lands. For example, when Alexander's Macedonians heard about female warriors living in the vicinity of Hyrkania or Chorasmia, they were quick to identify them with the Amazons of legend (Curtius 6.5.24 ff.; Plutarch *Alexander* 46; Arrian *Anabasis* 4.15.4; cf. Diodorus 17.77.1 ff. et al.).

But let us return to the more "scientific" geographers. As already mentioned, in the years after Hecataeus, geographers quarreled over the size of the *oikoumenê*, which was represented as a rectangle by the majority. Democritus proposed a ratio of 3:2 for the ratio between the *oikoumenê*'s length (from west to east) and its breadth (from north to south); Eudoxus preferred 2:1; Dicaearchus chose 3:2; and according to Eratosthenes – who called the *oikoumenê* "chlamys-shaped" after the Greek name of a type of cloak – it was more than 2:1. Aristotle estimated on the basis of data from land and sea that the ratio was more than 5:3 (cf. Agathemerus 1.2; Strabo 1.4.5; Aristotle, *Meteorology* 2.5, 362b). According to the Hellenocentric view, the central position of Greece had endowed the Greeks with the best qualities of the northern and southern peoples. By such geographical reasoning, Aristotle could derive a legitimate claim for Greek supremacy over the "barbarians" (cf. Aristotle *Politics* 1.6, 1255a–b; 3.14, 1285a).

Aristotle was also one of the first to inform us about attempts to measure the circumference of the Earth, and cites unknown "mathematicians" (Eudoxus?) for the figure 400,000 stades (*De Caelo* 2.14, 298a15). Some years later, Archimedes (287–212 BCE) shortened it to 300,000 stades (*Arenarius* 1.8), which was still much too high, since these numbers are equivalent to about 63,000 or 47,000 km instead of about 40,000 km.

The man who put the measurement of the circumference of the Earth onto a scientific, i.e. mathematical, basis was Eratosthenes of Cyrene (276–194 BCE), the versatile scholar and head of the famous library of Alexandria. Eratosthenes' method (see esp. Cleomedes 1.7.51–110) was based on a simple proportion: he equated the difference between the astronomically observed latitudes of two places – Alexandria and Syene, lying roughly on the same meridian – with their known terrestrial interval, i.e. the measured distance on the ground. By this means, he arrived at an accurate estimation of the circumference of the Earth (assuming that he used the "itinerary" stade of 157.5, his calculation circumference – 250,000×157.5 – arrived at 39,375 km). Eratosthenes also drew a new map of the *oikoumenê*, which was a major improvement on the older maps with their speculative elements and characteristic appeal to principles of analogy and symmetry. He was also the first to draw parallel circles and meridians and to integrate them into a proper system.

In other respects too, his approach was more geometrically objective than that of his predecessors. The cornerstones of his map consisted no longer of peoples (Indians, Celts, etc.) whose geographical extent was ill defined, but of important cities (Carthage, Meroe, etc.) and landmarks (Pillars of Herakles, Cape Notou Keras, etc.). It was, however, only one step. Eratosthenes was unable to determine more than a handful of parallels (and even fewer meridians) running through some of the most important cities of early Hellenistic times (Alexandria, Rhodes, Byzantion, Carthage, Massalia [i.e. Marseille], Gades [i.e. Cadiz]). Thus, since these cities were chosen for their civil prominence or for other nongeographical reasons, they formed no fully abstract and geometrical set of coordinates. Hence, the modus operandi of Eratosthenes must not be judged as a complete break with geographical tradition, but as a compromise between ideal and reality.

Later astronomers and geographers criticized Eratosthenes for his extensive use of itineraries (i.e. records of road systems) and *periploi* (analogous documents for sea routes) in determining the latitudes and meridians of important locations. Hipparchus of Nicaea, working in the latter half of the second century BCE, is said to have demanded that the relative positions of places be determined solely by astronomical observations (cf. frs. 35, 26, 39 Dicks). As understandable and correct as such a demand was, it was one which simply could not be carried out in ancient times. Only a handful of locations had ever been measured astronomically. The preserved fragments make it clear that Hipparchus tried neither to condemn everything Eratosthenes proposed nor to replace Eratosthenes' geographical system with his own. Rather, Hipparchus' criticism was aimed mostly at particular cartographical aspects of Eratosthenes' geography. In consequence, he advised his readers to use the older Ionian maps (cf. fr. 14 Dicks) since in his view they represented a standing hypothesis that Eratosthenes had failed to prove incorrect.

Hipparchus' work, aptly entitled "Against the *Geography* of Eratosthenes," was a commentary on Eratosthenes from an astronomical point of view. This view is confirmed by Strabo (2.1.41 = test. F Dicks):

> Therefore, for Hipparchus who was not writing a geographical treatise, but was making a critical examination of the statements made by Eratosthenes in his *Geography*, it would have been fitting for him to have gone into further details of correction.

After Eratosthenes, Greek scholars went separate ways. Some, like Posidonius (ca. 135–51 BCE), were interested in physics (a part of philosophy) and wrote about geological matters; some, like Agatharchides (ca. 200–after 131 BCE), Polybius (ca. 200–120 BCE), and Artemidorus (first century BCE), produced ethnographical accounts about peoples and regions of the *oikumenê*; some, like Pseudo-Skymnus (second century BCE) and Dionysius of Alexandria (also called Periegetes; early second centry CE), rehashed spatial information for a broader audience by putting it into verse. None of them produced a new map of the world (Strabo 2.4.1).

Even Strabo (64/63 BCE–ca. 24 CE), who wrote a huge "geographical" (in the modern sense) work in seventeen books, never drafted a map. Accordingly, he refers to his own work as *chôrographia*, *periêgêsis*, *periodos gês*, and *periodeia tês chôras* (3.4.5; 6.1.2; 9.5.14). The only exception seems to come at the beginning of his third book (Strabo 3.1.1), where he indeed speaks about *geôgraphia*, but he clearly means by this the earlier cartographical efforts like that of Eratosthenes, which he had discussed in Books 1–2. Hence, a better title for Strabo's work would be "chorography" instead of "geography." As the titles of the works of Pomponius Mela and Pappus of Alexandria, which employ *chôrographia* rather than *geôgraphia*, show, the term "chorography" was used in antiquity for a comprehensive description of the whole world in a *non-cartographical* mode. Strabo's work, in fact, contains a thorough description of the entire *oikoumenê* which overlaps to a large extent with the *Imperium Romanum* of Augustan times. He advanced and discussed cartographical theories on the Okeanos, the three continents, the *klimata*, and the five zones as important elements of *oikoumenê*. But by imposing on this rich material other spatial, political, and ethnographical concepts (for example, Hellenes and barbarians, or centers and periphery) for every region, Strabo's work gained a new outlook, one that we nowadays would not hesitate to call "geographical."

But again, this was not how the ancients would have described it. A final hint can be detected in the famous definition of geography at the beginning of the *Introduction to "Geography"* (*geôgraphikê hyphêgêsis*) written by the famous astronomer Claudius Ptolemy (Klaudios Ptolemaios) around 150 CE:

> Geography is an imitation through drafting (*dia graphês*) of the entire known part of the world together with the things which are, broadly speaking, connected with it. It differs from chorography . . .

The phrase "together with the things which are, broadly speaking, connected with it" surely allows for some margin of interpretation, but Ptolemy was entirely concerned with producing maps, especially the "world-map," in his *Introduction to "Geography."*

By eliminating all political, historical, and ethnographical aspects and data (2.1.8) that he found in his source material (called "research derived from travel," *historia periodikê*, i.e. *periploi* and itineraries; 1.2.2; 1.4 tit.; cf. 1.5.2; 1.6.1; 7.7.4), Ptolemy arguably produced the most influential work in the history of geography. His methods and goals, expounded in detail in the first book, are clear and distinctly explained. The major part of his *Geography* consists of a catalog of no less than 8,000 localities with 6,300 pairs of coordinates of latitudes and longitudes. The geographical work of Ptolemy is an impressive demonstration of applied mathematics. In addition to this, it introduced some long-lasting cartographical innovations such as the north-at-top orientation of modern maps and the use of symbols. Ptolemy's advancements are important not only in terms of the enormous increase in the quantity of information represented, there is also a truly astonishing qualitative shift, in that unlike the former cartographers Anaximander, Hecataeus, and Eratosthenes, Ptolemy did not need to fall back on geometrical figures and estimates of distances to draft a map of the *oikoumenê*.

At this point, Greek geography, the foster child of geometry, appeared likely to emancipate itself as a discipline in its own right. But Ptolemy's groundbreaking work was not taken up by any later scientist in antiquity. In fact, it was almost completely forgotten in late antiquity and the Middle Ages, only to be rediscovered in the Byzantine Empire in the thirteenth century, and in the Latin world only at the beginning of the fifteenth. Why such an ingenious work as Ptolemy's fell into oblivion is hard to fathom. There was probably more than one reason. Apart from the required mathematical expertise, the technical difficulties in producing maps, and the costs of copying the long text and the maps, we may also point to the little-known fact that Ptolemy's work was epitomized by later authors. For example, while describing the Sasanian empire, the Latin historian Ammianus Marcellinus (*History* 23.6) clearly drew on a list of regions which was organized according to the structure of Ptolemy's *Geography*; and Pappus of Alexandria (fourth century CE) wrote a *Chôrographia oikumenikê* (*Description of the "oikumenê"*), extant in Armenian under the title *Ašxarhac'oyc'* (i.e. "worldview"), which is seemingly an abbreviated version of Ptolemy's work. The fact is that world maps played a much smaller role in antiquity than today. For orienting and navigating, the Greeks and Romans used other means, especially *periploi* and itineraries. Accordingly, geography in its original sense, "the art of drafting maps," never found much resonance with the Greeks and Romans.

21

GREEK OPTICS

A. Mark Smith

Although it can be fairly said that, as a science in the modern sense, Greek optics had its wellsprings in ray theory, three caveats are in order. First, the goal of Greek optics was to explain vision, not the physics of light, so the classical ray represented a path for sight, not light.[1] On that account, second, Greek optics must be evaluated within the broader context of concurrent epistemological thought. Finally, the dearth of contemporaneous sources makes it difficult to trace the development of Greek optics with any confidence. These caveats in mind, we shall begin with an overview of the evolution of ray theory from its presumed beginnings around 400 BCE to its culmination in the later second century CE. Next, we shall discuss various models of visual perception formulated over roughly the same period, taking into account underlying physical, physiological, and psychological issues. We shall then narrow our focus upon mathematical optics. The resulting account will reflect the tripartite analytic structure that evolved over the period between Euclid (ca. 300 BCE) and Ptolemy (ca. 160 CE). Accordingly, we shall examine the development of optics proper first (involving unimpeded visual radiation), then of catoptrics (involving fully broken or reflected visual radiation), and, last, of dioptrics (involving partially broken or refracted visual radiation).

HISTORICAL OVERVIEW

Precisely when and why the ray concept was developed is a matter of speculation, but its origins probably lie in the early fourth century BCE and certainly no later than mid-century. Aristotle, for instance, includes optics among the "more physical" of the mathematical sciences in *Physics* 2.2, and

[1] On this point, see Gérard Simon, *Le regard, l'être et l'apparence dans l'optique de l'antiquité* (Paris: Seuil, 1988).

he uses ray analysis to account for the rainbow in *Meteorology* 3.2–5.[2] It is therefore clear that the conceptual framework of mathematical optics was erected well before the appearance of the *Optics* and *Catoptrics* attributed to Euclid.[3] With Euclid, in short, we find ourselves not at the beginning but at a relatively advanced stage in the evolution of mathematical optics. If so, it follows that the key principles of visual-ray analysis – the visual ray, the visual cone, the equal-angles law of reflection, and the principles of image location in plane, spherical convex, and spherical concave mirrors – were common currency by the end of the fourth century BCE.

The course of post-Euclidean optical development is almost as difficult to trace as its pre-Euclidean trajectory.[4] If, for instance, Apuleius (*Apologia* 16) is to be credited, Archimedes (fl. ca. 250 BCE) wrote extensively on optics, undertaking in the process a close study of refraction. Still, given the lateness of Apuleius' testimony (late second century CE) and the lack of direct textual evidence, we must treat this claim with skepticism. Perhaps the earliest extant post-Euclidean optical source is *On Burning Mirrors*, ascribed to Diocles (early second century BCE?). Devoted to the focusing properties of spherical concave and parabolic mirrors, this work betrays a mathematical sophistication far beyond that of its closest known antecedent, the last proposition of Euclid's *Catoptrics*. Nonetheless, any claim about its significance in post-Euclidean optical development must be tempered by recognition of the questionability of its attribution and dating.[5]

As far as theoretical foundations are concerned, virtually no progress seems to have been made after Euclid until the appearance of Hero of Alexandria's *Catoptrics* (mid-first century CE). As we shall see, Hero's contribution lay in the attempt to justify the equal-angles law of reflection on the grounds of mathematical necessity. Because of its current fragmentary state, however, Hero's work is of little use as a gauge of optical development to its time. This leaves us pretty much in the dark about where Ptolemy's *Optics*, the final surviving source in line from Euclid, fits in the evolutionary scheme. That this treatise represents a significant advance over Euclid's two works is indisputable. For a start, its analytic structure is far tighter and more elegant. Being more empirical in orientation, moreover, Ptolemy's analysis is considerably broader in scope. And, finally, unlike Euclid, Ptolemy aimed for comprehensiveness, bringing physical, psychological, and physiological factors to bear in his explanation of sight.

[2] See Carl B. Boyer, *The Rainbow: From Myth to Mathematics* (New York: Yoseloff, 1959).

[3] On the problem of attribution, see Wilbur Knorr, "Pseudo-Euclidean Reflections in Ancient Optics," *Physis* 31 (1994), 1–45.

[4] For the standard account, see Albert Lejeune, *Euclide et Ptolémée, deux stades de l'optique géométrique grecque* (Louvain: Bibliothéque de l'Université, Bureaux du "Recueil," 1948).

[5] For details, see J. Toomer, *Diocles on Burning Mirrors* (Berlin: Springer, 1976).

THE PROBLEM OF VISUAL PERCEPTION

Every theory of sight proposed in antiquity was predicated on the assumption that vision requires physical mediation: that is, no visible object is seen unless it sends something to the eye for its delectation (intromission) and/or unless the eye sends something to it in order to sense it (extramission).[6] At least two intromissionist accounts of vision are known from antiquity. The atomists, for their part, had visible objects emitting representations of themselves (*eidola* or *simulacra*) in all directions. Consisting of either impressions in the air (Democritus, ca. 430 BCE) or atom-thick husks (Epicurus, ca. 300 BCE, and Lucretius, ca. 70 BCE), these representations were supposed to pass into and through the eye to the soul for perceptual scrutiny. For the atomists, then, the mediating entities in vision were taken to be replicas of the objects themselves. Aristotle (ca. 350 BCE), on the other hand, rejected such crass materialism in favor of a more subtle, qualitative explanation of sight. By his account in *De anima* and *De sensu*, what passes to the eye is a formal effect arising from the inherent color of visible objects. In order to reach the eye, though, this color effect must be assimilated by a transparent medium, such as air or water. But such media are only potentially transparent; to become actually transparent they must be rendered so by light, which enables them to take on color and transmit it instantaneously in all directions. Sight occurs when the propagated color makes a sensible impression in the eye. Unlike the atomists, then, Aristotle found the crucial mediating entity for sight in color rather than in replicas or *simulacra*.

The definitive example of extramissionism is, of course, to be found in the visual-ray theory of Euclid, Hero, and Ptolemy, which will be discussed in detail in this chapter. The likeliest conceptual sources for this theory are the line-of-sight techniques of surveying and the perspectival demands of scenography. In physical terms, the visual-ray theory probably harks back to the notion that the visual organ sensibly illuminates external objects with its own fire. Empedocles (ca. 460 BCE), for instance, is said to have likened the eye to a lantern whose fire shines through a protective transparent covering (presumably the corneal sheath), and the "Pythagoreans" are credited with a similar idea. Likewise, the notion of ocular emission underlies the account of nearsightedness in the pseudo-Aristotelian *Mechanical Problems* 31, where myopes are described as squinting in order to concentrate the visual flux (*opsis*) for a more intense visual impression of distant objects.

One burden facing extramissionist theorists but not borne by their intromissionist counterparts was to explain how the visual flux, once in

[6] For details on what follows, see J. I. Beare, *Greek Theories of Elementary Cognition from Alcmaeon to Aristotle* (Oxford: Clarendon, 1906); David C. Lindberg, *Theories of Vision from Al-Kindi to Kepler* (Chicago, IL: University of Chicago, 1976).

contact with external objects, abstracts visual information from them and returns it to the eye. Responding to this issue in the *Timaeus* 45b–c and 67d–e, Plato (ca. 380 BCE) posited complementary but opposing radiations. The eye, for its part, was assumed to emit a gentle fire that, having mingled with the fire of external light sources, flows out toward visible objects. Those objects, in turn, radiate particles of various sizes characteristic of their color. Colliding with the ocular flux, the largest such particles dilate it to create the impression of white. The smallest such particles cause the ocular flux to contract, the resulting impression being of black. Color is thus essentially subjective in the Platonist scheme, a perceptual effect rather than an objective cause of vision.

The Stoics, too, understood the visual process in terms of complementarity. For them the visual process starts with *pneuma* (a sort of fire) emanating from the eye into the surrounding air. Itself pneumatized by this efflux and impregnated by the fire/pneuma of external light sources, the air is transformed into a sympathetic extension of the eye. Through the resultant pneumatic links, the eye "feels" objects visually while they, in turn, make themselves felt to the eye, much as obstacles announce themselves to the hand through a walking stick. As an interplay of perceiver and perceived, sight was thus reduced by the Stoics to a species of touch.

Pneuma does more than physically link external objects and eye; it also establishes a physiological bond between eye and soul, where perceptual adjudication is assumed to occur. Drawing upon a tradition of anatomy and physiology established in third-century BCE Alexandria by Herophilus and Erasistratus, Galen (ca. 180 CE) mapped this connection in exquisite detail. For a start, Galen located the font of sense perception and intellect in the brain (the Stoics placed it in the heart), whose ventricles are, by his account, suffused with psychic *pneuma*. A distillate of the lifegiving vital *pneuma* coursing through the arterial network, psychic *pneuma* flows from the forefront of the brain into the two hollow optic nerves, which cross at the optic chiasma then branch out again to each eyeball. Streaming through the ocular humors, this *pneuma* reaches the crystalline lens to imbue it with visual sensitivity. Having finally pervaded the air at the corneal surface, the pneumatic flux forges sympathetic links through it in radial form. Those links, Galen concluded, provide a visual pathway between the eye, specifically the crystalline lens, and external objects, whose inherent color is the proper sensible for sight.

Few if any of the theorists so far discussed believed that vision consists in a simple sense act. How, after all, do we "see" such spatial characteristics as depth or sphericity through mere sensation? Some perceptual inference must occur. Vision, in short, requires psychological mediation. For the most complete account of such mediation in antiquity, we must turn to Aristotle's *De anima*. As a sense act, according to Aristotle's theory, vision starts with the eye's apprehension of color. The color impression arising from that apprehension is then conveyed to the common sensibility, where all the primal sense data gathered from a given object are integrated with the so-called common sensibles – that is, the spatial or spatially-determined characteristics – defining that object. The result is a perceptible representation that is passed on to the imagination for intellectual scrutiny. By means of such scrutiny, finally, the perceptible representation's ulterior implications are apprehended in the form of "incidental sensibles," which are inferred by reasoning. Visual perception was thus understood by Aristotle as a three-phase process involving: (1) brute sensation, during which color, the special sensible of sight, is apprehended; (2) perception, during which the common sensibles are inferentially grasped; and (3) apperception, during which the intelligible properties are inferentially grasped.[7] As we shall see in the next section, Ptolemy's model of visual perception reflects this tripartite scheme.

VISUAL-RAY THEORY – *OPTICS* PROPER

The primary analytic device for ancient visual-ray theory – the visual cone – is defined in the first three postulates of Euclid's *Optics* in terms of discrete rays emanating in an ever-diverging fashion from the center of sight within the eye. Any opaque object struck by these rays is seen, so visual apprehension would seem to be a function of touch. The Euclidean visual cone was expressly designed to explain variations in visual clarity with distance and the perception of space. Visual clarity, on the one hand, depends on the amount of radiation striking a given visible object. The closer the object to the eye, the more rays strike it, so the more intense the resulting visual impression. The farther the object from the eye, the fewer rays strike it until, eventually, it falls into the interstices between rays to disappear from sight (*Optics*, propositions 2 and 3). In the case of spatial perception, on the other hand, two variables are at play. First and foremost is the visual angle subtended by a given visible object, for it is by apprehending the extent of this angle, Euclid claims, that we judge the object's size (*Optics*, definition 4). Second, Euclidean visual rays possess directional privilege. Hence, by dint of their relative placement within the visual cone, they distinguish rightwardness, leftwardness, upwardness, and downwardness (*Optics*, definitions 5–6). On the basis of these properties of the visual cone and its component rays, Euclid accounts for such diverse spatial perceptions as the apprehension of relative size and the apparent convergence of parallel lines viewed endwise.

[7] See A. Mark Smith, "Picturing the Mind: The Representation of Thought in the Middle Ages and Renaissance," *Philosophical Topics*, 20 (1992), 149–70, esp. 151–6.

Although it has much in common with its Euclidean counterpart, Ptolemy's visual cone manifests certain crucial differences.[8] For one thing, because it is absolutely continuous, its constituent rays are imaginary rather than real. For another, Ptolemy's visual flux is sensitive not to the object itself but to its inherent color. Seeing, for Ptolemy, thus arises from a qualitative change – the "passion" of coloring – suffered by the flux, not from a physical impulse passed through it. Consequently, spatial perception is not immediate; it is a by-product of color perception. Finally, unlike its Euclidean forebear, the Ptolemaic ray conveys a sense of its own length back to the perceiver, who is thereby enabled to perceive distance.

Ptolemy characterizes visibility according to three levels. At the lowest tier lie the "instrinsic visibiles." Comprising luminousness and opacity, these attributes are not per se visible. They are simply prerequisites for visibility: luminousness to render an object's color effectively visible, and opacity to prevent the visual flux from passing through without sensible effect. Next comes color. Being "primarily visible," it constitutes the true sensible for sight, as indeed it does for Aristotle and Galen. Last come the "secondary visibles," which include, but are not necessarily limited to, such spatial characteristics as shape, size, place, and so forth. Perception of these attributes is carried out by the Governing Faculty, which makes all perceptual and apperceptual judgments and is presumably situated at the origin of the optical pathway in the brain.

Altogether, then, the visual process entails three phases for Ptolemy, much as it does for Aristotle. The initial phase occurs when the visual flux suffers coloring by visible objects and transmits the ensuing color effect to the cornea (or "viewer" in Ptolemy's parlance). The consequent sense impression inaugurates the second phase, during which things are spatially defined according to certain geometrical relations that are specified by the visual cone. Size perception, for instance, depends not just on visual angle, but also on distance and even orientation. Consequently, although two objects may subtend the same visual angle, whichever is perceived to be more distant will be judged larger. If, however, one of two objects subtending the same visual angle and lying the same apparent distance from the eye is noticeably slanted, it will be judged larger (*Optics* 2.53–63). The culminating stage in the visual process involves secondary inferences based on these spatial perceptions. Such inferences, for example, figure in the intellectual rectification of facing mirror images, when we mentally exchange real left for apparent right and vice versa.

With this three-phase model of visual perception in mind, Ptolemy develops a remarkably systematic typology of visual illusions. The lowest

[8] For details on Ptolemy's theory of sight, see A. Mark Smith, *Ptolemy's Theory of Visual Perception* (Transactions of the American Philosophical Society 86.2; Philadelphia, PA: APS, 1996), pp. 21–35.

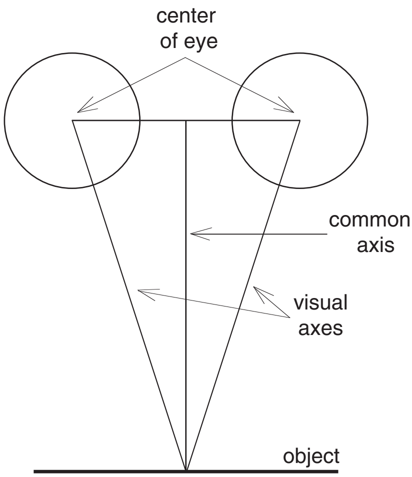

Figure 21.1. Binocular vision according to Ptolemy.

order of such illusions consists in misperceptions arising from anomalous physical circumstances, such as too much brightness or excessive distance. The next order of illusion involves misperceptions due to the sense of sight itself, an instance being the appearance of images behind plane mirrors, which results from the breaking of the incident visual flux. The final order of illusion involves deep-seated psychological misperceptions, the most notorious being the so-called Moon Illusion, in which the moon or sun appears much larger at the horizon than at the zenith (*Optics* 2.59).

Surely Ptolemy's most original contribution to the ray analysis of visual perception is to be found in his account of binocular vision (*Optics* 2.30–44, 3.26–56). We normally see one image with both eyes, he asserts, because our eyes are providentially designed to focus on precisely the same spot of any given object. This tracking requires that both visual axes converge at that spot, the convergence itself being regulated according to the common axis, which is controlled by the Governing Faculty. Thus, under normal circumstances, all three axes intersect at a single point that is ultimately specified by the common axis (see Figure 21.1). Diplopia involves a failure of such natural convergence. In that case a given object appears double, one or both images being displaced because of the axial displacement of either or both eyes with respect to the common axis. As ingenious and geometrically compelling as this account is, however, it suffers from two serious flaws. First, it is based on an implicit assumption that the visual field is planar rather than curved; and, second, it ignores the role of binocular vision in the three-dimensional perception of space.

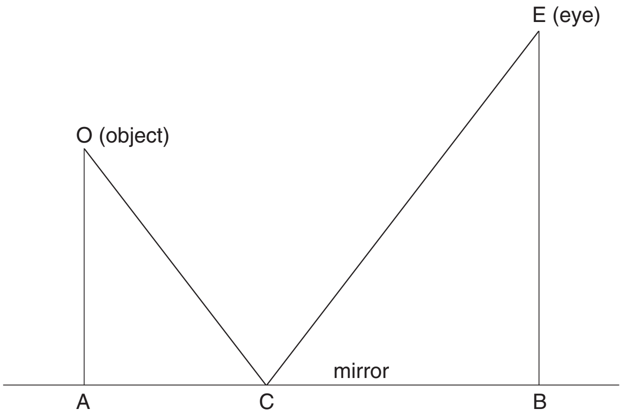

Figure 21.2. Equal-angles law of reflection according to Euclid.

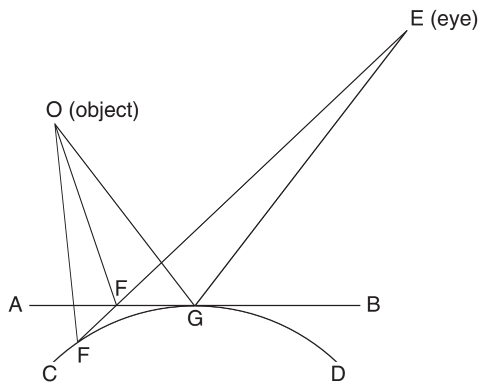

Figure 21.3. Equal-angles law of reflection according to Hero of Alexandria.

CATOPTRICS – THE ANALYSIS OF REFLECTION

Two main problems of reflection occupied Greek theorists: determining image location and, on that basis, explaining image distortion.[9] The cardinal principle governing image location in reflection is the equal-angles law, which asserts the equality of the angles of incidence (*i*) and reflection (*r*). A clumsy attempt to demonstrate this principle is found in proposition 1 of Euclid's *Catoptrics*, the equality of angles following from the postulated ratio EB:BC = OA:AC of triangles ECB and OAC in Figure 21.2, where EC is the incident ray and CO its reflected branch. Far more elegant is the proof offered by Hero of Alexandria in chapter 4 of his *Catoptrics*. Based on the stipulation that visual

[9] For the standard account, see Albert Lejeune, *Recherches sur la catoptrique grecque* (Brussels: Palais des Académies, 1957).

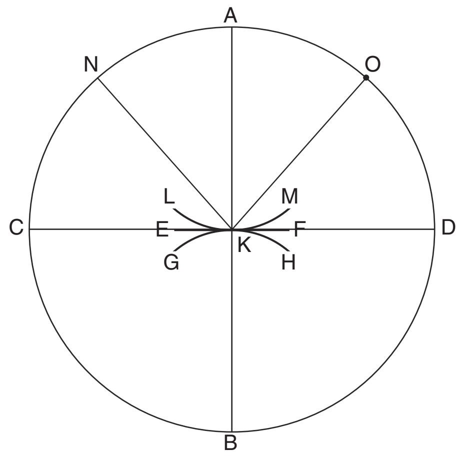

Figure 21.4. Equal-angles law of reflection according to Ptolemy.

radiation follows the shortest possible path (the Principle of Least Lines), this proof is as follows. Suppose that AB in Figure 21.3 is a plane mirror or the plane tangent to convex mirror CD at G. Of all possible ray couples reflecting from AB, Hero concludes, ray couple EG,GO, which forms equal angles with AB, covers the shortest distance (by contrast, for example, to EF,FO). In a sense, then, reflection is resolved into a special case of unimpeded radiation, which also follows the shortest path by virtue of its rectilinearity.

In contrast to Euclid and Hero, Ptolemy sought to establish the equal-angles law inductively (see *Optics* 3.8–12). His experimental apparatus consists of a bronze disk (ACBD in Figure 21.4) divided into quadrants by diameters AB and CD, each quadrant then subdivided into degrees. Three mirrors – plane, circular convex, and circular concave – are placed in turn along EKF, GKH, and LKM, respectively; a line of sight is established along NK; and a marker is moved along arc AD until it reaches O, whose image appears to lie directly in line with K along line of sight NK. When angles NKA and OKA are measured with respect to normal AK, they will invariably prove to be equal, no matter the shape of the mirror.

Of the three remaining principles determining image location in reflection, Euclid and Hero make explicit use of at least two: that the image (I in Figure 21.5a) is seen along the extension of incident ray ED and that it appears to lie where normal OAI dropped from object point O to where reflecting surface AD intersects ED extended. Furthermore, as both Euclid and Ptolemy recognize, this normal passes through the center of curvature C in the case of spherical concave or convex mirrors (see Figure 21.5b). Although implicit in Euclid's and Hero's analysis of reflection, the third principle is explicitly articulated for the first time in Ptolemy's *Optics*: that is, that the reflected ray couple and all the other

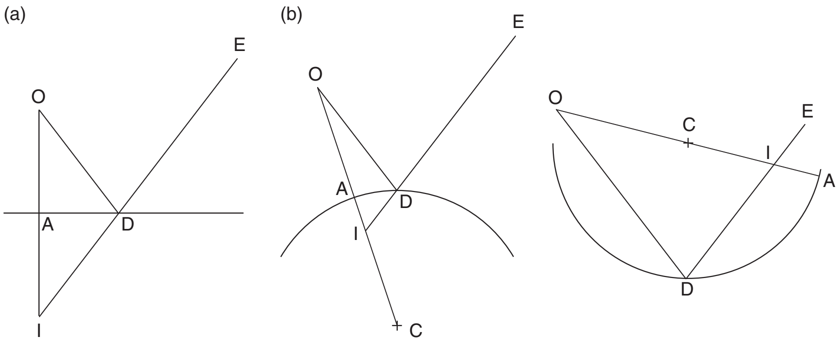

Figure 21.5. Principles determining image location in reflection.

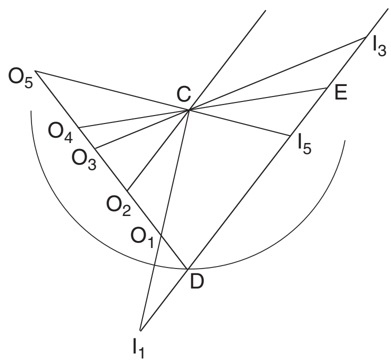

Figure 21.6. Analysis of reflection in concave mirrors.

referential lines lie on a single plane, itself standing normal to the plane of reflection.

Armed with these principles, Euclid and Ptolemy address a host of problems involving image formation and distortion (or lack thereof) in plane and convex mirrors. Both, for instance, prove that, although reversed, images formed in plane mirrors appear to be the same size and shape as their generating objects. By the same token, images in convex mirrors undergo left–right reversal but are distorted in size and shape, the amount of distortion being contingent upon the severity of curvature. Far more complex is the analysis of concave mirrors, for, unlike plane and convex mirrors, concave mirrors have more than one possible image location. A look at Ptolemy's synoptic account of image formation in concave mirrors in *Optics* 4.71–3 suffices to show how problematic the visual-ray analysis of such mirrors is. Let the arc in Figure 21.6 represent a segment of a concave mirror with center of curvature C, E represent the center of sight, and D the point of reflection for ray couple ED,DO_5. Let the object point be placed at various positions O on the reflected ray, and let the respective images I be located along incident ray ED where the normals passing from O through C intersect it. When the object is placed at O_1, it will be seen behind the mirror at I_1, and it will continue to be seen behind the mirror until it reaches O_2. At that point, normal O_2C will be parallel to incident ray ED, so there will be no intersection and thus no image. As the object moves from O_2 toward O_4, its image (for instance, I_3 of O_3) will fall behind the eye where it will be unseen, and at O_4 it will lie at the center of the eye, where it will also be unseen. Yet in all three cases, when the object is at O_2, O_3, and O_4, the reflected ray makes contact with the object, so the object *must* somehow be seen. Given this theoretical imperative, Ptolemy is forced to invent ghost images, the visual faculty supposedly shifting the image to the mirror's surface (or to a point between the eye and mirror in the case of I_3) so that it coalesces with the reflecting surface and takes on its color. In short, the image is reduced to invisibility by psychological transposition. Thus, not only does Ptolemy recognize the limitations of ray analysis, but he also acknowledges the critical role psychology plays in image formation.

DIOPTRICS – THE ANALYSIS OF REFRACTION

The effects of refraction were well known in antiquity, even at a popular level, perhaps the most remarked-upon example being the apparent breaking of a submerged oar at the waterline. In the sixth postulate of his *Catoptrics*, Euclid notes that any object viewed obliquely in a bowl appears to be raised when water is poured in. The earliest known attempt to articulate a physical model of refraction is found in Hero of Alexandria's *Catoptrics*, where the radiation of visual flux is likened to extremely swift projection. Within the framework of this analogy, Hero likens reflection to the rebound of visual flux from a perfectly impermeable surface, much as a swiftly hurled ball rebounds from an unyielding wall. Refraction in turn represents imperfect rebound, which is tantamount to passage through a permeable but resistant medium. Depending on the amount of resistance posed, the visual flux is more or less deflected toward the normal. Thus, as represented in Figure 21.7, when ray EC passes from air into water, its refracted branch CO inclines toward normal AB.

This is essentially the model adopted by Ptolemy, who emphasizes the systemic relationship between refraction and reflection. Especially telling, by his account, is that in both cases the image lies at the junction of the extended incident ray EC in Figure 21.7 and normal DO dropped from object point O to the surface of refraction/reflection. In both cases, too, all of the pertinent line and ray segments lie in a single surface normal to the surface of reflection/refraction. Hence, all that differentiates reflection from refraction is the relationship between the angle of incidence i and the angle of reflection/refraction r: in reflection both are

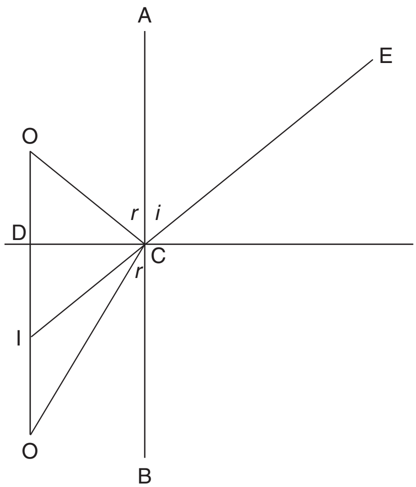

Figure 21.7. Analysis of refraction.

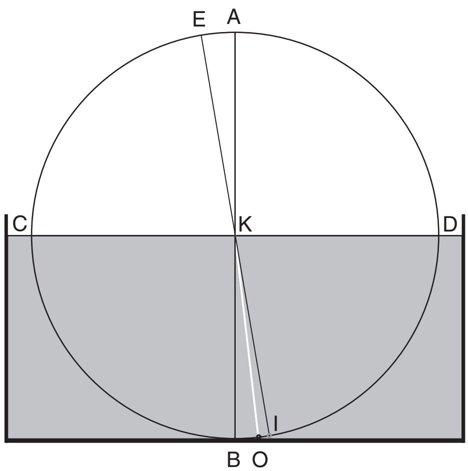

Figure 21.8. Ptolemy's experiment to determine the angle of refraction from air to water.

equal; in refraction they are not. But precisely how do *i* and *r* relate in refraction? This question takes us to Ptolemy's notorious effort in *Optics* 5.8–21 to determine *r* experimentally for refraction from air to water, from air to glass, and from water to glass (we will limit our discussion to the first case). The requisite apparatus consists of the bronze plaque used before in reflection and a vessel deeper than the radius of the plaque. The plaque is inserted vertically into the vessel and water is poured in until it coincides with diameter CD, as represented in Figure 21.8.

i	*r*	*r(modern)*
10°	8.0°	7.47°
20°	15.5°	14.85°
30°	22.5°	22.04°
40°	29.0°	28.82°
50°	35.0°	35.05°
60°	40.5°	40.50°
70°	45.5°	44.80°
80°	50.0°	47.60°

Figure 21.9. Ptolemy's tabulations for refraction from air to water.

Sighting along line AK, the experimenter moves a marker O along quadrant DB of the submerged half of the plaque until it reaches point B, where it appears to lie directly behind K. This indicates that no refraction has occurred. Realigning the line of sight EK at 10-degree intervals from normal AKB, the experimenter moves O until its image I appears directly in line with EK, angle EKA thus representing *i* and angle BKO *r*. Measuring *r* = BKO at each reprise up to *i* = 80 degrees, Ptolemy comes up with the tabulations given in Figure 21.9.

A close look at these tabulations shows that, for the most part, the values for *r* are quite close to those predicted by the modern sine law (sine *i* : sine *r* is constant). Note, however, that while *i* increases uniformly, *r* increases nonuniformly by decrements of 0.5 degrees for every 10-degree interval of *i*. In short, *r* increases by a uniformly decreasing amount, and precisely the same holds for Ptolemy's values for *r* in refraction from air to glass and from water to glass. It is evident, then, that Ptolemy was not tabulating raw data; he was adjusting his observations with the expectation that although *r* might not vary uniformly with *i* (i.e., $r_1 - r_2$ is not constant as *i* approaches 90 degrees), the variation in its variation will be constant (i.e., $r_1 - r_2$ decreases uniformly as *i* approaches 90 degrees).[10]

Having laid the theoretical and analytic foundations for refraction, Ptolemy addresses two more or less practical issues: the effect of atmospheric refraction on astronomical observation and refractive image distortion. Contained in *Optics* 5.23–30, Ptolemy's discussion of atmospheric refraction presupposes that the earth (with center C in Figure 21.10) is surrounded by an envelope of air that is denser than the ether above it. Under these conditions, horizontal

[10] For a more detailed analysis, see A. Mark Smith, "Ptolemy's Search for a Law of Refraction: A Case-Study in the Classical Methodology of Saving the Appearances and its Limitations," *Archive for History of Exact Sciences*, 26 (1982), 221–40.

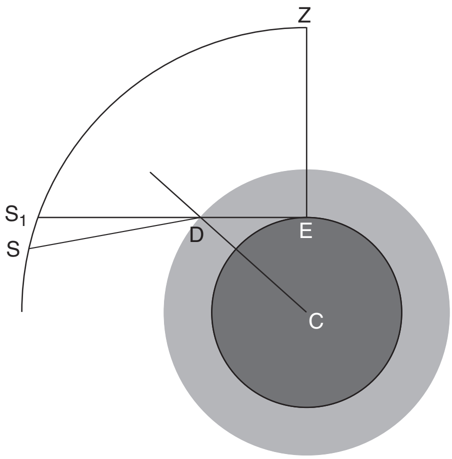

Figure 21.10. Atmospheric refraction according to Ptolemy.

visual ray ED will be refracted away from normal CD toward DS when it passes through the air–ether interface. Whatever celestial body lies at S will therefore appear at S_1, its image having been raised according to arc SS_1. Ptolemy goes on to show that this upward displacement diminishes constantly as the line of sight approaches the zenith (Z), where the visual ray strikes the air–ether interface orthogonally to pass straight through. The promise of this analysis is clear. If we could only determine the amount of upward displacement caused by atmospheric refraction, we could hone our celestial observations accordingly. Unfortunately, Ptolemy concludes, such a determination is impossible until we know the depth of the atmospheric sheath that causes refraction.

As to image distortion, the most obvious example consists in the magnification or diminution of objects viewed through various refractive interfaces. Ptolemy's analysis is as follows. Let AD in Figure 21.11 be a plane refractive interface, E the center of sight, and OO_1 the object. If the medium below AD is denser, then visual rays EB and EC will be refracted toward the normal along BO and CO_1. Since the resulting image will be seen at II_1, it will subtend a larger visual angle (IEI_1) than the object itself would if it were seen directly along EFO and EGO_1. Conversely, by the principle of reciprocity (which Ptolemy makes explicit in *Optics* 5.17), if II_1 were the object and the medium above AD were denser, then II_1 would be seen at OO_1 and would appear smaller because it would subtend a smaller visual angle (OEO_1) than the object itself if seen directly along IBE and I_1CE. As in reflection, therefore, so in refraction, appearances are absolutely deceptive because the object is misperceived in terms of both place and size.

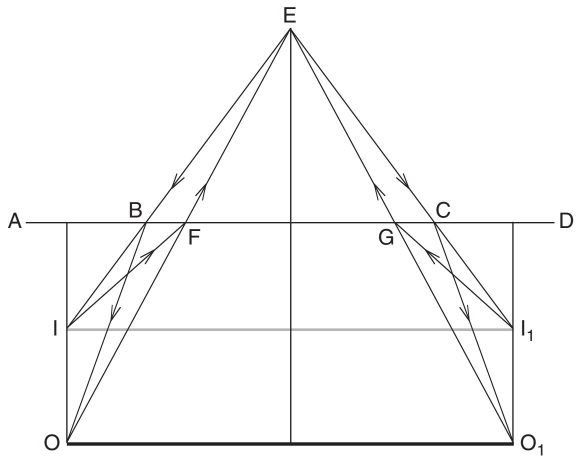

Figure 21.11. Refractive image distortion according to Ptolemy.

THE FATE OF GREEK OPTICS

Much like the *Almagest* in mathematical astronomy, Ptolemy's *Optics* brought the ray analysis of sight to an extraordinarily high level of sophistication. Unlike the *Almagest*, however, the *Optics* was virtually stillborn, attracting little or no apparent interest in its day or for the next several centuries. Whether this indifference to Ptolemy's work was cause or symptom, the period between roughly 200 and 900 CE shows clear signs of having been one of, at best, stagnation and, at worst, regression for the science of optics. As late as the mid-ninth century, in fact, Euclid seems to have served as the authoritative source for ray analysis among such newly Hellenized Arab thinkers as al-Kindi. Only with the rediscovery of Ptolemy's *Optics* in the tenth century, most notably by Ibn Sahl (ca. 950) and Ibn al-Haytham (ca. 1000), did the study of mathematical optics regain and eventually surpass the level it had achieved some eight centuries earlier with Ptolemy.[11]

11 For a brief account of the post-Ptolemaic optical tradition to the time of Ibn al-Haytham, see Smith, *Ptolemy's Theory*, pp. 5–8 and 49–57.

22

HARMONICS

Andrew Barker

INTRODUCTION

The science the Greeks called *harmonikē* was the study of the structures underlying musical melody, their elements and substructures, and the principles governing their organization.[1] The status of harmonics in one of its two main forms was comparable to that of astronomy or geometry; it appears as the last, and perhaps the most intellectually elevated, of the *mathēmata* (mathematical sciences) that Plato prescribes for a philosopher's education (*Republic* 530c–531c), and it was studied and discussed by eminent philosophers and scientists from the fourth century BCE to the end of antiquity. At the high point in the development of the other form it was a fully-fledged science in the Aristotelian mould. In both it was assumed that its principles and their consequences were as unchangeable, as intelligible, and hence as open to scientific inquiry as those of the subjects studied by other genuine sciences.[2]

Both versions of the science originated in the fifth century BCE, or possibly the late sixth. They emerged from two quite different enterprises. One tradition (which I shall call "mathematical") stems from the mathematical and metaphysical reflections of Pythagorean philosophers, for whom the striking mathematical features of certain fundamental musical relations provided an admirable model for the "harmonious" integration of the universe and its contents. Their approach to the subject was fruitfully developed in the fourth century, especially in the hands of Archytas and Plato. The origins of the other, "empirical" tradition lay with practical

[1] Annotated translations of many of the relevant texts are in A. Barker, *Greek Musical Writings, vol. 2: Harmonic and Acoustic Theory* (Cambridge: Cambridge University Press, 1989).

[2] There were also other scientific or quasi-scientific disciplines connected with music, including rhythmics, metrics, and the study of instruments; but metrics typically considers poetic texts in abstraction from any associated music, and the others were never developed to the same extent as harmonics.

musicians, who probably used it primarily in the instruction of their young apprentices.[3] So far as we can tell, none of them had serious philosophical or scientific pretensions until the late fourth century, when Aristoxenus embarked on a complete reconstruction of the discipline.

It may be worth stating here that there is very little surviving written music to provide supporting evidence of actual practice, although new fragments of score are still being discovered.[4] The notations seem to have originated in the fifth century BCE and were further developed in the fourth. So far as we can tell, musicians never played from scores in performance, but they may have assembled collections of them as reminders of their repertoire. Scores were also sometimes memorialized by inscription in stone, and in at least one case the music of a song was inscribed on a gravestone. Notational symbols do not appear in the surviving theorists' writings until the Roman period, though Aristoxenus refers to their use by some of his predecessors.

Writings in both traditions of harmonics continued to be produced in the Hellenistic period, but almost nothing of their contents survives. The first five centuries CE, on the other hand, have left us a substantial number of relevant and, in some cases, sophisticated texts; and, though empirical harmonics fared less well than its mathematical cousin, both were passed down into the medieval West (in defective forms) through Boethius' and Martianus Capella's Latin paraphrases of work by Nicomachus and Aristides Quintilianus. The Greek texts sprang back into life when they reached the musicologists of the Renaissance, where they provoked controversies about tuning and helped to fuel the enterprise of recreating ancient Greek culture.

This essay will focus much more on the theorists' methods and objectives than on the substance of their analyses, but I shall need to refer to some of the musical structures they discuss.[5] The theorists' first points of reference are the musical concords (*symphōniai*), of which only three were recognized within the span of an octave: the fourth, the fifth, and the octave itself.[6] The octave was regularly analyzed into two subsystems with a small gap between them, one running from the lowest note (called *hypatē*) to a note a fourth above it (*mesē*), the other down from the highest note (*nētē*) to the note a fourth below it (*paramesē*). Since an octave is the sum of a fourth and a fifth, *mesē* is also a fifth below *nētē* and *paramesē* is a fifth above *hypatē*. The gap between *mesē* and *paramesē* is thus the difference between a fourth

[3] There is a report by the late fourth-century scholar Phaenias (frag. 32 Wehrli, recorded at Athenaeus 348d) about the activities of Stratonikos, a prominent musician in that century's earlier decades.

[4] For a collection of the surviving scores see E. Pöhlmann and M. L. West, *Documents of Ancient Greek Music* (Oxford: Oxford University Press, 2001).

[5] For fuller treatments see e.g., Barker, *Greek Musical Writings*, vol. 2, 11–17; M. L. West, *Ancient Greek Music* (Oxford: Clarendon, 1992), pp. 160–89.

[6] *Symphōnia* was distinguished from *diaphōnia* (discordance) by the way in which its two components seemed to blend together in the hearer's perception, so that they were heard as a single unified sound. It was this that gave them their status as paradigms of the harmonious integration of many things into one.

and a fifth, and this defines the interval called the *tonos*, our "tone" or "major second."

The relations between the notes bounding these intervals were regarded as unchangeable, and the notes were called "fixed" or "standing" notes. To complete an octave scale, two other notes were inserted between the boundaries of the fourths at the top and the bottom, forming two subsystems of four notes each, called tetrachords; the lower is called the tetrachord *mesōn* and the higher the tetrachord *diezeugmenōn*. In any regularly formed scale, the pattern of intervals in each of the two tetrachords was the same, but the Greeks had many more types of scale than we do, differing principally in the placing of the intermediate notes of the tetrachord; because their positions were variable they were called "moveable" notes.

Scales were classified into three groups or "genera": enharmonic; chromatic; and diatonic. In an enharmonic scale the intermediate notes were very close to the bottom of the tetrachord, which therefore consisted of two tiny intervals (about a quarter-tone each) lying below a large one (roughly a ditone).[7] In a chromatic tetrachord the intervals at the bottom are rather larger; their sizes vary, but they always occupy jointly less than half the span of the fourth. The pair of small intervals at the bottom of an enharmonic or a chromatic tetrachord is known as the *pyknon*, meaning something "compressed" or "compact." The diatonic genus is defined by the fact that the corresponding pair of intervals at the bottom of its tetrachords always adds up to at least half the span of the fourth, so that it no longer counts as a *pyknon*; and it never amounts to more than about one and a half tones, leaving roughly a tone for the highest interval. As the vagueness of these specifications will have suggested, not only did enharmonic, chromatic, and diatonic scales differ from one another, but scales in any one genus could also take subtly different forms; and theorists argued incessantly about the minutiae of their structures.

Perhaps the most significant of the other questions the theorists debated were those dealing with the ways in which identically structured scales could be placed at different levels of pitch, to form a set of what we would call "keys," in Greek *tonoi*.[8] Theorists' conceptions of the *tonoi* are notoriously hard to grasp, and there were (and still are) controversies about their number, the distances between them, the relations between them, and the forms or "species" of the octave, and the musical purposes for which they were designed.[9] I shall not tackle these or any other technical issues here, but shall refer to notes, intervals, concords and discords, tetrachords, scales, the

[7] Depending on whether one was a mathematical theorist or an empiricist, a fourth amounted either approximately or exactly to two and a half tones.

[8] Here the word *tonos* refers to something very different from the interval of a "tone," which I mentioned above. For the word's various musical meanings see e.g., Porph. *In Ptol. Harm.* 82.1–14.

[9] The seven species are the various arrangements of the intervals that will appear in octave scales in any genus, when each begins from a different note of the scale.

three melodic genera, and the *tonoi* (just occasionally).[10] I shall deal separately with the classical and the later periods ("classical" meaning the fifth and fourth centuries BCE, with a little overspill at each end); and in each case I shall devote separate sections to the empirical and the mathematical versions of harmonic science.

EMPIRICAL HARMONICS IN THE CLASSICAL PERIOD

Some of the fifth-century musicians from whose teaching-aids this tradition begins probably wrote essays on the subject, but not even a fragment of these survives. We have to rely on comments by later writers, especially Plato and Aristoxenus – both of them hostile or contemptuous witnesses whose evidence must be cautiously handled. But the overall purpose of the endeavor seems clear. Musicians had previously taught mainly by example and demonstration, but in the fifth century the scale systems became increasingly diverse and complex, sometimes differentiated by tiny nuances of tuning, and musical styles developed that modulated freely between the various attunements and scales.[11] Students could hardly be expected to master their intricacies in the absence of precise definitions and formulations of their patterns of notes and intervals and the distinctions and relations between them.

Theorists (whom I shall call "the *harmonikoi*," as Aristoxenus does) set out to fill the gap, and above all to specify, in exact, quantitative terms, the intervals that figured in each of the systems and the sequences of intervals that gave them their structure; and for this purpose they needed to identify a determinate unit of measurement. They used no special equipment, but simply listened attentively to notes played on two strings of an instrument such as a lyre, gradually adjusting one string's tension until the pitches differed by the smallest amount their ears could detect. The ear was the sole judge, and inevitably they disputed about the point at which this state of affairs had been reached.[12] In order to turn this interval into a unit of measurement, it was compared with the *tonos* – the interval by which the perfect fourth (*dia tessarōn*) differs from a perfect fifth (*dia pente*). We do not know how any of the *harmonikoi* set about determining how many times the "minimal" interval will fit into a *tonos*; we know only that it came to be identified as a quarter-tone.[13]

[10] On everything to do with the extension of the scale system to two octaves, associated developments in musical notation, *tonoi*, and modulation see especially S. Hagel, *Ancient Greek Music: A New Technical History* (Cambridge: Cambridge University Press, 2010), a technically demanding but very rewarding book that sets many topics in Greek musicology in a new perspective. Cf. also e.g., Barker, *Greek Musical Writings*, vol. 2, 17–27.

[11] On the musical innovations of this period see e.g., West, *Ancient Greek Music*, pp. 356–72; E. Csapo, "The Politics of the New Music," in P. Murray and P. Wilson (eds.), *Music and the Muses: The Culture of "mousikē" in the Classical Athenian City* (Oxford: Oxford University Press, 2004).

[12] Plato *Rep.* 531a4–b1.

[13] This can be inferred from e.g., Aristox. *El harm.* 53.2–9. A passage at 14.21–5 shows his partial agreement with the *harmonikoi*: he states that the quarter-tone is the smallest interval that can be either produced by the voice or detected by the ear. But he does not treat the quarter-tone as a unit of

The *harmonikoi* could then set about their principal task: producing analyses of the kind recorded by a much later writer, Aristides Quintilianus (perhaps third century CE), in a collection of "very ancient scales."[14] By later standards their structures are strangely irregular; but all the intervals they include are quarter-tones or their multiples, and all are based on a substructure characteristic of the enharmonic genus. The *harmonikoi* described scale structures verbally but also in diagrams, as Aristoxenus bears witness. So far as we can tell these were simply lines marked off at distances representing quarter-tones, twenty-eight of them in all, forming a grid on which the positions of the notes were marked.[15] Aristoxenus also criticizes their use of a form of musical notation (*parasēmantikē*) that shows only the sizes of the intervals between the notes.[16] His most fundamental objection was that the *harmonikoi* aimed to provide only an empirical record, without any explanation or investigation of underlying principles.

Aristoxenus set out to remedy this state of affairs. In his *Elementa harmonica*[17] he undertakes to reconstitute empirical harmonics as a science in the Aristotelian mould, taking the *Posterior Analytics* as his principal guide.[18] Its essential starting point is data provided by the ears of a listener, who must be familiar with every musical style and whose hearing has been "trained to accuracy." Aristoxenus repeatedly insists that music exists only in the domain accessible to hearing, and that harmonics is exclusively concerned with musical sounds and structures in the guise in which we perceive them. He therefore dismisses the mathematical approach as irrelevant to harmonics, since the items whose interrelations it investigates are not sounds but only the inaudible and hence nonmusical material events underlying them. After listening carefully to a great deal of music and analyzing the structures on which each type is built, the scientist's next task is to abstract from the data the principles by which the structures are governed, principles which effectively define the nature (*physis*) of the science's subject, *melos* (melody) or *to hērmosmenon* (that which is attuned).[19] Such principles are stated in passages near the ends of books 1 and 2. From these principles,

measurement; several of the systems he describes involve intervals which are not among its multiples, such as one-third of a tone.

[14] Arist. Quint. *De mus.* 18.5–19.10; his immediate source was probably Aristoxenus. He goes on to say that they are the *harmoniai* or patterns of attunement to which Plato refers in book 3 of the *Republic* (398d–399c).

[15] Aristoxenus refers also to something he calls the *katapyknōsis* ("compression") of the diagram, apparently the procedure of bringing a collection of scales together in the same picture, set in the same range of pitch, to make it easier to compare one scale with another.

[16] *El. harm.* 39.4–40.24. This is the earliest reference to a notation in our sources. Whether their notation was the one familiar from later treatises and the surviving scores is still under debate.

[17] In its surviving form this work is in three books, of which the third is incomplete. Book 1 probably comes from an earlier or more elementary treatise on harmonics than books 2 and 3.

[18] He was admirably qualified for the task. He was initially trained for a musical career, but then turned to philosophy, learning first from eminent Pythagoreans and later studying and teaching with Aristotle in the Lyceum.

[19] Compare Aristotle's account of this process of abstraction at *An. Post.* 99b17–100b17.

finally, Aristoxenus derives logical demonstrations (*apodeixeis*) proving, for a fundamental series of propositions, that the facts *must* be as he has described them and explaining why (book 3).

Musical melody, Aristoxenus says, is always something coming into being, at no time existing as a whole.[20] The main focus of his analyses is therefore on scalar or melodic sequences conceived as developing in time; here again he differs from the mathematical theorists, who typically described static structures corresponding (in principle) to the patterns of attunement set up on an instrument prior to performance. In a very influential passage (8.13–10.20), Aristoxenus describes the way in which the singing voice is perceived as "moving" from pitch to pitch, distinguishing this form of movement from that of speech; he devotes thoughtful discussions to the concepts of "continuity" (*to syneches*) and succession (*to hexēs*);[21] and all the *apodeixeis* of book 3 are designed to prove propositions about which intervals can follow which others, and in which musical contexts.

In the course of books 1 and 2 Aristoxenus discusses various fundamental items such as pitch, notes, and intervals; and he presents detailed, quantitative analyses of six forms of scale (one enharmonic, three chromatic, and two diatonic), expressed in units of a tone and its fractions. But these, he says, are only the most familiar examples; there is a potentially unlimited number of others which are equally recognizable as acceptable instances of scales in these genera, since the moveable notes can lie anywhere in the unbroken pitch continuum between the boundaries he has assigned to them. Further, the perceived musical identity of a note can remain the same regardless of small variations in its distance from its neighbors; and hence the sizes of the intervals of scales are not fully determinate. But the sciences, as Aristoxenus conceives them, cannot deal with anything indeterminate, and it follows that quantitative analyses of musical structures are both musically and scientifically unsatisfactory.[22] The determinate aspects under which the science can and should investigate notes and intervals are "their functions [*dynameis*], their forms [*eidē*], and their positions [*theseis*]."

The crucial concept here is that of *dynamis*, literally "power" or "potential." A note's musical identity is not constituted by its pitch or by the sizes of the intervals surrounding it, but by the power it exercises, in a particular melodic context, over the routes that a melodic sequence can follow. An analogy with more modern musical conceptions gives us the easiest

[20] "The understanding of music comes from two things, perception and memory; for we have to perceive what is coming to be and remember what has come to be. There is no other way of following the contents of music" (38.33–39.3). The reliability of the scientist's memory is as important as that of his ear.

[21] See 27.15–29.1, 52.34–53.32. The main issue is the sense in which a sequence such as a scale can be said to be "continuous," allowing no additional notes to be inserted in it, and a note can be said to be the direct "successor" of another, given that there is always space left in the continuum of pitch where intervening notes could be placed.

[22] See e.g., 39.26–40.23, 68.13–69.28, and cf. Plato *Phileb.* 16c–e.

way of understanding this idea. We learn little about a note's identity if it is specified merely by reference to its pitch, but much more when it is characterized by its "function," for instance as the "tonic" or the "leading note" of the prevailing scale. In practice its pitch may vary a little, but so long as it retains the same function, it fixes the identities (though again, not the exact pitches) of other notes in the same environment; it influences the course that the current melody can follow and the way in which it will be musically interpreted. It is under this aspect that melodic phenomena can become objects of scientific study.

Aristoxenus' work in harmonics was typically ignored or dismissed by later philosophers, and its status as a pioneering recreation of the discipline as an Aristotelian science was rarely recognized. But among practical musicians it was hugely influential right through to the end of classical antiquity, providing much of the language and the conceptual equipment through which they discussed and taught their art.

MATHEMATICAL HARMONICS IN THE CLASSICAL PERIOD

This form of the science began from a set of simple observations, or perhaps a nugget of instrument makers' traditional lore. When the relevant dimension of a sounding object stands in the ratio 2:1 to that of another which is otherwise identical, the former gives a pitch an octave below the latter; when the ratio is 3:2 the interval between them is a perfect fifth; and when it is 4:3 the interval is a perfect fourth.[23] To a reflective mind these results are remarkable: the ratios are the simplest possible; they use only the first four numbers, which add up to the perfect number ten (forming the Pythagorean "*tetraktys* of the decad"); the intervals with which they are correlated are the three fundamental *symphōniai*; and the *symphōniai*, as we have seen, are not only the foundations of musical organization, but are directly characterized by their fusion of two sounds into a unified whole. Thus they are paradigmatic instances of the harmonious integration of different items into one, and the basis of this lies in the relations existing between numbers. Here we arrive squarely on Pythagorean territory: it was the early Pythagoreans who first used these musical ratios in the service of philosophical theorizing.[24]

[23] In the simplest case, half the length of a stretched string of even thickness and tension sounds an octave above the whole string. Curiously, our sources on the early theorists give the case of a string no special prominence, and the first writer to discuss the causes of pitch variation (Archytas frag. 1) does not even mention it. Our earliest allusion to the instrument based on it, the monochord, dates from the end of the fourth century.

[24] Later writers commonly attribute the "discovery" of the musical ratios to Pythagoras himself; they could be right, but there is nothing to prove it. Some of their accounts are unbelievable because physically impossible (notably the tale of the "harmonious blacksmith" told by Nicomachus in *Harm.* 6 and repeated word for word by Iamblichus *Vit. Pyth.* 26), and none of the others seem to rest on good authority. Even our earliest source for the attribution, Xenocrates frag. 87, inspires little confidence.

The dubious credentials of much of our alleged information make the topic of early Pythagoreanism notoriously treacherous.[25] Here it must suffice to say that the Pythagoreans are credited with a plethora of experiments with sounding bodies of different sorts, all of which confirmed the consistent importance of these ratios.[26] The experiments' diversity suggests that they were trying to show that the ratios remain the same regardless of the nature of the bodies involved; what mattered were the numbers and their association with attributes of the sound itself, independent of the method of its production. By the early fourth century, if no earlier, the Pythagoreans had developed the (essentially correct) theory that a sound is constituted by a movement in or of the air, and that its pitch is determined by an attribute of this movement that varies quantitatively. In the dominant version of the theory this was its speed of transmission, while for others it was the amount of time between the impulses created by successive impacts of an object on the air.[27]

One important facet of Pythagorean theory is illuminated by the surviving fragments of Philolaus, from around the end of the fifth century. Here the universe is made up of unlimiteds and limiters, integrated with one another to form a *kosmos* by the action of *harmonia*; in a later fragment, *harmonia* is clearly understood in musical terms, identified with the octave and its internal organization in a particular form. One of Philolaus' main aims was to show the symmetry of this internal structure, first in musical terms, as a pattern of overlapping fourths and fifths perfectly balanced around a central point, and secondly in numerical terms, representing the fourth, fifth, and octave by their ratios, 4:3, 3:2, and 2:1, and introducing also the ratio of the tone,

[25] The starting point for all modern research on the matter is W. Burkert, *Weisheit und Wissenschaft: Studien zu Pythagoras, Philolaus und Plato* (Nuremberg: H. Carl, 1962), translated by Edwin L. Minar, Jr. as *Lore and Science in Ancient Pythagoreanism* (Cambridge, MA: Harvard University Press, 1972), which gave reasons for mistrusting the great bulk of the putative evidence, leaving Aristotle's testimony as almost the only survivor (cf. C. Huffman, *Philolaus of Croton: Pythagorean and Presocratic* (Cambridge: Cambridge University Press, 1993), pp. 17–33; C. Huffman, *Archytas of Tarentum: Pythagorean, Philosopher and Mathematician King* (Cambridge: Cambridge University Press, 2005), pp. 91–100). A more generous view, particularly in respect of the Platonist sources, is proposed in C. H. Kahn, *Pythagoras and the Pythagoreans: A Brief History* (Indianapolis, IN: Hackett, 2001); a much more ambitious rehabilitation of the later traditions had already been put forward in L. J. Zhmud, *Wissenschaft, Philosophie und Religion im frühen Pythagoreismus* (Berlin: Akademie Verlag, 1997). The matter is still under debate.

[26] See e.g., Theo Smyrn. 56.9–61.17.

[27] Dependence on speed was probably the earlier idea: propounded by Archytas (frag. 1), repeated with variations by Plato (*Tim.* 67b, 80a) and Aristotle (*De an.* 420a–b) and mentioned by Aristoxenus (*El. harm.* 32.23–6). The thesis that a sound is a movement of air is found also in the work of several non-Pythagorean Presocratics, as reported in Theophrastus *De sensu*, but they do not seem to have allied it with a theory of pitch. Our earliest explicit sources for the impact theory come in the late fourth or early third century; see the introduction to [Eucl.] *Sect. can.*; Porph. *In Ptol. Harm.* 30.1–31.21 (quoting a Heraclides who may be Heraclides of Pontus); [Aristotle] *Prob.* 19.39; cf. [Aristotle] *De audib.* 803b–804a. But Porphyry's report on a curious Pythagorean procedure, mentioning Archytas as his source, may best be interpreted as presupposing the same view; see e.g., Barker, *Greek Musical Writings*, vol. 2, 35 n. 29.

9:8.[28] Mathematically, the identification of the crucial central point was problematic. The interval of a tone in the middle of the system cannot be divided into two intervals with equal sub-ratios, at whose common boundary the central point would lie.[29] Evidence in two passages of Boethius, both probably based on a lost treatise by Nicomachus, suggests that Philolaus was aware of this unsatisfactory aspect of his system, and attempted to remedy it by deploying numerical measurements in a different way.[30] This involved representing intervals both as ratios and as "distances" or quantities represented by individual numbers; the mixture of the two approaches now appears mathematically unintelligible, but it served the purpose of allowing Philolaus to divide a tone into two equal parts and identify the midpoint of the octave.[31] Whether or not this report is authentic, it is plausible in terms of what we know of the development of Pythagorean symbolism, with Philolaus' analysis of the octave's structure closely tied to cosmological speculations. While the starting point for his analysis must have been a real musical system, there is nothing to suggest that he was concerned with the investigation of such systems for their own sake or with a practical musical outcome in mind.

We know of only one mathematical theorist of the classical period who designed his investigations for the analysis of practical music: Plato's Pythagorean contemporary Archytas. Socrates' denunciation of Pythagorean theorists was very probably directed at Archytas personally; we also have more detailed evidence from Ptolemy, who describes Archytas as "the most dedicated to music of all the Pythagoreans" and tabulates and discusses his analyses of three systems of attunement, one in each of the genera.[32] They take the form of "divisions of the tetrachord," setting out the ratios of the intervals of an enharmonic, a chromatic, and a diatonic tetrachord, as do many later writers; but this is enough to establish the ratios of a system spanning a whole octave, since its two tetrachords are identically structured, and the interval separating them is

[28] This can be calculated by simple arithmetic, since the tone is defined as the difference between the fifth and the fourth. Philolaus does not in fact use the term *tonos*; he always refers to the interval as *epogdoos*, that is, "in the ratio 9:8."

[29] This follows from a theorem subsequently proved by Archytas.

[30] Boethius *Inst. mus.* 3.5 = H. Diels and W. Kranz, *Die Fragmente der Vorsokratiker* (7th edn; Berlin: Weidmannsche Verlagsbuchhandlung, 1961) ("DK"), 44A26, test. A26; Huffman, *Archytas of Tarentum*; Boethius *Inst. mus.* 3.8, printed as an appendage to DK44B6 and in Huffman, *Archytas of Tarentum* as frag. 6b. The authenticity of the reports is in serious doubt; my own view is that they have better credentials than is usually admitted, and that their strictly mathematical peculiarities and incoherencies may weigh in favor of an early date. For a judicious assessment of the arguments, see Huffman, *Philolaus of Croton*, pp. 364–74.

[31] For details see A. Barker, *The Science of Harmonics in Classical Greece* (Cambridge: Cambridge University Press, 2007), pp. 271–83.

[32] Plato *Rep.* 530c–531c. On the passage's connection with Archytas see Huffman, *Archytas of Tarentum*, pp. 414, 423–34, and on the relations between Archytas and Plato see pp. 32–42. Huffman's book should be consulted on everything to do with Archytas. For Ptolemy see Ptol. *Harm.* 1.13, with a critique in 1.14.

invariably a tone (ratio 9:8).[33] The divisions have various peculiarities, most notably that the lowest interval in each of them is the same, 28:27; this feature is unique to Archytas, among the Greeks. He also associates them with what he calls "the three means in music": the arithmetic; the geometric; and the harmonic. Modern discussions of these divisions are complex, but there seems to be a consensus that they are based on a rather subtle combination of empirical observations and mathematical principles, and are probably quite accurate representations of systems in practical use.[34] This fits well with our broader picture of Archytas' work, as his studies in astronomy, optics, and perhaps mechanics show genuine scientific interest in the behavior of things in the material world. As a pioneering and very distinguished mathematician, he applied his mathematics to real-world investigations.

Some other writers of the period, including Aristotle and other Peripatetics, also thought of mathematical harmonics as applicable to the study of musical practice. But they differ from Archytas: they use only a small number of very elementary propositions; they make no attempt to develop the science itself in any way; and the only division of the tetrachord presupposed by their discussions is a diatonic system whose ratios had originally been worked out for quite different purposes.[35] None of them tried to pursue Archytas' line of research either by modifying his conclusions or by supplementing them with analyses of other variants of the diatonic, chromatic, and enharmonic.

The first theorist to specify the ratios of the diatonic system in question was probably Philolaus, but it was Plato who made it famous and gave it the context for most of its subsequent career. The task he sets for harmonics in the *Republic* is to "ascend to problems, to investigate which numbers are concordant and which are not, and in each case why" (*Rep.* 531c). It is not to study notes and intervals; the "problems" are mathematical, not musical; and the explanations of the concordance or nonconcordance of numbers will be worked out from mathematical principles.[36] Plato's Socrates treats all

[33] Ptolemy provides tables spanning the full octave in *Harm.* 2.14, including his own analyses together with those of Archytas and three other theorists.

[34] See especially R. P. Winnington-Ingram, "Aristoxenus and the Intervals of Greek Music," *Classical Quarterly* 26 (1932), 195–208; Barker, *Greek Musical Writings*, vol. 2, 46–52; Huffman, *Archytas of Tarentum*, pp. 410–23; Barker, *Science of Harmonics*, pp. 292–302.

[35] That is not to say that it failed to correspond, even approximately, to any system in practical use. It can be attuned on a stringed instrument by the simplest of methods and was almost certainly in common use by musicians.

[36] For fuller discussions of the passage see particularly A. Meriani, "Teoria musicale e antiempirismo," in M. Vegetti (ed.), *Platone, la Reppublica: traduzione e commento*, 7 vols. (Naples: Bibliopolis, 2003), vol. 5: *Libri 6–7*, pp. 565–602; cf. Barker, *Science of Harmonics*, pp. 314–18, where I argue, inter alia, that the adjective *symphōnos*, "concordant," does not have its technical sense here, but refers to a wider range of musically (mathematically) acceptable relations. On Plato's similar treatment of astronomy see e.g. A. P. D. Mourelatos, "Plato's 'Real Astronomy': *Republic* 527d–531d," in J. P. Anton (ed.), *Science and the Sciences in Plato* (Buffalo, NY: Caravan Books, 1980), pp. 33–73, and on both astronomy and harmonics see I. Mueller, "Ascending to Problems:

other forms of harmonics as trivial and pointless; if it takes a purely mathematical form, on the other hand, it becomes "useful in the quest for the fine and the good"; it is the last of the five mathematical disciplines essential to the philosopher's education.

The details come into play in his analysis of the structure of the World Soul in the *Timaeus* (34b–36d). This extraordinary construction amounts, in musical terms, to the specification of the ratios in a diatonic scale spanning four octaves plus the interval of a sixth, a much larger range than anything used in Greek practice. The standard pattern of concords is preserved, as are the relations between the tetrachords; and these are divided in accordance with the diatonic system mentioned above. But Plato says nothing here about musical practice, and uses no musical language in his description of the process of division. It is instead seen as the result of actions of the *dēmiourgos*, the divine "craftsman," cutting divisions on the strip of substance according to a succession of numerical ratios before bending the strip into circles and setting them in motion. Thus, for Plato the value of music lies in its kinship with the cosmic order and in the help this enables it to give us in the enterprise of retuning our souls. Plato's attitude to harmonics was massively influential among philosophers in later antiquity.[37]

There are uncertainties about authorship and date of the last important text in mathematical harmonics that survives from classical times, the pithy little treatise called *Sectio canonis*; I believe that the whole essay was written at a date around 300 BCE.[38] The writer sets out to prove the most basic propositions of mathematical harmonics on solid mathematical grounds, in a set of sixteen theorems in Euclidean style which form the core of the treatise. Two additional propositions are concerned specifically with notes in the enharmonic genus. At the end there is a two-stage "division of the canon," which shows how a single string can be divided into lengths corresponding to all the notes of a two-octave diatonic scale, on the basis of the ratios that the first sixteen theorems have established.[39] All this

Astronomy and Harmonics in *Republic* VII," in Anton (ed.), Science and the Sciences in Plato, pp. 103–22.

37 Even Aristotle seems to have been mightily impressed by it at one stage of his career; see frag. 908 Gigon (= frag. 47 Rose, frag. 25 Ross), quoted at [Plut.] *De mus.* 1139B. For comments on the fragment and its context see Barker, *Science of Harmonics*, pp. 329–38.

38 For arguments to this effect see Barker, *Science of Harmonics*, pp. 364–70; further considerations are added in the course of my discussion of the work itself, pp. 370–410. The introduction to A. Barbera, *The Euclidean Division of the Canon* (Lincoln, NE: University of Nebraska Press, 1991) is a masterly study of the text's very complex history and the problems about when and by whom it was written. Porphyry quotes a large part of it and attributes it to Euclid, but no author is credited by the various other ancient writers who quote or paraphrase excerpts. Some of the manuscripts assign it to Euclid but others to a certain Cleonides, and in others it is anonymous. Cleonides is unlikely; Euclid is a more plausible candidate, since the *Sectio* has obvious affinities with his work, but it is more probable that it was written by one of his students or associates. Certainty seems unattainable, and we must be content with "anonymous."

39 It has the same basic structure as the scale in the *Timaeus*, with the 4:3 ratio of its tetrachords divided as 9:8x9:8x256:243, i.e., tone, tone, and *leimma*. The string is that of the instrument called the

mathematical material is preceded by an introduction, in which the author expounds a series of propositions in physical acoustics. He believes that sounds are movements of the air, caused by impacts; those formed of more closely packed sequences of movements are higher in pitch, so the pitch relations between notes can be expressed as numerical ratios. He then identifies the three types of ratio – multiple, epimoric, and epimeric – and argues that the ratio of every musical concord must be either multiple or epimoric, a principle that is essential to his reasoning in the theorems.[40]

None of the mathematics in propositions 1–9 is original to this author, nor are any of the conclusions reached in propositions 10–18. He merely integrated a body of well-known elementary propositions of harmonics with a set of established mathematical theorems. Later writers drew frequently on selections of the work's propositions and borrowed its division of the canon (with various modifications); in a large sector of mathematical harmonics its proofs and principles, despite their flaws and weaknesses,[41] became part of a regular orthodoxy.

FOURTH-CENTURY ATTITUDES TO THE TWO HARMONIC SCIENCES

There was no consensus in this period about the status of the two styles of harmonic theory, or about the relations between them. The first writer to comment on their credentials is Plato, who dismisses both of them as relying on the evidence of human ears, whereas the only worthwhile testimony is that of reason. Aristotle treats mathematical and empirical harmonics as legitimate enterprises with different but complementary roles, the latter providing the facts and the former their explanations, and both relating to phenomena in the world of the senses rather than metaphysical constructions. Both he and other early Peripatetics borrow ideas from mathematical harmonics, sometimes to fuel discussions of music itself, and sometimes in other domains. By contrast Aristoxenus insists on the sole authority of empirical harmonics, in the new form he had given it, and labels the mathematical theorists' work as irrelevant: it may perhaps be acceptable in the field of physical acoustics, but has nothing to do with harmonics. The *Sectio canonis* says nothing explicit to suggest that there is any kind of harmonics apart from its own, but some of its later propositions (e.g. that the

monochord, though it is not named here. On this instrument and related devices see D. E. Creese, *The Monochord in Greek Harmonic Science* (Cambridge: Cambridge University Press, 2010).

40 Here an epimeric ratio is simply any ratio that is neither multiple nor epimoric. Some later writers give more complex classifications.

41 Specifically the invalidity of prop. 11, and the unconvincing reasoning about the ratios of concords, which entails the octave plus a fourth being a discord. So far as I know, subsequent writers never mention this reasoning, but almost all (except Ptolemy) accept the principle and very rarely do they attempt to give it more secure foundations.

octave amounts to less than six tones, that the interval of a tone cannot be halved, and that the intervals in an enharmonic *pyknon* cannot be equal) have no intelligible role in the context except as polemical points against the empirical theorists, and so the work clearly focuses on a mathematical approach as superior.

Despite the general acceptance of mathematical harmonics in Peripatetic circles, it came under sustained attack from Aristotle's immediate successor, Theophrastus. But his arguments seem to have made little or no impact in his own time, and his contention that pitch is a qualitative attribute of sounds resurfaces only very occasionally amongst later writers. They almost always adopted the position implicit in the *Sectio canonis*, emphasizing the points at which their results conflict; as most were working in the mathematical tradition, readers may get the impression that they concluded that the empirical, "Aristoxenian" approach is simply wrong. But this impression is only partly correct; the truth, as we shall see, is a good deal more complicated.

THE LATER CENTURIES

Only a few scraps of evidence about harmonics in the Hellenistic period survive.[42] Discussions of the *Timaeus* continued; there was some adaptation of Aristoxenus' ideas; and the versatile third-century scholar Eratosthenes discussed musical issues in his *Platonicus*, working out new sets of ratios for the enharmonic and chromatic genera. Both forms of harmonics attracted attention from Stoic philosophers, notably Diogenes of Babylon.[43] But, so far as we can tell, these scholars added nothing substantial to the content of harmonic theory.

By the end of the period we can see signs of a more reflective attitude, which surfaces first in passages that Porphyry quotes from the theorists Ptolemaïs of Cyrene and Didymus.[44] I shall concentrate on Ptolemaïs (the

[42] At the beginning of his commentary on Ptolemy, Porphyry lists a group of theorists who apparently belonged to Hellenistic times; but he says that even in his own day they survived only "as names" (*Harm.* 3.1–12).

[43] Most of the evidence on Diogenes comes from Philodemus' criticisms of him in his *De musica*; see particularly cols. 135.23–137.27, 152.2–30 Delattre, with the notes in D. Delattre (ed.), *Philodème de Gadara, Sur La Musique*, (Paris: Les Belles Lettres, 2007), and cf. e.g. cols. 84.42–85.16, 109.28–45, 144.19–145.30. For further discussion of Diogenes' theories see A. Barker, "Diogenes of Babylon and Hellenistic Musical Theory," in C. Auvray-Assayas and D. Delattre (eds.), *Cicéron et Philodème: la polémique en philosophie* (Paris: Editions Rue d'Ulm, 2001), pp. 353–70; and on harmonics in the wider domain of Stoic ethics see A. A. Long, "The Harmonics of Stoic Virtue," *Oxford Studies in Ancient Philosophy* supp. vol. (1991), 97–116, reprinted in A. A. Long, *Stoic Studies* (Cambridge: Cambridge University Press, 1996), pp. 202–23.

[44] Porph. *Harm.* 22.22–28.26. For a fuller discussion of these passages see A. Barker, "Shifting Conceptions of 'Schools' of Harmonic Theory," in M. C. Martinelli (ed.), *La Musa dimenticata* (Pisa: Edizioni della Normale, 2009), pp. 165–90, though I now withdraw my contention (p. 184) that the text at *Harm.* 25.25–26.1 is corrupt.

only known female musical theorist of antiquity); we have no definite information about her life or dates, but I would tentatively place her in the first century BCE.[45] Her work employs the question-and-answer form, which suggests that it was intended as a teaching text at a fairly elementary level. Ptolemaïs classified harmonic theorists into five groups, distinguished by their epistemological standpoints and specifically by the roles they assign to the "criteria" of reason and sense perception.[46] In brief they are: extremist Pythagoreans who reject sense perception altogether; practical musicians who ignore theory; those like Aristoxenus who allow reason and perception roughly equal merit; Pythagoreans who recognized that reasoning must take some perceptual information as its starting point; and some of Aristoxenus' followers who favor perception but allow reason an auxiliary role. This fine-grained set of distinctions appears unprecedented.[47] A focus on the criteria became canonical in the Roman imperial period, though most later writers content themselves with a rather coarse contrast between champions of reason (usually called "Pythagoreans") and champions of sense perception ("Aristoxenians"), with the finer nuances being lost.

EMPIRICAL HARMONICS

This version of the discipline had mixed fortunes in the centuries after Aristoxenus. Among intellectuals who were not specialists in harmonics his reputation remained very high. But although a number of treatises in empirical harmonics survive from the imperial period, none of them attempts to develop it in new ways. The most substantial texts are: the *Introduction to Harmonics* attributed to Cleonides (perhaps from around 100 CE); the first book of Aristides Quintilianus' *De musica* (probably third century); the short "handbooks" of Bacchius and Gaudentius[48] (probably fourth century); and the three even shorter essays known collectively as Bellermann's *Anonymi* (dates anywhere between the second century and the sixth).

The smell of the schoolroom hangs about most treatises of this kind, for example the manuals by Bacchius and Gaudentius. This helps to explain why Aristoxenus' reputation remained so firmly established. But the

45 For an attempt to reconstruct the outlines of her life (though inevitably some of it is speculative) see F. R. Levin, *Greek Reflections on the Nature of Music* (Cambridge: Cambridge University Press, 2009), pp. 229–40. Chapter 7 of Levin's book (pp. 241–95) examines Ptolemaïs' work at length, bringing it into connection with Aristoxenus; it also includes useful comments on Didymus. This Didymus is probably, but not certainly, the scholar and musician of that name who lived in the time of Nero (*Suda* Δ 875).

46 Porph. *Harm.* 25.3–26.5.

47 Didymus gives a fuller account, though Porphyry (*Harm.* 25.1–26.29, 27.17–28.26) only includes his discussions of Ptolemaïs' first three groups. His contributions are thoughtful, but as Porphyry hints (25.1–4), he is dependent on Ptolemaïs and proceeds by elaborating what she had already said. He may indeed have been Porphyry's source for the Ptolemaïs quotations.

48 Although some passages in Gaudentius' work adopt the "mathematical" stance.

surviving texts have real importance for modern scholars for at least two reasons. First, they reveal aspects of Aristoxenus' theories that have not survived in his own words. Second, they show that Aristoxenian propositions were still being reformulated and reinterpreted in slightly different ways. Although the discipline had ceased to be a focus of intellectual debate, it was not entirely fossilized.

It received more interesting treatment at the hands of people working in the mathematical tradition. In the second century even Ptolemy, the most dogged and vociferous of Aristoxenus' critics, could not avoid smuggling a few Aristoxenian tenets into his work (notably at *Harmonics* 33.22–7). And there are two other writers who deal with the mathematical and Aristoxenian approaches fairly even-handedly and apparently subscribe to both. One is Adrastus of Aphrodisias (around 100 CE). He is best known as a commentator on Aristotle, but he also wrote a commentary on Plato's *Timaeus*; it has not survived as a whole, but substantial excerpts are quoted or paraphrased by his approximate contemporary Theon of Smyrna, in his *Mathematics Useful for Reading Plato*. I discuss Adrastus more fully in the context of mathematical harmonics. An alternative standpoint – based on the Aristotelian view that sciences with different subject matter must proceed from different principles – was adopted by Aristides Quintilianus in his *De musica*. His work has affinities with that of the Neoplatonists, who regularly tried to bring Aristotelian and Platonic doctrines into line with one another.

Ptolemy's comments show clearly how empirical harmonics could be regarded from the perspective of an exponent of the mathematical discipline. In *Harm.* 1.2 he launches a detailed attack upon Aristoxenians, accusing them of undervaluing and misusing reason and producing quantifications of intervals that are inconsistent with aural evidence; the theme is continued in *Harm.* 1.9, where they are condemned for defining intervals rather than notes, and only characterizing intervals in relation to other intervals.[49] This criticism is arguably unfair, as is the one that follows (20.23–21.8), which ludicrously conflates Aristoxenus' conception of the distances between pitches with the distances between the positions to which the bridge of a monochord must be moved in order to sound the relevant notes.[50] Ptolemy thus found ways of justifying the elimination of empirical harmonics from serious consideration; his approach was unlikely to be challenged, as the empiricists had already been relegated to the sidelines of philosophy and science.

[49] He returns to the charge once again in connection with the *tonoi* in 2.9–10, though this time he does not name his opponents.

[50] For further discussion of this confusion see A. Barker, *Scientific Method in Ptolemy's Harmonics* (Cambridge: Cambridge University Press, 2000), pp. 96–100, 252–4.

MATHEMATICAL HARMONICS

Mathematical harmonics fared much better in the Roman period than its empirical counterpart. Most of its exponents were Platonists; they often commented on the music of human practice, but were concerned primarily with the mathematics and metaphysics of abstract musical structures. Our estimate of their contribution will thus depend on where we place the boundaries of the science of harmonics. Does it include only propositions relevant to the structures of musical systems or extend also to remoter issues in number theory, and to questions of numerology and symbolism such as those addressed by Plutarch in his essay on the World Soul in the *Timaeus*? Any decision on these matters is bound to be fairly arbitrary. For present purposes I shall adopt a restricted view of the science's scope, as it would take an extensive examination of late Platonist metaphysics to make the theorists' views on the questions about "significance" intelligible.

It became customary for these writers to expound the elements of mathematical harmonics as a preface to discussions of the numerological and philosophical topics that really preoccupied them, but there is rarely anything original either in their remarks about the science or in the ways in which they deploy it. The most interesting of them is the influential *Timaeus* commentary of Adrastus, for which Theon of Smyrna is our best source. Theon's project was to provide Plato's readers with a guide to the mathematical disciplines at work in the dialogues; he follows Plato's order closely, but discusses "*harmonia* in numbers" separately from "*harmonia* in the cosmos," and says that both must be preceded by an understanding of "the *harmonia* perceptible in instruments" (Theo Smyrn. 16.24–18.2 and 46.20–47.17). This view seems to have been shared by Adrastus, who discussed audible music, using a broadly Aristoxenian approach, as a preliminary to his main business of mathematical harmonics; it was from Adrastus that Theon also derives most of the detail of his harmonics. Theon then cites him as saying

> that when the instruments used for the discovery of the concords are prepared beforehand in accordance with the ratios, perception bears witness to their correctness; and when perception takes the lead, reason is in tune with it (61.20–3).

This is an important and controversial point, not made in any earlier source: sense perception and mathematical reason are not in competition with one another but in partnership, and each will agree with the other's conclusions.[51]

[51] The theme is pursued in detail down to 63.24. It is the mainstay of Ptolemy's harmonic enterprise in the next century. But, for Ptolemy, accepting the testimony of perception absolutely does not mean accepting an Aristoxenian articulation of it, as it does for Adrastus.

Afterwards Aristoxenus fades from the picture almost completely and the discussion becomes progressively more focused on mathematical issues and matters remote from harmonics. But before we leave Adrastus I should like to mention two short passages which, so far as I know, have no parallel in other sources. The first is a strange argument (65.10–66.11) in which, contrary to all other theorists, he contends that the larger term of a ratio should be assigned to the lower note.[52] Secondly, he gives an elaborate discussion of the proposition that the interval of a tone cannot be divided into equal parts (69.17–72.20). He distinguishes four ways in which the tone can be conceived: in thought (*noēsis*); in numbers; in distances (*diastēmata*, not musical "intervals" here but lengths between points on a string or pipe); and in audible sounds. He offers entirely different (and variously plausible) arguments for saying that equal division is impossible in any of the last three cases; it is possible only in thought. It is an extraordinary passage. Given the Platonic context of Adrastus' treatise, all I can suggest is that it is based on the Divided Line of the *Republic*: the sounds are "images" of the physical lengths; these are their visible originals; they in turn are reflections of the numbers; and the thinkable or intelligible tone is the corresponding form.

It is now time to leave the Platonists behind and turn to Nicomachus. His lost longer essay on harmonics, as reflected in Boethius's version,[53] was quite closely related to the Platonist tradition and concerned above all with the minute mathematical dissection of elementary systems of attunement. It has an intriguing introduction, and occasional interjections of (largely speculative) nonmathematical material, but the bulk of it is unoriginal and unappetizing. His short surviving treatise is another matter. It purports to be written as an introduction to the subject for the benefit of a noble lady, an amateur scholar; as a baptism in harmonics it is an eccentric if engaging document. The gist is as follows. After a formal preface, there is a close paraphrase of Aristoxenus' discussion of the "continuous" and "intervallic" forms of vocal movement, but attributed to the Pythagoreans and recast in their terminology; then a derivation of the names of the notes from the relative positions of the seven "planets,"[54] with a brief introduction to the

[52] See A. Barker, "Adrasto e l'altezza: un argomento eccentrico e il suo contesto intellettuale," in D. Castaldo, D. Restani, and C. Tassi (eds.), *Il Sapere musicale e i suoi contesti* (Ravenna: Longo, 2009), pp. 43–56. Most theorists assign the larger number to the higher pitch, on the grounds that the value of the relevant property of its movement (e.g. its speed or its vigor, or the rapidity of successive impulses) is greater. With the exception of Adrastus, all those who assign the larger number to the lower pitch are quantifying a variable feature of an instrument (e.g., the length of a string or pipe), rather than of the sound itself.

[53] Nicomachus is certainly the main source for Boethius' first three books and perhaps the fourth; book 5 is a rather feeble attempt at a paraphrase of Ptolemy. For an excellent annotated translation, see C. M. Bower, *Fundamentals of Music: Anicius Manlius Severinus Boethius* (New Haven, CT: Yale University Press, 1989).

[54] They are Moon, Venus, Mercury, Sun, Mars, Jupiter, Saturn, in that order; the sphere of the fixed stars is not involved. The lower planets, i.e., those nearer the earth, are assigned the higher notes (in most ancient accounts of the "harmony of the spheres" it is the other way round); thus the Moon is lowest in position but is assigned the highest note.

acoustic theory on which such constructions are based and its instantiation in various instruments; then a description of the heroic achievements of Pythagoras, in his work on the scale and concords, with a cursory treatment of the *Timaeus* system and citation of Philolaus. Chapter 10 goes back to the topics of chapter 4, describing the ways in which the ratios of concords can be found on a monochord and on wind instruments; then two final essentially Aristoxenian chapters describe the complete two-octave system, the concepts "note," "interval," and "system," the perceptible difference between concords and discords, and the three melodic genera.

Given Nicomachus' Pythagorean allegiance the scheme makes a sort of sense; perhaps he presents the last two chapters as he does because the mathematics of a Pythagorean version would be too demanding in an introductory work. But it has no parallels elsewhere. Other writers (especially Neoplatonists) borrow from Nicomachus from time to time, but no one takes the form of his peculiar essay as a model.

Ptolemy's *Harmonics* is also unique, and vastly more important.[55] Its guiding theses are epistemological, focusing on the roles of the "criteria" of reason and perception. Perception is vague and imprecise, but it provides (mathematical) reason with data which reason will "bring to accuracy." Perception will then fall into line with reason, recognizing the superiority of its constructions. Finally, though in principle unerring, human reasoning is fallible; its results cannot be accepted until they have been judged by perception. If the two disagree, there must be something wrong either with our logic, or with the rational (mathematical) principles from which our derivations began.[56]

Reason, presented with rudimentary data, must establish the rational principles governing the construction of musically perfect systems of attunement, using the concept of "simplicity of comparison." The relation between two pitched sounds is quantitative, and any quantitative relation is beautiful and pleasing if the values of its two terms are easily compared; they are more easily compared where one differs from the other either by being a multiple of it, or by exceeding it by a "simple part," such as one-half or some other fraction (1.1). The comparison is simpler when the ratio is "nearer to equality," that is, when the multiple is smaller or when the "simple part" is greater; thus the most beautiful musical intervals must correspond exactly to the best and simplest ratios, and they become progressively less musically excellent as their ratios become harder to assess. Taken at face value, Ptolemy's principles imply that all concords must have

[55] I have discussed his work in detail elsewhere: Barker, *Scientific Method in Ptolemy's Harmonics*. For annotated translations see Barker, *Greek Musical Writings*, vol. 2, 270–391; J. Solomon, *Ptolemy, Harmonics: Translation and Commentary* (Leiden: Brill, 2000); M. Raffa, *La Scienza Armonica di Claudio Tolemeo* (Messina: A. Sfameni, 2002).

[56] These points are sketched in 1.1–2. I shall cite passages in the *Harmonics* by book and chapter numbers only.

multiple or epimoric ratios, but his treatment of the notes of an octave as equivalent to a single note allows him to make an exception in the case of the octave plus a fourth (ratio 8:3).[57] For all melodic intervals that are musically acceptable but less excellent than concords, the ratios must be epimoric (1.7).[58]

The second step is to work out, mathematically, all the ways in which a tetrachord can be divided in conformity with these principles. Ptolemy's derivations (1.15) are ingenious: he defines six different divisions (one enharmonic, two chromatic, and three diatonic), and rightly says that there are no other ways of dividing the tetrachord consistently with his principles. The remaining task is to bring these results to the judgment of the musical ear by using specially designed instruments, all of them based on the monochord. In Ptolemy's view, the monochord itself (described in 1.8) can be used successfully to demonstrate the ratios of the concords, but beyond that point becomes unreliable (see especially 2.12–13). He therefore calls for the use of various more complex instruments, some with eight strings, some with fifteen, together with a remarkable device on which the ratios are produced by using a single, angled bridge and sliding the strings sideways across the face of the instrument into positions determined by the ratios.[59]

Ptolemy introduces each instrument as the need arises, so the chapters discussing them are scattered throughout the *Harmonics*. The descriptions are not just mathematical but include pretty well everything a craftsman would need to know in order to build them, with discussions of the ways in which they can be used and modified to produce the acoustically best results. In fact he confidently offers his readers the chance to conduct the experiment for themselves and check his results; in principle he lays his mathematical constructions open to empirical falsification. In that sense his approach, unlike that of any other Greek theorist, can be reckoned genuinely experimental.

But Ptolemy recognizes that his mathematical constructions do not correspond exactly to the tuning systems used by contemporary musicians (1.16, with a continuation in 2.1). He explains that the enharmonic and one of the chromatics are not in fact used; and then, after discussing another possible tuning system, he describes seven different ones used by players of

[57] Unlike the Pythagoreans and other mathematical theorists (whom he criticizes on this score) he insists that this interval is a concord.

[58] These points support arguments establishing the ratio of each concord; here Ptolemy's conclusions are identical with those of other mathematical theorists. See also his discussion of the Pythagoreans in 1.5–6. The smallest concord is the fourth, with a ratio of 4:3. The melodic intervals follow in a simple sequence of progressively declining excellence, 5:4, 6:5, 7:6, and so on.

[59] This and the mathematical construction (the *helikōn*) which inspired it are described in 2.2; I have discussed it in A. Barker, "Ptolemy and the Meta-Helikon," *Studies in the History and Philosophy of Science* 40 (2009), 344–51. For a study of the monochord and its family of "experimental" instruments see Creese, *The Monochord in Greek Harmonic Science*.

the lyre and the *kithara*. Four of them diverge from his mathematical constructions by combining two of Ptolemy's divisions into a single system, with one kind of tetrachord at the top of the octave and another at the bottom.[60] He goes on to explain his conception of the *tonoi* or "keys," of which he recognizes only seven, corresponding to the seven species of the octave (2.3–11). Then he returns to the previous topic in 2.15–16, the climax of the strictly musicological part of the *Harmonics*, making it clear that the mathematical constructions were designed only to define the differences between the genera, and that they go beyond his proper task of explicating the musical systems familiar to our experience. These chapters then locate each of the "familiar" systems in the *tonos* in which musicians use it, first through tables of ratios and then discursively, with added comments on practical issues. There is every reason to believe that Ptolemy's analyses reflect very closely the practices of musicians in his environment: Alexandria in the second century CE.

Interspersed with the themes I have discussed (mainly in book 1) are vigorous criticisms of other theorists, especially Aristoxenus, but also the "Pythagoreans," who here include Platonists. Ptolemy's opinion of Archytas (1.13–14) is more favorable, though not wholly so. After 3.2 he moves into wider territory, with 3.3 a fascinating study of the metaphysical status of *harmonia* and the characteristics and functions of harmonic reason, with its dedication to the study of what is beautiful (*to kalon*); it is aided in its task by sight and hearing, the only senses that have access to beauty. These sister senses give birth to a pair of cousins, astronomy and harmonics, who are themselves brought up under the tutelage of arithmetic and geometry. All these sciences and the two senses that serve them are instruments through which we achieve an understanding of *to kalon*. The remainder of book 3 is devoted to discussions of the ways in which *harmonia* is manifested in the human soul, and in the patterns woven by the stars and planets; here scientific astronomy is inextricably entangled with astrology.[61] This material has no special importance in the tradition of speculation on these topics. Ptolemy's distinctive contribution to harmonics ends with 3.3.

As far as we know, Porphyry's commentary on Ptolemy's *Harmonics* is the only ancient commentary ever written on a musicological text.[62] Its form, however, is familiar from comparable philosophical writings, discussing the text passage by passage in an almost continuous sequence from the beginning to partway through 2.7. The author was probably

[60] He admits one exception: while singers performing pieces in what he calls the "tense diatonic" use the Ptolemaic ratios (10:9x9:8x16:15), string players tune their instruments to the more practically convenient pattern of tone, tone, and *leimma* (9:8x9:8x256:243), which is strictly improper, but the differences are mathematically minute and not audibly appreciable.

[61] Ptolemy's astrological work, the *Tetrabiblos*, has rather little in common with these chapters.

[62] For annotated translations into English and Italian, see A. Barker, *Porphyry's Commentary on Ptolemy's Harmonics: A Greek Text and Annotated Translation* (Cambridge: Cambridge University Press, 2015).

drawn to the subject by Plato's discussions in the *Republic* and the *Timaeus*. In an introduction he explains why he chose to write specifically on Ptolemy: because of Ptolemy's mathematical difficulty, his keen judgment of earlier treatments of harmonics, and his mastery of mathematical disciplines and ancient philosophy, and because he has thus, uniquely, brought the science to perfection.

Modern scholars have valued the work mainly for its enormous collection of quotations from other writers. But there are other good reasons for studying it. Porphyry is above all a philosopher, and it is very clear that the parts of Ptolemy's work that most interested him are those that raise serious philosophical issues. The commentary is thus heavily weighted towards Ptolemy's opening chapters, where epistemological, methodological, and logical issues are most explicitly in play. Many of Porphyry's reflections on these matters are original and are important to philosophers, notably his discussions of 1.1 and 1.3; in respect of the latter he takes issue with Ptolemy's assignment of pitch to the category of quantity and insists that it is a qualitative attribute.[63] There is good meat for students of harmonics elsewhere, for instance in the discussions of continuous and intervallic sound, of the proper definition of the concept "note" in 1.4, and of the relations between *diastēma* (interval), *logos* (ratio), and *hyperochē* (the "excess" of one term over another) in 1.5. But, as Ptolemy's treatment becomes progressively more technical and more closely focused on the details of harmonic construction, so Porphyry's interest wanes. I suspect that the commentary is incomplete because Porphyry gave up and never finished it, not because the remainder was written and has been lost.

Though Porphyry is not the latest of the theorists I have mentioned, this seems a fitting place to end. Of the two main branches of harmonic theory, the empirical has degenerated into scholasticism, and the mathematical has been absorbed into the works of philosophers, to be used for their own purposes rather than to be developed for its own sake or to promote a greater understanding of audible music. In a sense it has returned to its origins. Ptolemy had reinvigorated it as an authentically musical science, but Porphyry now draws his *Harmonics* too into the philosophical ambit. The flame of harmonics as an autonomous science with its own agenda flickers and goes out, not to be rekindled until the musicological controversies of the Renaissance.

[63] In Ptolemy's text, 1.3 runs to about three pages, in Porphyry's commentary to over forty-eight. It is clearly his dedication to the logical issues raised by Aristotle (whom he champions here against Plato) that prompted him to deal at such length with a passage which for Ptolemy is only a preliminary bit of ground clearance. It is also the only part of the commentary in which he criticizes Ptolemy's argumentation and rejects his conclusions.

23

GREEK MECHANICS

Serafina Cuomo

The original meaning of the Greek word *mēchanē* is "trick," "ruse," or "cunning device"; it was associated both with a mental quality, resourceful intelligence (*mētis*), and with material artifacts such as instruments, produced by (*mēchanikē*) *technē*, an "art" that could be learned and taught.[1]

Mētis helped in situations where the odds were against you. For instance, in the *Iliad* the son of Nestor, Antilochus, overtakes Menelaus in a chariot race even though his horses are not as good as those of his rival.[2] Another classical personification of *mētis* was the *polumēchanos* (resourceful, many-resourced) and *polumētis* Odysseus. The epithet of *polumēchanos* is in Homer restricted to Odysseus, and is connected to his ability to survive.[3]

But, both with Antilochus and with Odysseus, *mētis* and *mēchanē* are also associated with *dolos*, "trap" or "treachery," "stratagem." Even though he has come second in the chariot race, the son of Nestor is practically forced to surrender his prize to Menelaus, because the latter stands up and accuses him of foul play.[4] There was thus a sense in which a ruse was not fair; cunning devices indicated ethically questionable behavior. Thus, "to machinate" is used by Homer to indicate what Penelope's suitors are doing against Telemachus, and by Demosthenes to denote the crafty but deceitful

[1] The Latin *machina* derived from the Doric form *machana*, common in Southern Italy; cf. Astrid Schürmann, *Griechische Mechanik und antike Gesellschaft. Studien zur staatlichen Förderung einer technischen Wissenschaft* (Stuttgart: Steiner, 1991), p. 35.

[2] The episode, quoted in Marcel Detienne and Jean-Pierre Vernant, *Les Ruses d'Intelligence* (Paris: Flammarion, 1974), translated by Janet Lloyd as *Cunning Intelligence in Greek Culture and Society* (Hassocks: Harvester Press, 1978), p. 12, is *Iliad* 23.306 ff.: [Nestor to Antilochus] "The horses of the others are swifter, but the men know not how to devise more cunning counsel than thine own self. Wherefore come, dear son, lay thou up in thy mind cunning of every sort, to the end that the prizes escape thee not" (trans. A. T. Murray, 2 vols. (Loeb Classical Library; Cambridge, MA: Harvard University Press, 1924–5), vol. 2, 517).

[3] He is called *polumēchanos* at, e.g., *Iliad* 2.173; 4.358; 8.93; *polumētis* at, e.g., *Iliad* 1.311; 3.200; 3.216. This has been contrasted with Achilles' qualities and with the fact that he is *ōkumoros*, "short-lived." Cf. Anthony T. Edwards, *Achilles in the Odyssey* (Königstein: Hain, 1985).

[4] *Iliad* 23.570–85.

discourses of his opponent.[5] Analogously in Latin, a *machinator* was both a person who built machines or an architect, and a person who engineered plots and spun webs of deceit, such as Catiline (at least as seen by Cicero).[6]

In other words, throughout ancient culture, *mēchanē* carried both positive and negative overtones. Sometimes it was praised for its capacity to find a way out of apparently desperate straits, sometimes reviled because of its element of base cunning. In any case it subverts what should *normally* happen, thus interfering with, or disrupting, the natural course of events, the established state of things where the strong prevail over the weak.

FROM HOMER TO ARISTOTLE

Homer is one of the earliest literary sources to mention drills and levers, employed for example by Odysseus and his men to launch their raft to sea.[7] The textual tradition often registers what has already been in use for some time: Thucydides contains references to siege engines such as battering rams and Aeschylus refers to the *mēchanas* deployed by Xerxes to bring the Persian troops across the Bosporus. At the same time, the term retained its more general sense of military "stratagems."[8] Lifting or carrying devices must have been used quite early on in architecture to move around heavy blocks of stone. Herodotus reports that "levers made of short wooden logs" were employed in the construction of Cheops' pyramid, and machines are mentioned in numerous Greek building inscriptions (which were set up to list the items of expenditure).[9] By the fifth century BCE, machines were in use in the theatre whenever changes of scene or spectacular exits were required. One of the characters in Aristophanes, for instance, reproaches the operator (*mēchanopoios*) for a sudden change in the backdrop which scared him.[10]

5 *Odyssey* 3.207; 3.213; 16.134; and Demosthenes, *Against Androtion* 35, respectively.

6 E.g. for the first meaning Livy, 24.34.2 about Archimedes, "machinator bellicorum tormentorum," and Tacitus, *Annals* 15.42 talking about one of the people who built Nero's *domus aurea*; for the second meaning e.g. Cicero, *Against Catiline* 3.6 and the *Codex of Theodosius* 10.24.3.

7 E.g. *Odyssey* 5.261; 9.385. Examples quoted in Gianni Micheli, *Le origini del concetto di macchina* (Florence: Olschki, 1995), pp. 10–11.

8 Respectively, Thucydides, e.g. 2.76; 2.77; 4.13; Aeschylus, *The Persians* 722; and Xenophon, *Cyropaedia* 1.6.38.

9 Herodotus, 2.125. See also Alison Burford, *The Greek TempIe Builders at Epidaurus* (Liverpool: Liverpool University Press, 1969) and inscriptions from, e.g., Delphi, in *Fouilles de Delphes* III.19, and from Didyma, in *Didyma. Die Inschriften* II.39. We have later iconographical evidence for machines used in architecture, e.g. a Roman bas-relief depicting two simple pulleys, or the famous bas-relief from the tomb of the Haterii, representing a huge crane operated by men walking inside a wheel. Evidence collected in J. P. Adam and P. Varène, "Une peinture romaine représentant une scène de chantier," *Bulletin archéologique* (1980), 213–38.

10 Aristophanes, *Peace* 174. Machines of this sort could also be used for other purposes: Livy, 39.13.13, has a story of a dubious mystery cult, where people who refused to join in once they discovered that the cult involved all sorts of orgies "were alleged to have been carried off by the gods, [while they] had been bound to a machine and borne away out of sight."

The emergence of mechanics as a form of knowledge can be seen as an attempt to understand the nature of some man-produced phenomena which were already "out there" as well as an inquiry aimed at producing machines "new" or "better" than the extant ones.

According to Diogenes Laertius, Archytas of Tarentum was the first person to deal with mechanics using mathematical principles. We also have this passage in Plutarch:

> the art of mechanics, now so celebrated and admired, was first originated by Eudoxus and Archytas, who embellished geometry with its subtleties [. . .]. But Plato was incensed at this, and inveighed against them as corrupters and destroyers of the pure excellence of geometry, which thus turned her back upon the incorporeal things of abstract thought and descended to the things of sense, making use, moreover, of objects which required much mean and manual labour.[11]

This is not the right place to discuss the very complex question of Plato's attitude to technology and "applied" knowledge. Although in late antiquity a solution to the problem of cube duplication which employed mechanical aids (thus "corrupting" geometry) went under Plato's name, Plutarch's interpretation has been very influential.[12]

Aristotle, on the other hand, mentions mechanics, together with optics and harmonics, as an example of the so-called "subordinate," "mixed," or "middle" (as they came to be called from the Middle Ages) sciences, which borrow first principles from arithmetic and (in the case of mechanics) from geometry.[13] Moreover, at the beginning of the *Metaphysics*, Aristotle says: "All men begin [. . .] by wondering that things are as they are, as they do about self-moving devices, or about the solstices or the incommensurability of the diagonal of a square with the side." Self-moving devices, which are a typical product of mechanics, are used as an example of astonishing phenomena, the causes of which are obscure and warrant further investigation. Philosophy itself is characterized as an inquiry into such "unnatural" things, in an effort to bring them under the control of human understanding.[14]

[11] Diogenes Laertius, *Lives of the Philosophers* 8.83 and Plutarch, *Life of Marcellus* 14.3–5 (trans. B. Perrin, *Plutarch Lives*, 11 vols. (Loeb Classical Library; Cambridge, MA: Harvard University Press, 1914–26), vol. 5, 470–3).

[12] The solution in Eutocius of Ascalon's commentary on Archimedes' *Sphere and Cylinder* 66.21 ff. Studies on this topic include Giuseppe Cambiano, *Platone e le tecniche* (Turin: Einaudi, 1971) and Helmuth Schneider, *Das griechische Technikverständnis* (Darmstadt: Wissenschaftliche Buchgesellschaft, 1989).

[13] *Posterior Analytics* 76a20–5 and *Metaphysics* 1078a10–18.

[14] *Metaphysics* 983a13–17 (translation by W. D. Ross (Oxford: Clarendon Press, 1928), with modifications).

THE PSEUDO-ARISTOTELIAN *MECHANICAL QUESTIONS*

The earliest mechanical treatise we have has sometimes been attributed to Strato of Lampsacus, Theophrastus' successor at the leadership of the Peripatos from ca. 288 to 269 BCE and, according to Diogenes Laertius, tutor to the Egyptian king Ptolemy II Philadelphus.[15] It opens with a definition of mechanics:

> When [. . .] we have to do something contrary to nature, the difficulty of it causes us perplexity and art has to be called to our aid. The kind of art which helps us in such perplexities we call mechanics. [. . .] Instances of this are those cases in which the less prevails over the greater, and where forces of small motive power move great weights – in fact, practically all those problems which we call mechanical problems. They are not quite identical nor yet entirely unconnected with natural problems. They have something in common both with mathematical and with natural speculations; for while mathematics demonstrates the how, natural science demonstrates the about what.[16]

Mechanics is at the interface between nature, *phusis*, and what lies beyond or against nature, *para phusin*; it originates from human resourcefulness in the face of a hostile or at least unsupportive environment. Mechanics is also at the interface between physics, or rather natural philosophy, the kind of inquiry that Aristotle has carried out for example in the *Physics*, and mathematics. It is an intervention on the world that surrounds us, as well a cognitive effort to understand it. Mechanics is also defined as the art of making the weaker prevail over the stronger, and it is accompanied by wonder (like the *automata* at the beginning of Aristotle's *Metaphysics*), by the sense that what is happening is extra-ordinary. The main body of the treatise is arranged in a sequence of questions and answers, ranging from "Why does a missile travel further from the sling than from the hand?" to "Why does a small rudder move a large ship with the exercise of little force?," and from "Why are pebbles round?" to "Why can a dentist extract a tooth more easily with a tooth extractor than with the hand only?"[17] Answers are generally along the lines of finding what forces are at work, and which is the point where they are applied. The *Questions* also contain a discussion of the following question: why are larger balances more accurate than smaller ones? First the author explains that, when two concentric circles move around their common center, although they are moved by the same force, the larger one will have to move faster in order to keep up with the smaller one. From

[15] Diogenes Laertius, *Lives* 5.58–64, esp. 58–9.

[16] *Mechanics* (henceforth "*Mech.*") 847a16–30 (translation by E. S. Forster, in W. D. Ross and J. A. Smith (eds.), *The Works of Aristotle Translated into English*, 12 vols. (Oxford: Clarendon Press, 1908–52), vol. 6, with modifications).

[17] *Mech.* 852a38 ff.; 850b30 ff.; 852b29 ff.; 854a16 ff., respectively.

this derives that larger balances are more accurate than smaller ones, because, if one imagines the concentric circles as balances with the center as fulcrum, in small balances the weight will hang closer to the fulcrum and will therefore produce a very small movement, while in large balances "the extent of the swing is much greater for the same weight."[18]

After considering the case in which a balance is supported from underneath the fulcrum, rather than suspended from above, the pseudo-Aristotle moves on to "Why is it that small forces can move great weights by means of a lever?" This is again explained on the basis of the two concentric circles, and the demonstration incidentally mentions that "the ratio of the weight moved to the weight moving it is the inverse ratio of the distances from the centre,"[19] i.e. a formulation of the so-called principle of the lever. The author of the *Mechanical Questions* does not aim at presenting new machines; rather, he wants his audience to understand why some things that are already familiar to them happen the way they do. Thus, the treatise takes machines common in everyday use, such as levers or pinchers, or everyday phenomena that do not seem susceptible of normal, "natural" explanation, and finds the cause of them all in the properties of the circle.

The circle is seen as the site of opposites and a source of wonder, because of its paradoxical properties: a point moving on its surface is at the same time getting back towards its starting-point; while the circle moves, one part of it (its center) remains at rest; also, the circle is both convex and concave. While in itself wondrous, the circle can help us understand the causes of mechanical phenomena, and thus dispel the wonder they inspire.[20] For instance, the author describes "wheels of bronze and steel which are dedicated in temples," such that the motion of one imparts a contrary motion to the next one, and says: "So making use of this property inherent in the circle, craftsmen make an instrument concealing the original circle, so that the marvel of the machine is alone apparent, while its cause is invisible."[21]

Knowing the cause of the movement of the wheels amounts, in more than one sense, to lifting the veil on the hidden mechanism that lies behind their marvellous appearance.[22]

[18] *Mech.* 849b30. The author also adds: "This is how sellers of purple arrange their weighing machines to deceive, by putting the fulcrum out of the true centre, and pouring lead into one arm of the balance, or by employing wood for the side to which they want it to incline taken from the root or from where there is a knot" (trans. W. S. Hett, *Aristotle Minor Works* (Loeb Classical Library; Cambridge, MA: Harvard University Press, 1936), 346–7, with modifications).

[19] *Mech.* 850a30–850b10.

[20] The paradoxical qualities of the circle had already been noted, e.g. in Plato's *Laws* 893c–d, where circular motion around a fixed center is seen as "the source of all marvels."

[21] *Mech.* 848a25–6 for the mention of the circles; the passage at 848a34–8. It is not entirely clear what these wheels looked like; similar objects, called Egyptian wheels, are described by Hero, *Pneumatics* 148.2–5 and Plutarch, *Life of Numa* 69.

[22] On the topic in general see Sylvia Berryman, *The Mechanical Hypothesis in Ancient Greek Natural Philosophy* (Cambridge: Cambridge University Press, 2009).

THE HELLENISTIC KINGDOMS

The *Mechanical Questions* depict mechanics as productive both of things useful to humankind and of wonder and astonishment. The testimonies we have from the Hellenistic world confirm this picture: mechanics is applied to war and agriculture and to public displays. Indeed, there is a sense in which mechanics for display serves purposes analogous to mechanics for utility: they both aggrandize the monarch and convey in both cases an image of effortless power.

Nearly every single episode from this period where mechanical devices are used sees the presence, either indirect or as a participant, of a political leader – indeed, the connection between power and mechanics in antiquity as a whole cannot be overstated.

For instance, the Aristotelian philosopher Demetrius of Phalerum, who ruled Athens as archon in 309–308 BCE, is reported as sending "a snail moved by machinery" and spitting out saliva in front of his procession.[23] Ptolemy II Philadelphus (whom we have mentioned in connection with Strato) is associated with another procession in Alexandria, held probably between 280 and 275 BCE as part of a public festival. Among the several grandiose displays, signifying the wealth and might of the sovereigns, there was a wagon pulled by sixty men with a statue on top which stood up "mechanically, without anyone laying a hand on it" and poured a libation[24].

But peacetime processions were not the only scenario for mechanics. War technology made huge steps in this period, according to the ancient sources themselves. The tyrant Dionysius of Syracuse is credited with the invention of the first catapult around 399 BCE. He realized that it was a favorable moment for him to enter war with the Carthaginians and,

> [a]fter collecting many skilled workmen, he divided them into groups in accordance with their skills, and appointed over them the most conspicuous citizens [. . .] not only was every space, such as the porticoes and back rooms of the temples as well as the gymnasia and colonnades of the market place, crowded with workers, but the making of great quantities of arms went on, apart from such public places, in the most distinguished homes. In fact the catapult was invented at this time in Syracuse, since the ablest skilled workmen had been gathered from everywhere into one place. The high wages as well as the numerous prizes offered the workmen who were judged to be the best stimulated their zeal. And over and above these factors, Dionysius circulated daily among the workers, conversed with them in kindly

[23] Polybius, quoting Demochares, 12.13.11. On the snail, see Schürmann, *Mechanik*, pp. 239–40.

[24] Athenaeus, quoting from Kallixeinos of Rhodes, in *Deipnosophists* 198 f. (E. E. Rice's translation in *The Grand Procession of Ptolemy Philadelphus* (Oxford: Clarendon Press, 1983). The festival was in honour of Dionysus (the Ptolemies considered themselves his descendants), and the statue in question represented Nysa, the city in India where Dionysus was allegedly born.

> fashion, and rewarded the most zealous with gifts and invited them to his table. Consequently the workmen brought unsurpassable devotion to the devising of many missiles and engines of war that were strange and capable of rendering great service.[25]

Within a few years the engines were in common use: Diodorus himself reports a siege in 340 BCE where both sides (one of them being Philip of Macedon) were well equipped with *mēchanas*; Alexander the Great used siege engines extensively, and the name of one Hellenistic king, Demetrius of Ephesus Polyorketes (336–283 BCE), became synonymous with siege engines, especially with a huge attack tower called helepolis. It was apparently such a wondrous device that even enemies were struck by its beauty and on one occasion asked Demetrius for a truce just to have a closer look at it.[26]

To the extent to which mechanics on display symbolized a certain configuration of power – monarchic (when not tyrannical) – attitudes to mechanics took on political meanings too. For instance, Demetrius' snail was frowned upon as a risible trick on the part of a second-rate politician (at least according to Polybius' testimony), and Plutarch's anecdote, according to which the Spartan king Archidamus, "when he saw a missile shot by a catapult which had been brought then for the first time from Sicily, exclaimed, 'Great Heavens! Man's valor is no more!'," is also cast as a typical utterance from a leader of "egalitarian" Sparta.[27] Or, in the Roman world this time, a mechanical feat such as the revolving theatres of Gaius Curio (ca. 52 BCE), attracted this comment on the part of Pliny the Elder:

> [h]ere we have the nation that has conquered the earth, that has subdued the whole world [. . .] swaying in a machine and applauding its own danger [. . .] and the aim, after all, was merely to win favor for the speeches that Curio would make as tribune.[28]

To move to a superficially less grandiose level, the references to waterlifting machines from the Hellenistic period are too numerous to mention. They

[25] Diodorus, 14.41–2. The catapults are first seen in action at the siege of Motyê in 397 BCE: "Indeed this weapon created great dismay, because it was a new invention at this time" (14.50). On this passage see Serafina Cuomo, *Technology and Culture in the Greek and Roman World* (Cambridge: Cambridge University Press, 2007), chapter 2.

[26] Diodorus, 16.74–5. Philip owned "numerous and varied catapults," while the besieged inhabitants of Perinthus received reinforcements of men, missiles, and artillery from Byzantium. On Demetrius Polyorketes see especially Plutarch, *Life of Demetrius* 20–1; see also a description of the helepolis in Biton, *Belopoietics* 52 ff. (third quarter of third century BCE), in E. W. Marsden, *Greek and Roman Artillery, Technical Treatises* (Oxford: Clarendon Press, 1971). On catapults see Tracey Rihll, *The Catapult: A History* (2nd edn; Yardley, PA: Westholme, 2009).

[27] Plutarch, *Sentences of Kings and Emperors* 191e and *Laconic Sentences* 219a.

[28] Pliny Sr., *Natural History* 36.24 (116–20), trans. D. E. Eichholz, *Pliny Natural History, vol. 10* (Loeb Classical Library; Cambridge, MA: Harvard University Press, 1962), pp. 92–5, with modifications.

were used principally for irrigation purposes, but also against fires. Both Vitruvius and Hero describe pumps for this purpose.[29]

Our papyrological evidence depicts at least one of the Ptolemaic engineers with some clarity: Kleon, who lived around 262–249 BCE in Krokodilopolis in the Arsinoite district. He had quite high status, seems to have had direct access to the king, and commanded, apart from a hefty salary for himself, big sums for the paying and sustaining of his workforce.[30] Kleon was in charge of the irrigation of quite a large region, which involved supervising the good functioning of water-lifting machines as well as setting in place, repairing, and generally maintaining a system of canals and waterways. Work of this kind could be sub-contracted, which suggests a rather widespread level of mechanical expertise.

ARCHIMEDES

Archimedes of Syracuse will already be known to the reader; see chapter 18 in this volume. His identity as a mechanician has been the subject of some controversy, although he was known to the general public in antiquity less for his writings than for his feats of "engineering" – as the person who managed, single-handedly, to launch a huge ship into the sea or as the old man whose war engines kept the Roman army at bay for nearly two years. He built a cog-wheel device in the first case and, in the second, designed various types of catapults, large and small, all very powerful and accurate, an iron hand that could grab enemy ships and sink them into the sea, and, at least according to later sources, even burning mirrors.

Yet, according to Plutarch,

> to these [Archimedes] had by no means devoted himself as work worthy of his serious effort, but most of them were mere accessories of a geometry practiced for amusement, since in bygone days Hiero the king had eagerly desired and at last persuaded him to turn his art somewhat from abstract notions to material things. [. . .] the art of mechanics, now so celebrated and admired, was first originated by Eudoxus and Archytas [. . .]. But Plato was incensed at this, and inveighed against them as corrupters and destroyers of the pure

[29] The evidence, both literary and archaeological, extends well into the second century CE and is collected in J. P. Oleson, *Greek and Roman Mechanical Water-lifting Devices* (Toronto: University of Toronto Press, 1984). Other instances where mechanical devices were employed in agriculture were wine- and oil-presses, as described by, e.g., Hero, *Mechanics* (henceforth "Hero, *Mech.*") 3.13–20 and Cato, *On Agriculture* 18–19 (he also describes a mill at 20–1), and found on archaeological sites such as Pompeii and Cosa (evidence collected in K. D. White, *Farm Equipment of the Roman World* (Cambridge: Cambridge University Press, 1975) and Schürmann, *Mechanik*, pp. 127–37). The evidence for water pumps is collected in Schürmann, *Mechanik*, pp. 104–12.

[30] See the chapter dedicated to Kleon in Naphtali Lewis, *Greeks in Ptolemaic Egypt* (Oxford: Clarendon Press, 1986).

> excellence of geometry [. . .]. For this reason mechanics was made entirely distinct from geometry, and being for a long time ignored by philosophers, came to be regarded as one of the military arts.[31]

It will never be known, even though debate will probably continue, what Archimedes really thought about his extra-curricular activities, but the issue itself need not be as relevant as the question of what Archimedes became, of how his image was constructed on the one hand as that of the pure Greek mathematician, and on the other as that of the wizard mechanician and engineer.

In both cases his relationship to King Hiero is presented as crucial: it was Hiero who diverted Archimedes from his favorite pursuits towards more earthly ones; it was in front of Hiero that the huge ship was launched into the sea (the ship itself was a present by Hiero to one of the kings in Alexandria); it was as a follow-up to the fact that Hiero wanted to find out whether the golden crown he had paid for was really all made of gold that Archimedes jumped out of the bath shouting "Eureka!"

If we turn to Archimedes' extant writings, several of them deal with mechanics: he wrote two books *On the Equilibrium of Planes*, about centers of gravity of plane figures, and two books *On Floating Bodies*. He and others also refer to a *Mechanics*, which could be identified with a perhaps fuller version of the work on centers of gravity.[32]

Archimedes' mechanical works, such as we have them, are organized along axiomatico-deductive lines, and abstract from any physical property so much so that they provide formulations for the center of gravity of *plane* figures (triangle, parallelogram, trapezium).[33] Archimedes also gives a mathematical version of the principle of the lever ("Two magnitudes, whether commensurable or incommensurable, balance at distances reciprocally proportional to the magnitudes") and, of course, the so-called principle of Archimedes: a solid lighter than a fluid "if it is forcibly immersed, [will] be driven upwards by a force equal to the difference between its weight and the weight of the fluid displaced."[34]

The combination of mechanics and mathematics is again at work in the *Quadrature of the Parabola*, principally aimed at finding the area of the segment of parabola. The treatise accomplishes that both mathematically,

[31] *Life of Marcellus* 14.3–5. Descriptions of the siege of Syracuse also in Polybius, 8.3(5)–7(9) and Livy, from 24.33 to 25.23–31.

[32] He refers to a *Mechanica* in *Quadrature of the Parabola* 6 and 10; to *Elements of Mechanics* in *Floating Bodies* 2.2.

[33] Hero *Mech.* 1.24 reports that the Stoic Posidonius had defined the center of gravity and of inclination from a physical point of view, whereas Archimedes was more specific and drew a distinction between center of gravity and center of inclination. Archimedes is also indicated as the first to have inquired into the centers of gravity of plane figures.

[34] In Thomas L. Heath's formulation of props. 6 and 7 of the *Equilibrium of Plane* I and of prop.6 of *Floating Bodies* I, in *A History of Greek Mathematics*, 2 vols. (Oxford: Clarendon Press, 1921), vol. 2, 75 and 92, respectively.

with a proof along usual (for Archimedes) "exhaustion method" lines, and mechanically. Namely, the parabola is considered as posited on one arm of a balance and in equilibrium with other figures and eventually proved equal to one of them which has such-and-such characteristics. Moreover, a palimpsest rediscovered in 1899 revealed a hitherto unknown work, known as the *Method*, where, in the form of a letter to Eratosthenes, Archimedes explains that the background to some of his most famous discoveries, e.g. the volume of the sphere, implied the use of mechanics.

The "method" of the title

> conceives geometrical figures to be attached to a lever in such a way that the latter remains in equilibrium, and then draws up conditions for such equilibrium; and it is further based on the view that the area of a plane figure is to be looked upon as the sum of the lengths of all the line segments drawn therein in a given direction and of which the figure is imagined to be made up."[35]

PHILO OF BYZANTIUM

Philo of Byzantium is the author of the *Mechanical Syntaxis*, which originally comprised nine parts, about, among other things, the lever, the construction of harbours, *automata*, and sieges. The only extant parts are the fourth book, on construction of war machines (*Belopoietics*), and the fifth book, on pneumatics, which has come down to us in an Arabic translation, as well as fragments from books 7 and 8 on defensive and offensive works and siege operations.[36] It is not certain when Philo lived: he says that he knew people who knew Ctesibius, and "the linguistic analysis of his extant writings [. . .] shows that he is one of the first representatives of the literary κοινή, thus a man of the second century BC."[37] The *Pneumatics* is addressed to Ariston, a friend who had required an introduction to the subject. Philo protests that his account is for beginners and non-experts, and aims at conciseness over exhaustiveness. Even though he eschews reporting in full all the contrasting opinions on void and matter, he mentions one theory which he takes to be wrong and misleading, i.e. a version of atomism according to which void

[35] E .J. Dijksterhuis, *Archimedes*, trans. C. Dikshoorn (Princeton, NJ: Princeton University Press, 1981; 1st edition 1956), p. 318. See also Reviel Netz, William Noel, Natalie Tchernetska, and Nigel Wilson (eds.), *The Archimedes Palimpsest* (Cambridge: Cambridge University Press, 2011).

[36] Fragments of the seventh and eighth books, erroneously indicated as book five, in Y. Garlan, *Recherches de poliorcétique grecque* (Athens: École Française d'Athènes, 1974). The Arabic version of the *Pneumatics* in B. Carra de Vaux, "Le livres des appareils pneumatiques et des machines hydrauliques par Philon de Byzance," *Notices et extraits des manuscrits de la Bibliothèque Nationale de Paris* 38 (1903), 27–235. There is also a partial version of a Latin translation of the *Pneumatics* in the Teubner edition of Hero's *Pneumatics.*

[37] Schürmann, *Mechanik*, p. 7.

does exist in nature. Philo, on the contrary, thinks that there is no natural void, but that when a vessel looks empty, it is in fact full of air, which is a body.

Consequently, he first proves, by means of a vessel upturned in water, that there is something in the vessel, and then that this something is air. Philo's *Belopoietics* abounds in details, including cost considerations and attention to the general appearance of the completed machine, which has to look imposing, so as better to frighten the enemy. In the text, again addressed to Ariston, Philo talks about the discovery that there is a mathematically expressible relation between the various parts of a catapult and the diameter of the hole that contains the torsion spring of that catapult. That is, the relation can be expressed in terms of enlarging or diminishing cubes: to build a catapult proportionally twice as big as a given one, it is possible to use the duplication of the cube. Philo's *Belopoietics* is the first extant text to contain a solution to this problem.

The discovery occurred, Philo says, over time:

> The ancients [. . .] did not reach a conclusion, [. . .] since their experience was not based on many works, but they did decide what to look for. The more recent [technicians] drew conclusions from former mistakes, and from things that were experienced after those looking at the constant element, they reached the principle and basis of construction. [. . .] Alexandrine craftsmen achieved this first, being heavily subsided through the interest of kings fond of the arts and of fame [that is, the Ptolemies].[38]

VITRUVIUS

In his *Architecture*, written around the last thirty years of the first century BCE and dedicated to the Emperor Octavian Augustus, Vitruvius devotes some space to mechanics in chapters eight (on aqueducts), nine (on dials and clocks), and ten (on machines), on which latter I will focus.[39]

The tenth book describes both peacetime and war machines. To the first category belong above all weight- or water-lifting instruments, then the hydraulic organ, and the hodometer, a device which could be attached to the wheels of a cart and enabled one to measure the distance traveled. This last object embodies the two aspects of mechanìcs we have already encountered: *utilitas* and *delectatio* (enjoyment).[40] While stressing the importance

[38] *Belopoietics* 50.19–26; translation, with modifications, in Marsden, *Artillery*.

[39] Other treatises of mechanics directly dedicated to monarchs include Biton's *Belopoietics* (to King Attalus of Pergamon) and Apollodorus' *Polyorketiks* (to Hadrian, second century CE).

[40] *On Architecture* (henceforth "*Arch.*") 10.9.7. The hydraulic organ was one of Nero's favourite instruments, cf. Suetonius, *Lives of the Caesars* 6.41.2; 6.54.

of first-hand experience, Vitruvius explicitly relies on previous authors, many of them Greek. Nowhere is the existence of a well-established tradition more evident than in Vitruvius' report of measurements necessary to build or modify several types of catapult on the basis of weight and type of the projectile to be used. The data are presented as more or less standard, the result of years of research, probably along the lines described by Philo.

Whereas some machines are so common that their construction is completely unproblematic, war engines can perplex those who are not familiar with numbers and multiplications, Vitruvius says.[41] In any case, no matter how sophisticated the machine, the real power lies with the use one makes of it and thus, ultimately, with human intellect: in a famous anecdote, the architect Diognetus of Rhodes defeats Demetrius Polyorketes (and *his* architect, Epimachus of Athens) with the simple trick of a mass of mud, refuse, and debris which cause the huge helepolis of the king to get stuck, so that it can be easily destroyed.[42]

In other words, in Vitruvius the emphasis is not only on the machines, but also on the people designing or using them.

HERO OF ALEXANDRIA

Hero (second half of the first century CE) wrote on war machines (*Belopoietics*), on self-moving puppets (*Automata*), on devices operated by means of hot air (*Pneumatics*), on surveying instruments (*Dioptra*), on mirrors (*Catoptrics*), as well as on *Mechanics* "in general." He has also left several works dealing with more specifically mathematical matters. Hero is a particularly precious source in that he is very vocal in praising his forms of knowledge (mechanics and mathematics) and in comparing them (favorably, of course) with philosophy. For instance, in the introduction to the *Belopoietics*, he says:

> The largest and most essential part of philosophical study is the one about tranquility, about which many researches have been made and still are being made by those who pursue learning; and I think research about tranquility will never reach an end through reasonings. But mechanics has surpassed teaching through reasonings on this score and taught all human beings how to live a tranquil life by means of one of its branches, and the smallest – I mean, of course, the one concerning the so-called construction of artillery.[43]

[41] *Arch.* 10.11.1.
[42] Ibid. 10.16.3–8.
[43] Hero, *Belopoietics* 71–73.11; translation, with modifications, from Marsden, *Artillery*, pp. 18–19. A similar contrast with philosophy is in Athenaeus Mechanicus (maybe second century CE), *On Machines* 4–5, edited by R. Schneider in, "Griechische Poliorketiker III," *Abhandlungen der königlichen Gesellschaft der Wissenschalten zu Göttingen*, Philol.-hist. KI. N.F. 12, Heft 5 (1912).

The *Pneumatics* is addressed to future mathematicians. With respect to those who think that there is absolutely no void and those who think that a continuous vacuum exists in nature, Hero opts for a sort of compromise theory of void, according to which void is "distributed in minute portions" through all substances, and continuous vacuum can be obtained only in situations which are against nature. In other words, void can be "created" by machines, even simple ones like a doctor's bleeding cups.[44]

In the treatise Hero proceeds to describe several types of siphons and vessels, where the effects of air pressure and of steam produce all sorts of wondrous effects. A typical example is a vessel which remains full, although water is drawn from it, but there are also devices to produce the sound of a trumpet, the fire-pump and hydraulic organ we have already mentioned, some puppets which "drink," and a machine to open the gates of the temple when fire is lit on an altar. Utility and wonder are united by Hero himself as key themes: "various combinations are effected, some of which supply the most pressing wants of human life, while others produce amazement and alarm."[45]

The *Automata* is the only treatise surviving from antiquity on the topic. Examples of self-moving items for grand public occasions have already been noted; the *automata* Hero describes in his treatise are of smaller size, and to be used principally in the context of banquets or feasts. He stresses the importance of decoration and novelty, as well as of good "engineering," so that the device does not stop abruptly or break down altogether.[46] The *Mechanics* deals principally with what already in the *Mechanical Questions* had been presented as the quintessential mechanical problem: how to move a body. The treatise opens with a description of a cog-wheel device, the *poluspaston*, which should enable the person operating it to move a certain weight (Hero provides the example of one thousand talents) with a small force, say, five talents, and then proceeds to a discussion of the movements of circles set one next to the other (which are a sort of geometrical framework for the cog-wheels themselves).[47]

[44] *Pneumatics* Pref. 2.4–28.15 (translated by J. G. Greenwood, edited by B. Woodcroft, and reprinted with an introduction by Marie Boas Hall (London/New York: Macdonald & Elsevier, 1971). Cf. Karin Tybjerg, "Hero of Alexandria's Geometry of Mechanics," *Apeiron* 37 (2004), 29–56.

[45] The passages at *Pneumatics* 102.20–110.11; 96.7–100.15; e.g. 144.7–146.15; 174.10–182.6; and 2.1820, respectively. "Trick vases" dating from at least the sixth century BCE have been found, e.g., in Boeotia. They are very similar to the one described by Hero and employ the same siphon principle. See e.g. G. Argoud, "Hydraulique et siphons béotiens," in P. Roesch and G. Argoud (eds.), *La Béotie antique* (Paris: Éditions du Centre national de la recherche scientifique, 1985), pp. 137–42; K. Kilinski, "Boeotian Trick Vases," *American Journal of Archaeology* 90 (1986), 153–8. Athenaeus has a mention of a horn-like cup (a *rhyton)* made by the μηχανοποιός Ctesibius for the temple of the deified queen Arsinoë, such that a trumpet sound was followed by the flowing of wine from a spout, *Deipnosophists* 11.497 ff.

[46] An English translation is in S. Murphy, "Heron of Alexandria's On Automaton-Making," *History of Technology* 17 (1995), 1–44.

[47] Hero *Mech.* 1.1–5. The *Mechanics* has come down to us in a ninth-century Arabic translation; translations are mine from a German translation: W. Schmidt, *Heronis Alexandrini Opera Quae*

Hero also tackles the general problem of the force necessary to move a body along a horizontal plane:

> Some people think that a weight lying on the ground will be moved only by a force equivalent to it, because they trust false appearances. We thus prove that a weight situated in the above-said position will be moved by a force smaller than any known force, and explain the reason why this is not evident.

He goes on to comment that it will only take a minimal inclination to displace the body from its state of rest and make it move. This happens particularly when the body is cylindrical or spherical, when it is very cohesive, and when both its surface and the surface on which it rests are very smooth. When instead one has to move a body up a plane, whether perpendicular or simply inclined to the horizon, a force equivalent to the weight will be necessary to counterbalance it, and then the addition of a small force will suffice to move the body.[48] In fact, even a condition of rest is seen as analogous to one of movement, in that a force is necessary in the latter case to move the body, in the former one to prevent the body from following its natural course downwards, i.e. to support the body. A section of the first book of the *Mechanics* is thus devoted to supports, such as columns – the architectural context of these issues is evident.[49] The five mechanical powers (i.e. lever, pulley, wedge, screw, and wheel and axle, here for the first time listed in full), introduced in the second book, are once again linked to the main problem of mechanics, because they are powers to move a body. Although they are very different, they are all dependent on one principle, Hero says; he repeats the idea (already present in the pseudo-Aristotle) that the law of the lever has to do with concentric circles only eventually to affirm that the balance is more useful as explanatory principle than the circle.[50] One of the key-words is again "utility," and wonder plays a role too. Indeed, readers are told that

> we will wonder at things which, when we have proved them, are the contrary of what is manifest to us. But the things, the causes of which we can talk about only according to the simplest principles, will increase our wonder even more when we see that the things which we employ are the contrary of what we are used to and what we hold for certain.[51]

Supersunt Omnia, 5 vols. (Leipzig: Teubner, 1899–1914), vol. 2. The problem of moving a thousand talents with a force of five talents is picked up again at 2.21.

48 Ibid. 1.20–3.

49 Ibid. 1.25–30.

50 Ibid. 2.1–5 for the five mechanical powers; 2.20 for the balance. The equivalence of circle and balance is stated at 2.9.

51 Ibid. 2.33.

WAR TECHNOLOGY AND THE ROMANS

The Roman army, which we have encountered helpless when faced with Archimedes' wondrous machines, is thus described by Flavius Josephus, an eye-witness to the siege of Jerusalem in 70 CE:

> Wonderfully constructed as were the engines of all the legions, those of the tenth were supreme. Their quick-firers were more powerful and their stone-projectors larger [. . .]. The rocks which they hurled weighed a talent and had a range of two furlongs or more; and their impact not only on those who first met it but even on those considerably in the rear was irresistible.[52]

The Romans seem to have learnt their military technological expertise from the Greeks. There are numerous testimonies of early borrowings: for instance, after the successful siege of New Carthage in 210 BCE, Scipio Africanus acquired for the *res publica* as public slaves 2,000 craftsmen, who were encouraged to collaborate by the hope of eventual freedom. For the siege of Utica in 204 BCE, new artillery and engines were being made in an arsenal by the thus acquired workforce.[53] Sulla is depicted laying siege to Athens in 87 BCE with the support of Thebes. He built new engines with wood timber taken from the grove of the Academy, and demolished the Long Walls for materials like stones.[54] The appropriation of Greek resources, both material and technical, could not be more complete. As we have seen with Vitruvius, the set of attitudes to war technology was determined by the values one associated with it. Frontinus, in his treatise on the military art, declines to talk at length about "works and engines of war, the invention of which has long since reached its limit, and for the improvement of which I see no further hope in the applied arts."[55] Where no improvement is possible on the technological front, the balance in a conflict is tipped by the personal qualities of the leader. Although leadership is not determined uniquely by technical expertise, the leader must be well-informed about the business at hand, because "there is nothing so disgraceful for a decent man as to conduct an office delegated to him, according to the instructions of assistants."[56]

[52] Josephus, *The Jewish War* 5.269–270.
[53] Livy 26.47.2; 29.35.8. The episode of New Carthage is also in Polybius 10.17.6–9.
[54] Appian, *The Mithridatic Wars* (12), 30.
[55] *Strategemata* 3 Introd.
[56] Frontinus, *On Aqueducts* Pref. 2.

PAPPUS OF ALEXANDRIA

Pappus' main work, the *Mathematical Collection*, was written some time around 320 CE and deals with various subjects. Book 8 is entirely devoted to mechanics, defined as:

> The mechanical enquiry [. . .] not only investigates the causes of what moves according to nature, but also moves what goes forcibly against nature from its own place towards a contrary motion. [. . .] The mechanicians around Hero say that there are a discursive and a manual part of mechanics; the theoretical part is composed of geometry, arithmetic, astronomy, and discourses about nature, the manual part of work in metals, architecture, carpentry, and painting, and of manual practice in these. They say that someone who has been trained from a young age in the aforesaid sciences and in addition has reached mastery of the aforesaid arts and has a versatile mind for these things will be the best architect and inventor of mechanical devices. [. . .] Of all the arts, the most necessary for the uses of life are: that of the makers of mechanical powers, they themselves being called mechanicians by the ancients (for they lift great weights by mechanical means to a height, contrary to nature, moving them by a lesser force); that of the makers of engines necessary for war, also called mechanicians (for they hurl missiles both of stone and of iron and suchlike objects to a great distance, by means of the instruments, known as catapults, made by them); in addition, the art of those who are in their turn especially called makers of machines (for they raise water from a great depth more easily by means of the instruments for water-drawing which they build). The ancients also call mechanicians the wonder-workers, of whom some practice their art by means of air, as Hero in *Pneumatica*; some by means of strings and ropes, thinking to imitate the movements of living things, as Hero in *Automata* and *Balances*; others by means of bodies floating in the water, as Archimedes in *On Floating Bodies*; or by telling the time by means of water, as Hero in *Hydria*, which seems to have something in common with the enquiry on sun-dials. They also call mechanicians those who know about the making of spheres, who build models of the sky, by means of uniform and circular movement of water.[57]

Pappus' definition of mechanics is the most inclusive that has come down to us from antiquity. While he does not concern himself with all of the branches described here, he can show his reader that he is acquainted with them all and well aware of the tradition before him. The definition

[57] *Mathematical Collection* 1022.1–1028.3, trans. I. Thomas, *Selections Illustrating the History of Greek Mathematics*, 2 vols. (Loeb Classical Library; Cambridge, MA: Harvard University Press, 1939–41), vol. 2, 615–19 with modifications.

attributed to (the people around) Hero is a sort of educational programme: a mechanician should ideally master both arts and sciences, both the discursive part and the manual part of his discipline. The fact that specialization inevitably ensues is not a matter of subordination of one aspect of mechanics to the other, but just a consequence of our human limits: it is not possible to know everything. Although he is no engineer (he claims to have no first-hand experience, and the weight-lifting machine he describes is taken from Hero), Pappus often refers to practical applications of his mathematical results: mathematics and mechanics are in his view complementary, and help each other as far as both proofs and applications are concerned. For instance, the problem of the duplication of the cube (equivalent to the problem of finding two mean proportionals between two given lines) is a mathematical problem, with several practical applications (for instance, the construction of catapults, as we have already seen), but it needs mechanics for its proof, because conic sections, which should be applied to its solution, are not easy to draw in a plane. Or again, some problems, such as finding the thickness of a cylinder whose bases have both been chopped off along irregular lines, are "outside the domain of geometry" and can be solved by mechanical means alone. Pappus also adds that the problem is frequent in architecture.[58] The intersection of mechanics and mathematics is present in Pappus also in his study of mechanical curves, i.e. spiral, quadratrics, and so on. He devotes a good part of book 4 of the *Collection* to these unnatural objects and stresses once again their importance for the solution of geometrical problems which cannot easily be solved otherwise. If we recollect Plutarch's depiction of the alleged contrast between Eudoxus and Archytas on the one hand and Plato on the other, it would seem that, if indeed there were philosophical prescriptions against the use of mechanics in mathematics, they were often ignored.

MECHANICS IN THE LATE ROMAN EMPIRE

Late antiquity saw a rise in status of the mechanical professions. One famous example is Cyriades, the senator, as well as "comes et professor mechanicae" mentioned by Symmachus.[59] The legal texts collected in late antiquity under the titles *Codex of Theodosius*, *Codex of Justinian*, and *Digesta* contain several decrees about mechanicians-architects. All through antiquity and, it would seem, especially in this period, the distinctions between these two terms are not clear. The Cyriades above was in charge of building a bridge and a basilica, i.e. tasks typical of an architect, and Procopius refers to the people involved with the construction of Hagia

[58] *Coll.* 1070.1 ff. and 1074.3 ff., respectively.
[59] *Relations* 25–6, set in 382–7 CE.

Sophia in Constantinople as mechanicians.[60] A group of laws, under the general heading "On the exemptions of technicians," established full immunity from fiscal obligations for, among others, architects, doctors, painters, and carpenters, with the exhortation that, in the spare time from their activities, they teach other people, in particular their children, the profession.[61] One of these laws explicitly maintains that "one needs as many architects as possible" and exhorts the prefect of the African provinces to encourage towards that career any youths in their twenties who have had a taste of "liberales litteras."[62] The 344 CE law unifies all the previous categories under the names "mechanicos et geometras [here to be taken as land-surveyors] et architectos"; their collective duties are described as administering boundaries, measuring (it is not specified what) and looking after aqueducts. Moreover, they are compelled to teach and enable others to teach in their turn: "equalIy with our words we order pursuit of teaching and learning, so that they may enjoy the immunities and raise enough teachers to teach in their turn."[63] These measures have been linked to increasing demand for qualified personnel to look after existing public buildings and fortresses and to construct new ones.

Mechanics applied to war is found in Vegetius, who, writing in the late fourth or early fifth century CE about the Roman army, mentions as components of the troops a number of workmen: "wood-workers, masons, carpenters, blacksmiths, painters and all other workers for [. . .] making the machines, the wooden towers, and the other equipment by which the cities of the enemy might be attacked or their own defended." The officer in charge of these men was the *praefectus fabrorum*. Vegetius also lists a number of siege machines, including *ballistae*, battering rams, and *testudines*, and devotes a section to sickle chariots.[64]

Sickle chariots were also one of the war machines described by the anonymous author of the so-called *De rebus bellicis*, a letter to two emperors (Valentinian I and Valens, or Constans II with one of his Caesars) written towards the end of the fourth century CE. The author, declaring himself worried about the general state of things in the Empire, proposes several reforms and the introduction of some war machines: among others, the sickle chariot already mentioned; a ship propelled by oxen; and two kinds of *ballista*. Addressing the emperor, the author says,

[60] Procopius, *On Buildings* 1.1.24,71,76; 2.3.11.
[61] *Codex of Theodosius* (henceforth "*Cod. Theod.*") 13.4 and *Codex of Justinian* 10.64. The laws are dated 334, 337, and 344 CE and are issued by Constantine and his successors Constantius and Constans.
[62] *Cod. Theod.* 13.4.1 (334 CE).
[63] Ibid. 13.4.3.
[64] *Epitome of Military Art* 2.11 and 4.13 ff., respectively (translation by L. F. Stelten, New York: P. Lang, 1990).

> you will double the strength of your invincible army when you have equipped it with these mechanical inventions, countering the raids of your enemies not by sheer strength alone but also by mechanical ingenuity, particularly when with keen perception you find machines that will be effective on all the elements.[65]

While *utilitas* is one of the words that recur most often, the author is also concerned about the astonishing effects of his machines; for instance, he introduces the ship (*liburna*) claiming that, being guided by human *ingenium*, it can vanquish ten ships, with no need for a large crew.[66] The arts are explicitly contrasted to other forms of service to the state:

> In this connection one has always to examine what a person means rather than what he says; for it is universally agreed that not the greatest nobility, nor abundance of resources nor the powers rooted in the legai courts or the eloquence acquired with the letters has come up to the advantages of the arts.[67]

Ammianus Marcellinus also describes war machines, employed by Julian in his war against the Persians in 363 CE: these include the *ballista*, the *onager* (which, he says, was formerly called "scorpion"), and the helepolis, the same tower first built by Demetrius Poliorketes, whom Ammianus duly mentions.[68] The presence of the past indeed loomed large, but it did not prevent the anonymous author of *De rebus bellicis* from introducing at least one of his machines, the armoured sickle chariot, with the words: "This astonishing machine possesses a certain novelty." And Procopius comments, of a novel type of battering ram used in the siege of Petra in 550 CE: "thus, as time goes on, ingenuity is ever wont to keep pace with it by discovering new devices."[69]

[65] *On Things of War* 18.7 (translation by E. A. Thompson, *A Roman Reformer and Inventor* (Oxford: Clarendon Press, 1952). Cf. also Marco Formisano, *Tecnica e scrittura: le letterature tecnico-scientifiche nello spazio letterario tardolatino* (Rome: Carocci 2001).
[66] *On Things of War* Pref. 12.
[67] Ibid. Pref. 6 (translation in Thompson, *A Roman Reformer and Inventor*, with modifications).
[68] Ammianus Marcellinus, 23.4.
[69] *On Things of War* 14.1 and Procopius, *On the Wars* 8.11.28.

24

GRECO-EGYPTIAN ALCHEMY

Cristina Viano[1]

ORIGINS, IDENTITY, ETYMOLOGY, AND HISTORIOGRAPHY

Alchemy came into being out of the meeting of Greek and Egyptian culture that occurred at Alexandria in the first centuries of the Common Era.[2] It developed between the first and seventh centuries as the theory and practice of transmuting noble metals in Greco-Roman Egypt. Thence it was transmitted to the Byzantine world, where it was preserved by a generation of commentators, and then to the Arabic world, which gave it a more systematic and experimental orientation. In the western medieval world, Alexandrian alchemy was known only indirectly, through the filter of translations and compilations produced by Arabic alchemists. The corpus of Greek alchemists was rediscovered and reintroduced in Renaissance Italy, but outside of a small circle of scholars it was not disseminated widely among humanists and adepts.[3]

Greco-Alexandrian alchemists saw the origins of their art in Pharaonic Egypt, a thesis which most historians have accepted. Links with Mesopotamian, Indian, and Chinese alchemy have also been assumed.[4] However, besides some similarities in themes or processes, there is at this

[1] I wish to thank Marc Aucouturier, Michèle Mertens, and Matteo Martelli for their extremely helpful comments and bibliographical suggestions. The editors warmly thank Laurence Totelin for her translation of this chapter.

[2] See A. J. Festugière, *La révélation d'Hermès Trismégiste, vol. 1: L'Astrologie et les sciences occultes* (Paris: Lecoffre, 1944), p. 218: "L'alchimie gréco-égyptienne, d'où ont dérivé toutes les autres, est née de la rencontre d'un fait et d'une doctrine. Le fait est la pratique, traditionnelle en Egypte, des arts de l'orfèvrerie. La doctrine est un mélange de philosophie grecque, empruntée surtout à Platon et à Aristote, et de rêveries mystiques."

[3] For an overview of alchemy from its origins to the modern period, see M. Pereira, *Arcana sapienza: L'Alchimia dalle origini a Jung* (Rome: Carocci, 2001) and M. Pereira, *Alchimia: I testi della tradizione occidentale* (Milan: A. Mondadori, 2006). On the dissemination of Greek alchemy during the Renaissance, see S. Matton, "L'Influence de l'humanisme sur la tradition alchimique," *Micrologus* 3 (1995), 279–345.

[4] On the question of origins, see H. J. Sheppard, "Alchemy: Origin or Origins?," *Ambix* 17 (1970), 69–84; P. T. Keyser, "Alchemy in the Ancient World: From Science to Magic," *Illinois Classical*

point no concrete or decisive proof of such connections. Indeed, Greco-Alexandrian alchemy is a distinctive, complex, and unique phenomenon – it is very different, for instance, in its aims and methods from magic. Alchemists define themselves as "philosophers." Plato and Aristotle appear atop lists of old masters of the art; some alchemists are referred to as "exegetes of Plato and Aristotle."

The origin of the word "alchemy" is obscure. The use of the word *chêmeia* and its cognates is rare in the writings of Greek alchemists; instead they refer to their discipline with the phrase "divine and sacred art."[5] The Latin noun *alchimia*, which only appears in the twelfth century, is in fact composed of the Arabic article *al* and a root whose meaning is disputed. Greek alchemists mention Chymes, sometimes considered as the author of a book entitled *Chemeu*. *Chêmeia* has also been thought to derive from *cheô* (to melt); from *chumos* (sap extracted from plants); from the old name of Egypt, which is *chêmia* in Plutarch and KHME or XHMI in Coptic (that is, "black earth"); and from the Egyptian root *km*, which means "to achieve." Some even believe that "black" was an allusion to the first step of transmutation ("the black work").[6]

The works of the chemist Marcellin Berthelot (1827–1907) gave rise to an interest in Greek alchemy amongst modern scholars. *Les origines de l'alchimie* (1885) and the *Collection des anciens alchimistes grecs* (1888–9) (= *CAAG*), published in collaboration with the Hellenist Charles-Émile Ruelle, are characterized by their rationalist historical approach. Berthelot found in ancient alchemy the origins of the experimental method and saw it as the precursor to modern chemistry. The *Collection* has been criticized for its rather poor philological rigor, but it should be given due credit for disseminating these texts and stimulating scholarly interest. Between 1924 and 1932, the eight volumes of the *Catalogue des manuscrits alchimiques grecs* (= *CMAG*) were published, in preparation for a new, more complete and rigorous textual edition. Such an enterprise was at last launched in 1981 at Paris, with the Belles Lettres series, *Les alchimistes grecs*.[7]

Studies 15 (1990), 353–72. On Chinese alchemy, see J. Needham, *Science and Civilisation in China, vol. 5: Chemistry and Chemical Technology* (Cambridge: Cambridge University Press, 1974), part. 2.

5 The variants *chêmeia, chêmia, chumeia, chumia* are found in late Greek, especially in the works of Byzantine chroniclers. See, for instance, the Byzantine lexicon *Suda* (tenth century), which defines *chêmeia* as the "art of preparing silver and gold."

6 For the meanings of the word "alchemy," see J. Lindsay, *The Origins of Alchemy in Graeco-Roman Egypt* (London: Muller, 1970), pp. 68–89; R. Halleux, *Les textes alchimiques* (Turnhout: Brepols, 1979), pp. 45–6; D. Bain, "*Melanitis gê*. An Unnoticed Greek Name for Egypt: New Evidence for the Origins and Etymology of Alchemy," in D. R. Jordan, H. Montgomery, and E. Thomassen (eds.), *The World of Ancient Magic* (Bergen: Norwegian Institute at Athens, 1999), pp. 221–2.

7 The collection currently comprises four volumes: R. Halleux (ed.), *Les alchimistes grecs, vol. 1: Papyrus de Leyde, papyrus de Stockholm, recettes* (Paris: Belles Lettres, 1981); M. Mertens (ed.), *Les alchimistes grecs, vol. 4: Zosime de Panopolis, mémoires authentiques* (Paris: Belles Lettres, 1995); A. Destrait-Colinet (ed.), *Les alchimistes grecs, vol. 10: Anonyme de Zuretti* (Paris: Belles Lettres, 2000); A. Destrait-Colinet (ed.), *Les alchimistes grecs, vol. 11: Recettes alchimiques (Par. Gr. 2419; Holkhamicus 109) Cosmas le Hiéromoine – Chrysopée* (Paris: Belles Lettres, 2010).

Beside the historical and scholarly approach, which aims at a general overview based on textual criticism and the contextualization of alchemical authors, there are other approaches that focus on particular aspects, such as, for instance, psychology and cultural anthropology. Carl Gustav Jung interprets Zosimus' (the most significant figure in the history of Greco-Egyptian alchemy) *Visions* as subconscious archetypes.[8] Mircea Eliade compares the myths and rituals of alchemists with the symbols that characterize archaic societies.[9] Paradoxically, these erudite interpretations have tended to emphasize the irrational and mystical aspects of alchemy. They have also promoted the easiest, and also the least rigorous, approach to alchemy, that is, modern hermeticism, which studies alchemy in a completely non-critical manner as a revelation inherited from ancient civilizations and transmitted through initiation.

In this chapter, we will start with an introduction to the sources of our knowledge of Greek alchemy (the papyri, the manuscripts, the testimonies within and without the alchemical corpus), and to the alchemical authors and their texts. We shall then deal with practical aspects of alchemy, such as ingredients, processes, and apparatus, and with alchemical theory, its links with Greek philosophy, as well as its esoteric features. We shall conclude by presenting some methodological directions for future research, which are based on interdisciplinary collaboration.

SOURCES: PAPYRI, MANUSCRIPTS, TESTIMONIES

Our sources for Greek alchemy essentially consist of the manuscript tradition and the testimonies given by alchemists themselves; ancient testimonies external to the alchemical corpus are extremely rare.

The writings of Greco-Egyptian alchemists have been transmitted to us by means of two compilations on papyrus dating to the third and fourth centuries CE, now respectively at Leiden and Stockholm, and by a large corpus produced in the Byzantine period, which is preserved in multiple manuscripts. Considered the most important and most beautiful of these manuscripts by most scholars, the *Marcianus Graecus* ("*Marc. Graec.*") 299 (tenth or eleventh century) belonged to the library of Cardinal Bessarion (fifteenth century).[10]

Alchemical literature is in essence fragmentary, constituted of extracts, of collections of quotations, of commentaries and of précis composed by

8 See, for instance, C. G. Jung, "Einige Bemerkungen zu den Visionen des Zosimos," *Eranos-Jahrbuch* 5 (1937), 15–54.

9 M. Eliade, *Forgerons et alchimistes* (Paris: Flammarion, 1956). On the different approaches to alchemy, see Halleux, *Textes alchimiques*, pp. 50–8.

10 See H.-D. Saffrey, "Historique et description du Marcianus 299," in D. Kahn and S. Matton (eds.), *Alchimie: art, histoire et mythes* (Paris: SEHA, 1995), pp. 1–10.

compilators. Interpolations and additions by copyists, for the most part specialists who did not hesitate to intervene in the texts to make comments or to correct them, are numerous. The Greek of these texts is often incorrect, and the vocabulary relating to substances and transformations is still, in great part, in need of deciphering.

As for external testimonies, it is only in the fifth century that Proclus and Aeneas of Gaza started to refer to alchemy as a contemporary practice, aiming at producing gold from other metals.[11] Byzantine chroniclers mention the destruction of the books "concerning the chemistry [*chemeia*] of gold and silver," ordered by Diocletian in order to deprive the Egyptians of a source of wealth and thus prevent them from competing with the Romans.[12] This testimony is particularly interesting because it shows that in the third century CE, alchemical practice must have been important and sufficiently recognized for the Romans to seek the destruction of its books.

TEXTS AND AUTHORS OF THE ALCHEMICAL CORPUS

Greek alchemical literature is divided into three periods, according to a chronological development.[13] The first period includes the chemical recipes from the *Phusika kai mustika*, attributed to pseudo-Democritus (first to third century CE), who is not the philosopher from Abdera and who has been associated with a certain Bolos of Mendes,[14] as well as the anonymous papyri of Leiden and Stockholm (ca. third to fourth centuries CE). The recipes deal with imitating gold, silver, precious gemstones, and purple. Within these recipes the principles of a fundamental unity of matter and of sympathetic relationships between substances are expressed, through the famous formula frequently repeated in the alchemical corpus, "nature is delighted by nature, nature conquers nature, nature masters nature" (*hê phusis tê phusei terpetai, kai hê phusis tên phusin nika kai hê phusis tên phusin kratei*). In these recipes the model for producing gold appears to be that of imitation through coloring, which acts upon the external properties of bodies. This notion of imitation is the cornerstone of the ancient conception of the alchemical technique and is a precursor to the idea of transmutation. To this period also belongs a series of quotations or short treatises of the

[11] Proclus, *In Platonis rem publicam*, 2.234.14–22; Aeneas of Gaza, *Theophrastus*, 71, edited by Caspar Barth (Leipzig: Johannes Bauerus, 1655).

[12] John of Antioch, frag. 165; Carl Müller, *Fragmenta historicorum graecorum*, 5 vols. (Paris: A. Firmin Didot, 1868), vol. 4; *Suda, Delta* 1156 s.v. *Diokletianos; Khi*, 280, s.v. *chêmeia*.

[13] See H.-D. Saffrey, "Introduction," in Halleux (ed.), *Les alchimistes grecs*, vol. 1, xii.

[14] See M. Wellmann, *Die Physika des Bolos Demokritos und der Magier Anaxilaos aus Larissa, Teil 1* (Abhandlungen der Preussischen Akademie der Wissenschaften, Philosophisch-Historische Klasse 7; Berlin: de Gruyter, 1928); Halleux (ed.), *Les alchimistes grecs*, vol. 1; P. Gaillard-Seux, "Un pseudo-Démocrite énigmatique: Bolos de Mendès," in F. Le Blay (ed.), *Transmettre les savoirs dans les mondes hellénistique et romain* (Rennes: Presses Universitaires de Rennes, 2009), pp. 223–43.

mythical "old authors" such as Hermes, Agathodemon, Isis, Cleopatra, Maria the Jewess (according to the tradition, Maria invented a cooking technique still employed today in our kitchens called "*bain-Marie*"), Ostanes, Pammenes, and Pibechius (all between the first and third centuries CE).

The second period is that of the authors *stricto sensu*: Zosimus of Panopolis; Pelagius; and Iamblichus (third to fourth centuries CE). Zosimus, a native of Panopolis in Egypt, may have lived at Alexandria around 300 CE. We possess various fragments of his works, collected into four groups in the manuscripts: the *Authentic Memoirs*; the *Books to Eusebia*; the *Books to Theodorus*; and *The Final Quittance* with two extracts from the *Book of Sophe*. One of the main problems consists in identifying the twenty-eight books *kata stoicheion* (that is, in alphabetical order) that are mentioned in the Byzantine lexicon *Suda* and which appear to comprise the entire work of Zosimus, and associating them with the titles that have been transmitted directly or indirectly by the tradition. Among the most famous pieces one must mention *On the Letter Omega* and the three *Visions*, which are part of the *Authentic Memoirs*. The *Visions* describe the dreams in which the properties of metals were revealed to Zosimus.[15] The processes for metal transformations are accompanied by speculations concerning the nature of matter and by a ritual symbolism centered on the notions of death and resurrection, thus allowing two levels of interpretation – one technical and the other mystical.[16] As mentioned earlier with reference to the various approaches to alchemy, Jung took particular interest in the *Visions*, which he interpreted as archetypal images reflecting identification processes between the operator and his materials. Beside the Greek tradition, works attributed to Zosimus survive in Syriac, Arabic, and Latin versions.[17] The Syriac fragments (Cambridge University Library, Syriac MS 6.29) contain the only ancient recipes for the manufacture of the famous "black Corinthian bronze" much appreciated by the Romans and mentioned by Pliny the Elder.[18]

The third period, that of the commentators, begins in the fourth century with Synesius, who wrote a commentary on pseudo-Democritus' *Physica kai Mystica* in the shape of a dialogue entitled *Syneius to Dioscorus, Commentary on the Books of Democritus*, in which he explicitly declares his exegetic

[15] See M. Plessner, "Zosimus," in C. G. Gillispie (ed.), *Dictionary of Scientific Biography, vol. 14* (New York: Charles Scribner's Sons, 1976), pp. 631–2; Mertens (ed.), *Les alchimistes grecs*, vol. 4; H. M. Jackson, *Zosimos of Panopolis on the Letter Omega* (Missoula, MT: Scholars Press, 1978); F. Tonelli, *Zosimo di Panopoli: Visioni e risvegli* (Milan: Coliseum, 1988).

[16] Halleux, *Textes alchimiques*, p. 64.

[17] On this topic, see B. Hallum, "*Zosimus* Arabus: The Reception of Zosimos of Panopolis in the Arabic/Islamic World" (PhD thesis, Warburg Institute, London, 2008).

[18] Pliny, *Natural History* 34.8. See E. C. D. Hunter, "Beautiful Black Bronzes: Zosimos' Treatises in Cam. Mm. 6.29," in A. Giumlia-Mair (ed.), *I bronzi antichi: produzione e tecnologia. Atti del XV Congresso internazionale sui bronzi antichi* (Montagnac: Monique Mergoil, 2002), pp. 655–60.

intentions, namely, to disclose Democritus' thought, the sequence of his teachings, and his processes and materials.[19] In the sixth century, Olympiodorus produced a commentary on a lost treatise of Zosimus, the *Kat' energeian* (*On Action* or *According to Action*) and on some sentences by other ancient alchemists. Olympiodorus' commentary is characterized by a doxography in which the doctrines of nine monist Presocratic philosophers are compared to the alchemical doctrine of the unity of matter.[20] In the seventh century, Stephanus of Alexandria, in his nine *Lectures* (*Praxeis*) dedicated to Heraclius (r. 610–41), comments on the ancient alchemists in a highly rhetorical style and links alchemy with medicine, astrology, mathematics, and music.[21] According to the Arabic-Latin tradition, it was one of Stephanus' students, the monk Morienus (Marianus), who transmitted alchemy to the Arabic world by initiating, between 675 and 700, the Umayyad prince Khalid ibn Yazid.[22] These authors aim principally at clarifying the thought of their predecessors and they represent the most sophisticated stage in the theorization of Greek alchemy. We find in their work an attempt at defining and systematizing alchemical doctrine by means of Greek philosophy. Considering the high number of pseudepigraphic works in the alchemical tradition, scholars have wondered whether or not to identify Olympiodorus and Stephanus with their homonyms, the Neo-Platonic commentators. This identification – which is now commonly accepted – is fundamental as it undermines the alleged marginality of alchemy.[23] Close to Stephanus are also four poems *On the Divine Art*, attributed to Heliodorus,[24] Theophrastus, Hierotheus, and Archelaus (seventh to eighth centuries). Two anonymous commentators, usually referred to as the Christian Philosopher and the Anepigraph (sixth to eighth

[19] M. Martelli, *Pseudo-Democrito. Scritti alchemici con il commentario di Sinesio* (Paris: SEHA, 2011).

[20] See C. Viano, "'Olympiodore l'alchimiste et les Présocratiques: une doxographie de l'unité' (*De arte sacra*, §18–27)," in Kahn and Matton (eds.), *Alchimie*, pp. 95–150; C. Viano, "Gli alchimisti greci e l''acqua divina'," *Atti del VII Convegno Nazionale di storia e fondamenti della chimica* 115 (1997): *Memorie di Scienze Fisiche e Naturali*, pp. 61–70; C. Viano, "Olympiodoros of Alexandria" and "Stephanos of Alexandria," in P. T. Keyser and G. Irby-Massé (eds.), *The Encyclopedia of Ancient Natural Scientists* (London: Routledge, 2008), pp. 589–90 and 760–1.

[21] Stephanus' *Praxeis* have been edited by J. L. Ideler, *Physici et medici graeci minores*, 2 vols. (Berlin: Reimeri, 1841–2) and in parts by F. Sherwood Taylor, "The Alchemical Works of Stephanus of Alexandria," *Ambix* 1 (1937), 116–39, and 2 (1938), 39–49. See also M. Papathanassiou, "Stephanos von Alexandreia und sein alchemistisches Werk. Die kritische Edition des griechischen Textes eingeschlossen" (PhD thesis, Humboldt Universität zu Berlin, 1992); M. Papathanassiou, "L'Oeuvre alchimique de Stéphanos d'Alexandrie: structures et transformations de la matière, unité et pluralité," in C. Viano (ed.), *L'Alchimie et ses racines philosophiques: la tradition grecque et la tradition arabe* (Paris: Vrin, 2005), pp. 113–33; M. Papathanassiou, "Stephanos of Alexandria: A Famous Byzantine Scholar, Alchemist and Astrologer," in P. Magdalino and M. Mavroudi (eds.), *The Occult Sciences in Byzantium* (Geneva: La Pomme d'Or, 2006), pp. 163–204.

[22] E. Bacchi and M. Martelli, "Il principe Halid bin Yazid e le origini dell'alchimia araba," in D. Cevenini and S. D'Onofrio (eds.), *Conflitti e dissensi nell'Islam* ('Uyûn al-Akhbâr: studi sul mondo islamico 3; Bologna: Il Ponte, 2009), pp. 85–120.

[23] Viano, "Olympiodoros of Alexandria" and "Stephanos of Alexandria."

[24] G. Goldschmidt (ed.), *Heliodori carmina quattuor ad fidem codicis Cassellani* (Giessen: A. Töpelmann, 1923).

centuries?), lead us directly to the period when large compilations were produced, and in particular the most important alchemical manuscript, the *Marc. Graec.* 299 (tenth to eleventh centuries). The Byzantine alchemical tradition ends with Michael Psellus (eleventh century),[25] Nicephorus Blemmydes (thirteenth century), and Cosmas (fifteenth century).[26] Among the alchemical works in Greek, one must also mention the *Anonymous of Zuretti*, a Byzantine treatise produced in Calabria in 1378 and compiled from Latin sources.[27]

PRACTICAL ASPECTS: INGREDIENTS, PROCESSES, AND APPARATUS

The products listed in Greek alchemical texts are numerous and often expressed in a codified language that is difficult to interpret. In addition, one can assume that they were never chemically pure. Among the materials that can be identified, the most important are certainly metals: gold; silver; copper; mercury; iron; tin; and lead. In general, they are called "bodies" (*sômata*), in contrast to the "incorporeals" (*asômata*), a word which refers to other minerals. Also classified among the bodies are metallic alloys, such as, for instance, the debated *claudianos* and the *molybdochalcos* (copper-lead) or *electrum* (gold-silver). The alchemists also made use of numerous native ores, such as *natron* (a typically Egyptian ingredient from Wadi Natrun used notably in the process of mummification), marble, salt, and, above all, sulfur. Often, as in the case of sulfur, names are used in the broadest sense and also refer to similar products. Finally, one must mention some vegetable and organic ingredients, such as honey, urine, and milk. With regards to the production of noble metals, the processes that are described can be essentially interpreted as productions of alloys, which sometimes included gold and silver in varying proportions, and as surface-coloring processes. A typical process for falsifying gold and silver was the *diplôsis* (literally, duplication), which aimed at duplicating the mass of these noble metals by associating them with another metal. Falsification methods are closely linked to test methods.[28] This is the case for the process of cementation, which can be defined as "the process of heating a metal in the presence of a preparation that modifies its physical properties, either by attacking it or by combining it

[25] J. Bidez (ed.), *CMAG, vol. 6: Michel Psellus: épître sur la Chrysopée* (Brussels: Maurice Lamertin, 1928), pp. 26–47; F. Albini and Michele Psello, *La Crisopea ovvero come fabbricare l'oro* (Genoa: Edizioni Culturali Internazionali, 1988).

[26] Destrait-Colinet (ed.), *Les alchimistes grecs*, vol. 11.

[27] C. O. Zuretti (ed.), *CMAG, vol. 2: Les manuscrits italiens* (Brussels: Maurice Lamertin, 1930); Destrait-Colinet (ed.), *Les alchimistes grecs*, vol. 10.

[28] R. Halleux, "Méthode d'essai et d'affinage des alliages aurifères dans l'Antiquité et au Moyen-âge," in C. Morrisson et al. (eds.), *L'Or monnayé 1* (Paris: CNRS, 1985), pp. 39–77.

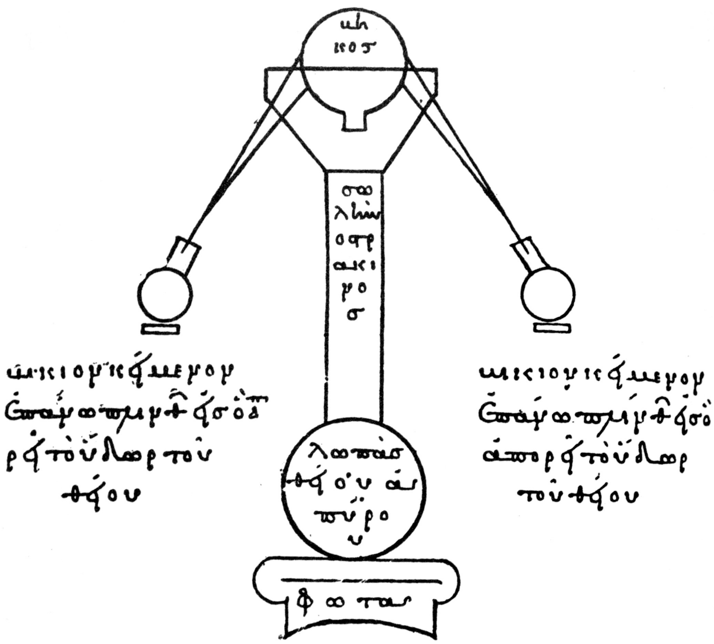

Figure 24.1. Alembic, *Marc. graec. 299*, fol. 193v (Berthelot 1889, p. 138).

with another metal that is not present in a metallic state."[29] For instance, a cement composed of salt and sulfurous products is intended to reduce the proportion of silver associated with native gold. As for surface treatments of metals, recipes also describe coloring processes by means of chemical reactions, as in the case of the recipes for "black bronze" preserved in the Syriac version of Zosimus' writings. These various processes share a common aim, namely a change in color. If one takes into account the fact that, in antiquity, color was thought to reflect the composition itself of a body, one understands better the close link between the idea of coloring and that of "making" a metal.[30] In the works of Zosimus are descriptions of most apparatus, among which the most famous are the alembic and the *kerotakis*. The alembic (in Greek, *ambix*, which will give us, via the intermediary of the Arabic *'al-'anbiq*, our word "alembic") is the most basic of distillation apparatuses (Figure 24.1). The process of distillation aims at converting into vapor a liquid mixed with a nonvolatile body, or liquids mixed with each other, in order to separate them. *Kerotakis* (from *kêros* = wax and *têkô* = to melt) appears to be a more complex apparatus, the original meaning of

[29] Halleux (ed.), *Les alchimistes grecs*, vol. 1, 40.

[30] Ibid., vol. 1, 77; A. J. Hopkins, "Transmutation by Color: A Study of Earliest Alchemy," in J. Ruska (ed.), *Studien zur Geschichte der Chemie: Festgabe E. O. von Lippman* (Berlin: Springer, 1927), pp. 9–14.

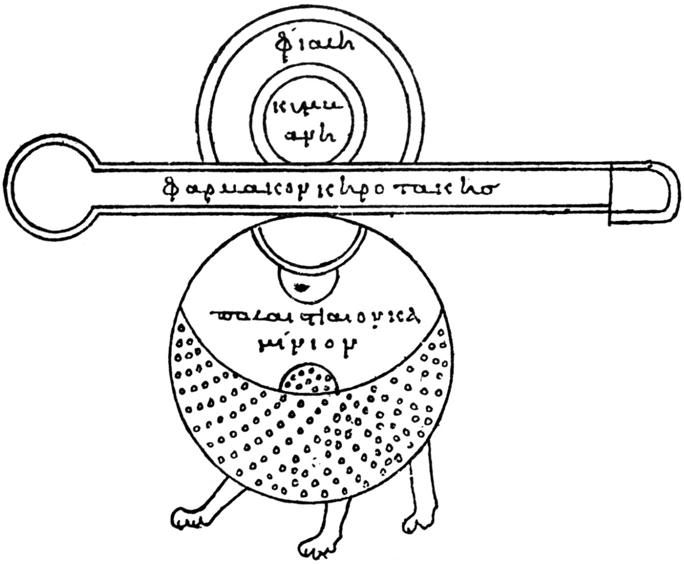

Figure 24.2. Kerotakis, *Marc. graec. 299*, fol. 195v (Berthelot 1889, p. 146).

whose name and function are more difficult to determine (Figure 24.2). In general, its aim seems to have been to dye gold leaves through cyclic vaporization of some alloys that are difficult to produce.[31]

THE THEORY OF TRANSMUTATION AND ITS MODELS

The entire development of Greek alchemy, from the first recipe compilations to the theoretical reflections of authors and commentators, is characterized by a dialectical tension between theory and practice. This tension is closely linked to the ambiguous relationship that exists between imitation and actual production of gold, between *aurifiction* and *aurifaction*, to use the famous distinction drawn by Joseph Needham.[32] The intention emerging from the earliest technical treatises is to produce a dye that would be an "imitation" of gold. Increasing theoretical complexity progressively transformed the notion of imitation into the ideal aim of total transformation. The idea of transmutation is based on the conception whereby all metals are constituted of a single matter. One needs first to remove the qualities that distinguish a metal by bringing it back to the primordial, indeterminate metallic matter, and then to give it the properties of gold.

[31] For a precise and complete description of Zosimus' apparatuses, see M. Mertens, "Introduction," in Mertens (ed.), *Les alchimistes grecs*, vol. 4, cxiii–clxix.

[32] Needham, *Science and Civilisation*, vol. 5.

The production of gold therefore results from a synthesis, which has as its starting point a primordial metallic matter, common and receptive, to which are added "qualities," that is, matters, according to sympathetic principles. It is difficult to identify the products allegedly responsible for the processes of coloration or transformation into gold. Among these products, "divine water" or "sulfuric water" (*theion hudôr*) plays a fundamental role. In the phrase *theion hudôr* lies a fundamental ambiguity, often exploited by alchemists, since the Greek word *theion* is at the same time the noun "sulfur" and the adjective "divine."[33] This water is often referred to in texts as the principal agent in the process of transmutation and as the true aim of alchemical research. The functions of this product are numerous and often antithetical, as for instance the fixing of color and the dissolution of metal. It is produced through distillation by means of alembics (see Figure 24.2). Divine water is sometimes assimilated with the primordial matter of metals. If that assimilation makes it even harder to identify this water, it nevertheless brings to light a fundamental characteristic of the alchemical doctrine: the qualitative affinity and the substantial unity between the principle of transmutation and the matter to be transmuted. In the manuscripts, the unity of matter is often represented by the image of a serpent biting its own tail, the *Ouroboros* (Figure 24.3).[34]

The theoretical reflection on alchemical processes clearly and openly borrows its conceptual apparatus and vocabulary from Greek philosophy. It was principally developed by the commentators, who explicitly claim the Presocratic philosophers, Plato, and Aristotle as their predecessors.[35] We are not dealing here with an unconscious and passive legacy, but rather with an active and constructive borrowing, with the view of building a new idea of transmutation.

If alchemy, as the study of metals and its transformations, can, from an epistemological point of view, find its place in an Aristotelian classification of natural sciences, the idea of transmutation is incompatible with some fundamental Aristotelian principles, the first of which is the fixity of species. Hence the famous passage in Avicenna's *De congelatione et conglutinatione lapidum*, long attributed to Aristotle himself: "Thus alchemists must know

[33] On divine water, see C. A. Wilson, "Philosophers, Iôsis and Water of Life," *Proceedings of the Leeds Philosophical and Literary Society* 19 (1984), 101–219; Viano, "Gli alchimisti greci"; M. Martelli, "'Divine Water' in the Alchemical Writings of Pseudo-Democritus," *Ambix* 56 (2009), 5–22.

[34] See H. J. Sheppard, "The Ouroboros and the Unity of Matter in Alchemy: A Study in Origins," *Ambix* 10 (1962), 83–96.

[35] The atomists and the Stoics have often been included among the philosophical models for Greek alchemy. It is true that tradition regards "Democritus" as one of its main authorities, but in reality, the atomic theory has very little bearing on Greek alchemy. As for the Stoics, if it is possible to find some similarities with some tenets of their physics (for instance, the doctrine of *pneuma* or the sympathy between substances), they are nevertheless conspicuously absent in the Greco-Alexandrian corpus. For the debate on this question, see S. Matton, "Alchimie et stoïcisme: à propos de récentes recherches," *Chrysopœia* 5 (1992–6), 6–144.

Figure 24.3. Ouroboros, *Marc. graec. 299*, fol. 188v (Berthelot 1889, p. 132).

that they cannot change the species of metals" (*quare sciant artifices alkimie species metallorum mutare non posse*).[36]

It is interesting to note that, in spite of theoretical obstacles, alchemy has always been dominated by Aristotelian physical and metaphysical models. Indeed, all medieval alchemy has attempted to negotiate with Aristotle the arguments that mainstream philosophy levels against transmutation: species fixity; unknowability of specific differences; and the "weakness" of the art compared to nature. This ambiguity is one of the principal characteristics of alchemy.

In view of these aspects of alchemical practice and theory, one can understand the enthusiasm of the chemist Marcellin Berthelot, who saw in Greek alchemy the foundations of modern chemistry. Indeed, although it later limited itself essentially to metals and minerals, Greco-Egyptian alchemy studied the composition of bodies and the rules of transformations; it carried out analyses and combinations by using and progressively perfecting its technical apparatus. In fact, the words *chymia* and *alchymia* remained synonymous until the seventeenth century. It is only during the Enlightenment that the two notions became separate and that alchemy was definitely discredited as a nonscientific practice.[37]

[36] E. J. Holmyard and D. C. Mandeville (eds.), *Avicennae de congelatione et conglutinatione lapidum* (Paris: Paul Geuthner, 1927), pp. 53–4.

[37] See D. Diderot (ed.), *Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers*, 35 vols. (Paris: Briasson, 1753), vol. 3, 417–24, s. v. "chymie."

ALCHEMICAL ESOTERISM: TECHNICAL SPECIALIZATION AND SPIRITUAL INTERIORIZATION

In Greek alchemy, the two fundamental aspects of "esoterism" can be found that are discernible in the etymology of the word (*esoterikos* = internal): technical specialization and mystical interiorization.

First, as is the case for all specialist knowledge, technical or philosophical, alchemy requires an apprenticeship, an initiation to "internal doctrines" in the Aristotelian meaning of the word, that is, "internal to a school." The alchemists are "initiates" (*mustai*) and have access to *mustêria* stemming from a revelation originally made by a god, a prophet, or some other mystical and ancient character. In the *Letter from Isis to Horus*, the secret of gold and silver production is the product of a revelation made by an angel to Isis, who in turn transmits it to Horus, her only legitimate son.[38] We have already mentioned that Zosimus received in dreams revelations concerning the properties of metals.

The practice of deliberately occulting truth is a constant feature in alchemical literature, acknowledged by alchemists themselves – hence the praise of silence, the lack of clarity in expression, the allegories, the metaphors, and the symbolic language.[39] In the *Final Quittance*, Zosimus accounts for the political and economic origin of this practice of secrecy.[40] It tells how the artisans of the Egyptian kings, appointed to increase royal wealth, were supposed to keep the secret in order not to share with others "the dominating power of wealth." This reminds us of the reason for the destruction of alchemical books given by Diocletian and reported by the Byzantine chroniclers. Similarly, says Zosimus, Democritus and the "ancients," who were the friends of the kings of Egypt, have hidden this art in the interest of the kings, but also as a result of a, so to speak, scientific jealousy.

The binomial jealousy/secrecy is common in the corpus of Greek alchemical texts. One often finds the contrast between the "jealous," who have hidden the truth under a multiplicity of words, and the philosophers "without jealousy," who express themselves clearly.

In the writings of the commentators, whose principal aim is to clarify the work of the ancients, the lack of clarity of those ancients is often considered to be in fact apparent. Olympiodorus, for instance, in the case of divine water, shows that the allegorical and philosophical language of the ancients

[38] M. Mertens, "Une scène d'initiation alchimique: la lettre d'Isis à Horus," *Revue d'Histoire des Religions*, 205 (1988), 3–24.

[39] The list and description of alchemical signs used in the corpus has been published by C. O. Zuretti, *CMAG*, vol. 8: *Alchemistica signa* (1932). On secret language, see B. Vickers, "The Discrepancy between *Res* and *Verba* in Greek Alchemy," in *Alchemy Revisited*, ed. Z. R. W. M. von Martels (Leiden: Brill, 1990), pp. 21–33.

[40] *CAAG*, vol. 2, 239, 12 ff.

is in reality clear and without jealousy. He finds the origin of enigmatic phrases in the writings of Plato and Aristotle and he explains that their aim is to encourage scholars to pursue their inquiry beyond physical phenomena.[41]

Stephanus distinguishes between a "mythical" alchemy and a "mystical" (symbolic, initiatic) alchemy. While the first is to be found in a multitude of discourses, the second, inspired by God, proceeds methodically.[42] In this perspective, the allegorical language of alchemy is linked to a rational methodology. It is one of the characteristic features that Greek thought lends to alchemy thus allowing, through the centuries, the constitution of a common ground for dialogue with philosophy.

The other aspect of alchemical esoterism is that of a road "within," not only within the discipline, but also within oneself. In Zosimus, one finds the hermetic, or more generally Gnostic, dogma of the knowledge of God and of the welcoming of God into one's self.[43] The separation of the soul from the body (see the frequent allusions to the "internal" or "pneumatic" man) becomes the ideal aim of an internal transformation of man, which appears to occur in parallel with transformations of metals.

Alchemy therefore assumes the characteristics of a revealed religion that aims at the salvation of the soul and the mystical union with divinity. However, one can observe in the Greek alchemical texts, and most particularly in Zosimus, that this aspect is accompanied by, and does not exclude, the elaboration of theoretical and technical principles according to a rational method.

DIRECTIONS FOR FUTURE RESEARCH: TOWARDS AN INTERDISCIPLINARY APPROACH

The nature of Greco-Alexandrian knowledge is undeniably dual: theoretical and practical. It consists of texts and recipes which concern mystical, cosmological, and physical reflections, and the production of concrete objects that are historically identifiable, such as the working and dyeing of metals, fabrics, and precious stones. Alchemy is not merely the ideal aim, dreamt of but never attained, of the *chrysopoeia*, that is, the fabrication of gold from other precious metals. The earliest texts are probably artisans'

[41] *CAAG*, vol. 2, 70, 4 ff.

[42] Ideler, *Physici et medici*, vol. 2, 208.

[43] On Zosimus' Gnosticism, see A. J. Festugière, "Alchymica," *Antiquité classique*, 8 (1939), 71–95, reprinted in his *Hermétisme et mystique païenne* (Paris: Aubier-Montaigne, 1967), pp. 205–29, p. 210; M. Mertens, "Alchemy, Hermetism and Gnosticism at Panopolis c. 300 AD: The Evidence of Zosimus," in A. Egberts, B. P. Muhs, and J. van der Vliet (eds.), *Perspectives on Panopolis: An Egyptian Town from Alexander the Great to the Arab Conquest* (Leiden: Brill, 2002), pp. 165–75; M. Mertens, "Zosimos of Panopolis," in N. Koertge (ed.), *New Dictionary of Scientific Biography, vol. 7* (Detroit, MI: Charles Scribner's Sons, 2008), pp. 405–8; K. Fraser, "Zosimos of Panopolis and the Book of Enoch: Alchemy as Forbidden Knowledge," *Aries* 4 (2004), 125–47.

Figure 24.4. Inkwell from Vaison la Romaine, first century BCE, Louvre. Cliché C2RMF Dominique Bagault. Cf. S. Descamps and M. Aucouturier, "L'encrier de Vaison la Romaine et la patine volontaire des bronzes antiques," *Monuments Piot*, 24 (2005), 5–30.

workbooks, produced in the circle of the Egyptian kings' goldsmiths. It is for that reason that the study of Greek alchemy is a topic at the boundaries between the history of philosophy, philology, and the history of science and technology. It is a composite field which requires the gathering of various skills, not only theoretical and historical, but also practical and technical – skills that deal with material culture, such as archaeology, metallurgy, and chemistry, which study materials and their transformation through artistic processes.[44] An excellent example of interdisciplinary collaboration, involving philologists, historians, archaeologists, and chemists in the study of a single common topic, is the study, already mentioned, of the Corinthian black bronze.[45] Indeed, the discovery of Syriac recipes attributed to Zosimus

[44] See, for instance, M. Beretta, *Counterfeit, Imitation and Transmutation in Ancient Glassmaking* (Sagamore Beach, MA: Science History Publications, 2009) for the history of glass technique, where he shows that the Egyptian art of glass production strongly influenced the establishment of alchemical theory.

[45] See R. Giumlia-Mair and P. T. Craddock, *Das schwarze Gold der Alchimisten: Corinthium aes* (Mainz: Philipp von Zabern, 1993); Hunter, "Beautiful Black Bronzes"; A. Giumlia-Mair, "Zosimos the Alchemist: Manuscript 6.29, Cambridge, Metallurgical Interpretation," in Giumlia-Mair (ed.), *I bronzi antichi*, pp. 317–23.

for the fabrication of this mythical alloy has attracted the attention of archaeologists and chemists, who have long wondered whether there is a link between the references to the alloy in classical authors and some objects preserved in museums that present a surprising black patina (Figure 24.4). Laboratory analyses have allowed researchers to make a start in retracing the history of the technique, which involved enriching a copper alloy with a small amount of gold and/or silver, thus leading to the formation, through a surface treatment by chemical reaction, of an artificial black patina that is particularly shiny and fit to make beautiful decorations stand out.[46] The Syriac recipes attributed to Zosimus are the only ancient recipes relating to this technique that have been preserved and their reproduction could provide the key to this process, as long as there is a close collaboration with philologists in the decipherment of the texts.

In the present state of studies in this field, it thus appears fundamental to pursue a multidisciplinary approach and to recognize the systematic and positive side of alchemy. This is legitimate, since ancient authors themselves often distinguish natural and rational research from deceptive practices subject to chance and the wishes of demons. Only such a methodological perspective will create a rigorous model of what alchemy was, in contrast to the currently prevalent abusive and partial interpretations of it as a crazed and quackish practice stemming from an esoteric hermeticism.[47]

[46] See F. Mathis, *Croissance et propriétés des couches d'oxydations et des patines à la surface d'alliages cuivreux d'intérêt archéologique ou artistique* (Sarrebruck: Éditions Universitaires Européennes, 2011).

[47] On this topic, see W. R. Newman and L. Principe, "Some Problems with the Historiography of Alchemy," in W. R. Newman and A. Grafton (eds.), *Secrets of Nature: Astrology and Alchemy in Early Modern Europe* (Cambridge, MA: Massachusetts Institute of Technology (MIT) Press, 2006), pp. 385–432.

Part IV

INDIA

25

ASTRONOMY AND ASTROLOGY IN INDIA

Kim Plofker

Astronomical computation and prediction in pre-modern South Asia (an area roughly comprising modern India, Pakistan, Nepal, Bangladesh, and Sri Lanka), like their counterparts in the classical world and its European inheritors, have left detailed though incomplete imprints in the historical record since the first millennium BCE. Founded in traditional cosmological and celestial lore and the requirements of maintaining a ritual calendar, the Indian "science of the stars" (Sanskrit *jyotiḥśāstra*) subsequently altered with the acquisition of various new scientific and social aims, as well as mathematical techniques and models. Its primary goal became the ability to predict, at any known time and terrestrial locality, the places of the heavenly bodies, partly for calendric purposes and partly to satisfy the needs of numerous systems for forecasting the future that may be lumped together as "astrology." The chief canonical texts of this tradition were written in Sanskrit, the dominant learned language of classical India, mostly though not exclusively by priestly and scholarly functionaries (Brāhmaṇas among the Hindus, members of monastic orders among Buddhists and Jains) whose responsibilities included the abovementioned timekeeping and forecasting practices, as well as the preservation of learning in general. But it was the ephemeral texts, the *pañcāṅgas* (calendars) and horoscopes produced in dozens of vernaculars as well as Sanskrit by local astrologers, that spread it throughout medieval South Asia and neighboring cultures.[1]

[1] Additional material on all these aspects can be found in, e.g., S. N. Sen with the research assistance of A. K. Bag and S. Rajeswar Sarma, *A Bibliography of Sanskrit Works in Astronomy and Mathematics* (Delhi: Indian National Science Academy, 1966); R. C. Gupta, "A Bibliography of Selected Sanskrit and Allied Works on Indian Mathematics and Mathematical Astronomy," *Gaṅita Bhāratī* 3 (1981), 86–102; Michio Yano, "Calendar, Astronomy, and Astrology," in Gavin Flood (ed.), *Blackwell Companion to Hinduism* (Oxford: Blackwell Publishing Ltd, 2003), pp. 376–92; and several of the surveys by David Pingree cited in the following notes.

PREHISTORY: THE INDUS VALLEY CIVILIZATION

Nothing definite is known about the practice of astral sciences in third/second-millennium BCE Indus Valley/Harappan societies. We can plausibly assume that these cultures' urban organization and wide-ranging trade required maintenance of a calendar that in turn demanded some astronomical knowledge. Archaeological evidence and texts from later periods have been adduced in support of various hypotheses attempting to reconstruct such knowledge; however, there is no known contemporary textual record that can conclusively confirm or correct these hypotheses (unless perhaps future scholarship should arrive at a definitive interpretation of the controversial "Indus script" symbols).[2]

ASTRONOMICAL IDEAS ATTESTED IN THE VEDIC PERIOD

The astronomical and cosmological concepts that survive from the period of the earliest known Sanskrit texts are preserved in the Vedic hymns and rituals, in their prose expositions or Brāhmaṇas, and to some extent in the Purāṇas, later cosmogonic texts in verse. Explicitly astronomical knowledge in the Vedic texts themselves is confined to scattered references concerning a few important topics: some star names, including in the later texts lists of the twenty-seven or twenty-eight *nakṣatras*, the constellations in the path of the Moon; the Moon and its phases (particularly new and full moon) and the synodic month as a unit of time (reckoned from its new or full phase); the Sun and its daily motion (from east to west with its bright side facing the earth during the day, and from west to east at night with its dark side showing) as well as its yearly motion (from south to north as winter changes to summer, and then from north to south); the causation of eclipses by demons; the ideal year of twelve months of thirty days each, with occasional intercalation of a thirteenth to maintain the seasonal characteristics (indicated by month names like "hot" and "cloudy") of the actual months (though the later Vedic texts also adopt names derived from the names of the *nakṣatras* occupied by each successive full moon); and the division of

[2] An overview of numerous studies and theories concerning Indus Valley cultures is given in Robert H. Dyson, Jr., "Paradigm Changes in the Study of the Indus Civilization," in Gregory L. Possehl (ed.), *Harappan Civilization: A Contemporary Perspective* (New Delhi: Oxford & IBH Publishing Co., 1982), pp. 417–27. Some hypotheses about Harappan astronomy and possible archaeoastronomical clues to reconstructing it are described in Mayank N. Vahia and Srikumar M. Menon, "Theoretical Framework of Harappan Astronomy," in Tskuo Nakamura (ed.), *Mapping the Oriental Sky: Proceedings of the Seventh International Conference on Oriental Astronomy* (Tokyo: National Astronomical Observatory of Japan, 2011), pp. 27–36, and several of the sources cited therein. Many salient points of the debates about decipherment of Indus script can be traced in the arguments and references in Richard Sproat, "A Statistical Comparison of Written Language and Nonlinguistic Symbol Systems," *Language* 80 (2014), 457–81.

the day and night each into fifteen units (*muhūrtas*). The occupational category of "star-observer" (*nakṣatradarśa*) is mentioned but not described.[3] Parallels to the concepts of the ideal year, the *nakṣatra*-lists, the yearly movement of the Sun's rising-point, and luni-solar intercalation are found in a Mesopotamian text from the end of the second millennium BCE; it is uncertain whether these similarities reflect parallel evolution of separate astronomical traditions or transmission between them.[4] Although the Vedic world recognizes divine agency in the good and ill fortune of humans, who can incite or appease these agents with prayers and curses, the idea of systematically using celestial events to predict such fortune appears to be a later development.

A more detailed physical (though not predictively mathematical) model of the cosmos and accounts of its chronology are revealed in the Purāṇas, a corpus of sacred texts compiled in their present form many centuries later than the Vedas, but preserving a tradition of sacred cosmology largely distinct from the spherical geometric models of later Indian astronomy.[5] The Purāṇic world or "egg of Brahmā" contains the vertically stacked layers of our universe, one of which is the disk of the earth. The earth itself is composed of a central circular continent surrounded by a circular ocean of salt water, which in turn is surrounded by the concentric rings of six other continents, separated from each other by other oceans. The earth's center is occupied by Mount Meru, on whose summit is the city of the gods; the Indian region lies in the land south of Meru. The flat disk of the earth extends 500,000,000 *yojanas* in diameter (the size of this traditional unit of length cannot be established precisely: a *yojana* is greater than one mile and probably less than ten).

The sun moves above the earth and parallel to its surface, the moon above the sun, and the *nakṣatras* higher still. Continuing outwards, the order of other major bodies is Mercury, Venus, Mars, Jupiter, Saturn (a sequence appearing in Greek sources near the beginning of this era[6]), the seven stars ("Seven Seers") of the Big Dipper, and the pole star. All the celestial bodies revolve about an axis terminating in the pole star, to which they are attached

[3] A. A. Macdonell and A. B. Keith, *Vedic Index of Names and Subjects*, 2 vols. (London: John Murray & Co., 1912; reprinted Delhi: Motilal Banarsidass, 1982). Cf. "Ahan" (1: 48–50), "Graha" (1: 243–4), "Candra" (1: 254), "Nakṣatra" (1: 409–31), "Māsa" (2: 156–63), "Muhūrta" (2: 169), "Sūrya" (2: 465–8).

[4] Various arguments for Mesopotamian influence on late Vedic astronomy are put forth in David Pingree, "MUL.APIN and Vedic Astronomy," in Hermann Behrens (ed.), *DUMU-E2-DUB-BA-A: Studies in Honor of Åke W. Sjöberg* (Philadelphia, PA: University Museum, 1989), pp. 439–45, and critiqued in Yukio Ohashi, "Development of Astronomical Observation in Vedic and Post-Vedic India," *Indian Journal of History of Science* 28 (1993), 185–251.

[5] The following summary is derived primarily from Books I and II of the Viṣṇupurāṇa; cf. *The Viṣṇu Purāṇa*, ed. and trans. H. H. Wilson, 2 vols. (Delhi: Nag Publishers, 1980; repr. 1989), vol. 1, 32–7, 249–349.

[6] O. Neugebauer, *History of Ancient Mathematical Astronomy*, 3 vols. (Berlin: Springer-Verlag 1975), vol. 2, 690–1.

by cords of wind, and appear to set as they pass behind the bulk of Mount Meru. The moon holds the divine nectar which is drunk by the gods and replenished by the sun every month; occasionally the celestial being Rāhu devours it all at once, producing a lunar eclipse. Above the region of the pole star and below the earth are the dwelling-places of the divine and demonic beings, respectively.

The existence of the earthly realms (extending as far as the region of the pole star) is terminated by universal destruction at the end of the period called a *kalpa*, or a single day of Brahmā, its creator, which lasts 4,320,000,000 earth years. A *kalpa* contains a thousand cycles of the period called a *mahāyuga* ("great age"), which in turn is composed of four unequal and progressively deteriorating ages: the Kṛtayuga (a sort of Golden Age comprising 1,728,000 years), the Tretāyuga (1,296,000 years), the Dvāparayuga (864,000), and the Kaliyuga (432,000). (The most recent Dvāparayuga is considered to have ended at the time of the great battle described in the epic *Mahābhārata*, and humanity is presently enduring the Kaliyuga.)

In summary, the available direct textual evidence suggests that before about the middle of the first millennium BCE, practical astral science in India focused on the use of various luni-solar calendric schemes to regulate the performance of monthly and seasonal sacrifices. (In fact, astronomy is classified in Brāhmaṇa texts as one of the six Vedāṇgas, "limbs of the Vedas," the purpose of whose existence is to support the Vedic ritual.) The basic units of the calendar were the solar year and the synodic month; a normal year had twelve months with intercalation of an extra one when necessary to keep the seasons and months synchronized. Except for the requirements imposed by the calendar, the traditional cosmology defined the dimensions and organization of the universe generally in accordance with mythological imagery rather than with observationally constrained models. Celestial objects besides the stars and luminaries eventually gained a place in this cosmos (although perhaps not until quite late in some cases), but they did not yet form part of complex schemes for forecasting the future.[7]

POST-VEDIC SYSTEMS OF MATHEMATICAL ASTRONOMY AND ASTROLOGY

Starting in the last half of the first millennium BCE more complex techniques for keeping track of time and the calendar begin to be documented,

[7] An alternative assessment of Vedic astronomy, based on interpretation of the textual references as much more observationally precise and theoretically sophisticated than they are here claimed to be, infers a much more ancient and complex astronomical tradition for Vedic India. See, e.g., B. N. Narahari Achar, "On the Vedic Origin of the Ancient Mathematical Astronomy of India," *Journal of Studies on Ancient India* 1 (1998), 95–108.

supplemented by ones for predicting ominous celestial phenomena. As was standard for didactic texts in Classical Sanskrit *śāstras* or learned disciplines, the astronomical and astrological works expounding these techniques were generally composed in Sanskrit verse, most often furnished (by the author or by later students of the text) with one or more commentaries in Sanskrit prose.

The seminal work *Jyotiṣavedāṅga* ("astronomical limb of the Vedas"), most likely compiled in the fifth or fourth century BCE, and similar texts from the succeeding centuries give rules for using the following instruments and methods: an out-flowing water-clock for measuring the length of day or night, with a linear zigzag function to determine the difference in the amount of water to be used depending on the time of year; a ratio for the length of the longest day of the year to that of the shortest (given as 3:2, which is ideally appropriate for a latitude of about 35°, with a linear zigzag function to give the constant difference per month in length of daylight); a vertical gnomon whose shadow could be measured for timekeeping, and a zigzag function giving the constant monthly difference in the length of the shadow at noon (which at the summer solstice is assumed in one text to be zero, which is appropriate for a latitude of about 24°); the unit of the *tithi* or "lunar day," one-thirtieth of a synodic month; a cycle of luni-solar intercalation, assuming that 5 solar years equal 62 synodic months or 1830 days; the use of the *nakṣatra* as a regular unit of arc, equal to 1/27 of a circle; and values for the mean motions of the sun and moon in *nakṣatras* per day. Some of these features (regular intercalation cycles, linear zigzag functions, the 3:2 daylight-length ratio, the thirtieth of a synodic month, and the gnomon and water-clock) are known from Babylonian texts, while others (the intercalation period of five years, the *nakṣatra* as a unit of arc, a noon solstitial gnomon shadow of zero) are not attested outside India. The development of post-Vedic mathematical astronomy thus may have involved an Achaemenid-era adaptation of Mesopotamian techniques and concepts to harmonize with some existing Indian ones, although no explicit attestation of such borrowing is known.[8]

The motions of the five star-planets emerge from obscurity in Indian astral science at about the same period, in systems of divination that include celestial omens. A fourth-century collection of sermons of the Buddha contains a list of certain forms of professional divination (repudiated by the Buddha as unethical) including various ominous meteorological and astronomical events: among the latter are lunar and solar eclipses and the risings and settings of the sun, moon, and stars (apparently meaning also the

[8] Overviews of some of the different interpretations of these developments include David Pingree, "The Mesopotamian Origin of Early Indian Mathematical Astronomy," *Journal for the History of Astronomy* 4 (1973), 1–12, and Yukio Ohashi, "Astronomy of the Vedic Age," in Clive L. N. Ruggles (ed.), *Handbook of Archaeoastronomy and Ethnoastronomy*, 3 vols. (New York: Springer, 2015), vol. 3, 1949–58.

star-planets). A text compiled in about the first century CE, the *Gargasaṃhitā* ("[omen] collection of Garga"), goes into much more detail about the effects on earthly fortunes of the planets' locations with respect to the stars of the *nakṣatras* and to one another. It also gives various quantitative estimates of the periods between the planets' risings, settings, stations, and so forth, reminiscent of earlier Babylonian planetary theory.[9]

In the first centuries of the Common Era, these sciences were greatly expanded by the introduction (via exchanges with the Roman Empire, most likely mediated by the Kṣatrapa rulers in western India) of Greco-Babylonian and Greek astronomical ideas predating or ignoring the system of Ptolemy, and by their development within the Sanskrit tradition. In texts of this period, the technical aims are enlarged beyond the earlier focus on mathematical schemes for keeping track of the divisions of the day and year.[10] Now the astronomer also needs to compute such quantities as the following: the mean positions, according to a system of spherical celestial coordinates with specified epoch positions, of the sun and moon corresponding to a given date; the true positions resulting from trigonometric correction of the mean positions; the time of first lunar visibility and the position and orientation of the crescent; and the times, locations, and appearances of solar and lunar eclipses. (In calculating such quantities as visibility and appearance, the local longitude and latitude on a spherical earth must be taken into account.) Thus instead of the earliest reference circle, that of *nakṣatras* in the moon's path, Sanskrit astronomical texts now refer to the ecliptic (on which the *nakṣatra* is a constant unit of arc, and which also contains the Greek zodiacal signs and degrees), and standard quantities such as longitude and latitude determine positions with respect to it and to the circle of the celestial equator. Time divisions now include, besides the *muhūrta*, sixtieths and 360ths of a nychthemeron. Plane trigonometry utilizing sines (developed from the Greek chord function) is used in projections of the celestial sphere to find the required arcs. Parameters for the planets' anomalies, in some cases involving geometric models that employ concentrics, eccentrics, and epicycles, are used to correct their mean motions.

Many of the desired astronomical quantities (e.g. position and orientation of eclipses and the lunar crescent; locations of planets at their transitions

[9] David Pingree, "Mesopotamian Omens in Sanskrit," in D. Charpin and F. Joannès (eds.), *La circulation des biens, des personnes et des idées dans le Proche-Orient ancien,* XXXVIII^{e} *R.A.I.* (Paris: Editions Recherche sur les Civilisations, 1992), pp. 375–9, and David Pingree, "Babylonian Planetary Theories in Sanskrit Omen Texts," in J. L. Berggren and Bernard R. Goldstein (eds.), *From Ancient Omens to Statistical Mechanics* (Copenhagen: University Library, 1987), pp. 91–9.

[10] The following summary depends upon a sixth-century compendium of most of the texts that survive from this time, of which the two chief editions and translations are *The Pañcasiddhāntikā of Varāhamihira*, ed. and trans. O. Neugebauer and D. Pingree, 2 vols. (Copenhagen: Kongelige Danske Videnskabernes Selskab, 1970), and *The Pañcasiddhāntikā of Varāhamihira*, ed. and trans. T. S. Kuppanna Sastry and K. V. Sarma (Madras: P.P.S.T. Foundation, 1993).

between visibility and invisibility) were sought on account of their ominous significance. Translations of Hellenistic Greek texts on nativity astrology, itself originally inspired by Mesopotamian natal omens, fundamentally shaped Indian mathematical astrology in this period and lent it many of its core technical terms.[11] The precise course of all these developments during the first half-millennium of this era is difficult to trace, since only a few texts representing this period survived the dominance of the "classical" form of this science during the next thousand years or so. But the end result was a powerful, flexible astronomical/astrological system incorporating a wide range of problem-solving techniques.

THE DEVELOPMENT OF THE *SIDDHĀNTA*

The fundamental type of text in classical Indian astronomy from about 500 CE onwards is the *siddhānta* ("astronomical system"), a treatise on the major problems and methods of astronomical computation. A *siddhānta* is distinguished from other astronomical texts not only by its more complete exposition but by its lack of a recent epoch: it reckons time from either the beginning of the present *kalpa* or the beginning of the present *yuga*. Thus a *siddhānta* ideally permits an astronomer to solve any recognized astronomical problem at any point during the lifetime of the universe. The work's presumption of timelessness is often reinforced by the omission of any datable references and/or by its ascription to a divine author, which can make its historical origin difficult to identify precisely.

The cosmology of the *siddhānta* preserves some features of the Purāṇic universe described above in the section "Astronomical Ideas Attested in the Vedic Period", modified by the exigencies of the geometrical models required by the later mathematical techniques. The universe described by these texts is spherical and centered upon a stationary, spherical earth (with Mount Meru at its north pole), about which it turns from east to west once a day. (The alternative hypothesis of a rotating earth is known, proposed most notably by Āryabhaṭa in the late fifth century, but almost unanimously rejected by other authors on the basis of intuitive physical arguments.)[12] Its

[11] One of the chief witnesses of this transmission is edited and translated in *The Yavanajātaka of Sphujidhvaja*, ed. and trans. David Pingree 2 vols. (Cambridge, MA: Harvard University Press, 1978); a revised interpretation of some of its elements arguing for a somewhat earlier amalgamation of Greek and Indian astronomy is offered in Bill Mak, "The Date and Nature of Sphujidhvaja's *Yavanajātaka* Reconsidered in the Light of Some Newly Discovered Materials," *History of Science in South Asia* 1 (2013), 1–20. It is suggested in Johannes Bronkhorst, *Buddhism in the Shadow of Brahmanism* (Leiden: Brill, 2011), pp. 30–2, that astrological prediction was one of the professional tools contributing to the resurgence of *Brāhmaṇas* and their religious traditions in a post-Vedic culture heavily influenced by Buddhism and Jainism.

[12] *Āryabhaṭīya of Āryabhaṭa*, ed. and trans. K. S. Shukla and K. V. Sarma (New Delhi: Indian National Science Academy, 1976), pp. 13–15.

present existence will endure for the period of one *kalpa*, of which somewhat less than half had passed at the beginning of historical time; the *Kaliyuga* of the current *Mahāyuga* began in 3102 BCE. The seven planets (Moon, Mercury, Venus, Sun, Mars, Jupiter, Saturn) also revolve concentrically about the earth – moving (usually) eastward among the fixed stars as well as sharing in their daily westward motion – but they are displaced from their uniformly moving mean positions by one anomaly (*manda* or "slow") which more or less corresponds to the ellipticity of the earth's or planet's orbit in modern theory, and by another (for the five star-planets only, termed *śīghra* or "fast") which accounts for the relative motion of the earth and the planet. The so-called apexes (corresponding to apogees in eccentric orbits) associated with these anomalies, the planets themselves, the nodes of their orbits, and the sphere of the fixed stars all complete integer numbers of revolutions in a *kalpa*, beginning at zero latitude upon the prime meridian at sunrise of the first day of creation. The orbital distances of the planets are assumed to be inversely proportional to their observed mean orbital speeds; since the mean motions of Mercury and Venus are identical to those of the Sun, their faster *śīghra*-apexes are used instead to place their orbits between those of the Moon and the Sun. The abovementioned displacements from the mean planetary positions are physically modeled by the assumption of a divine being standing at the position of each apex and pulling its assigned planet towards itself with a cord of wind.[13] (Mathematically equivalent models placing the true planet actually upon an epicycle or eccentric are also referred to.[14] Such merely mechanical divergences between models are unimportant in these texts, which focus chiefly upon mathematical prediction rather than physical dynamics.) When the astronomer can thus compute the corrected positions and motions of the planets, the problems of astrology and the calendar can be solved by determining these positions with respect to terrestrial locations, to other planets, and to various fixed stars.

THE STRUCTURE OF THE CLASSICAL *SIDDHĀNTA*

The organization of this text follows a general pattern that evolved around the middle of the first millennium: it consists of versified parameters and formulas for calculating various celestial phenomena for a known date and time. The example summarized below as roughly canonical is that of the first ten chapters of the *Brāhmasphuṭasiddhānta* of Brahmagupta (628).[15]

[13] This is described in the *Sūryasiddhānta* of about the eighth century: *Sūryasiddhānta*, ed. K. C. Dvived (Varanasi: Sampurnanand Sanskrit University, 1987), pp. 31–2.

[14] E.g., in the section on spherics of a *siddhānta* roughly contemporary with the *Sūryasiddhānta*: *The śisyadhvrddhidatantra of Lalla*, ed. and trans. Bina Chatterjee, 2 vols. (New Delhi: Indian National Science Academy, 1981), vol. 1, 198–200.

[15] *Brāhmasphuṭasiddhānta*, ed. S. Dvivedi (Benares: Medical Hall Press, 1901/1902).

Mean motions. The astronomer's first task is to find the mean longitudes of the planets at the given time at his own locality. In a *siddhānta*, this is done by computing the revolutions completed by a celestial body – planet, apogee, node, or fixed star – during the civil days (which are considered to start at the terrestrial prime meridian) from the beginning of the *kalpa* to the present moment. The fractional excess of this quantity over integer revolutions is its mean longitude. Thus the work commences by providing the number of complete sidereal revolutions of each celestial body in a *kalpa*; the ratio of elapsed years to total years of a kalpa is applied to this number, to give the body's elapsed revolutions. The calculation of the longitudinal distance (and hence the mean time-difference) between one's own geographical location and the prime meridian is also described in this first chapter. (This result depends of course upon the size of the earth, whose circumference is here said to be 5,000 *yojanas*.) Then the planet's computed mean longitude, plus or minus the arc corresponding to the time-difference due to geographical longitude, is the planet's mean longitude for the given time and place.

True motions. The true planets are considered to be physically displaced from their mean longitudes in the direction of the apex of the relevant anomaly, as described above. Simple trigonometry is required to compute this angular correction, so the second chapter provides a versified table of sines and rules for interpolation within the table, as well as the dimensions of the epicycle(s) centered on the mean planet, and the formulas for computing the correction. Since the planet's motion need not be geometrically uniform, the rules for applying the dual correction for a planet with two inequalities vary from text to text, in ways whose physical significance is not always apparent. Another physically obscure feature of this procedure is a variation (dependent on the amount of anomaly or on altitude) in the size of one or both epicycles, which is described (though not explained) in this chapter.

"Three Questions." Once the true positions of the planets relative to the center of the earth are thus known, the astronomer must be able to compute the direction, place, and time – the 'Three Questions" of the title of the third chapter – in which they will appear, relative to terrestrial position. Most of this chapter is devoted to the various trigonometric formulas for the quantities used in solving these problems (equation of time, equation of daylight, terrestrial latitude, solar amplitude, ascendant, etc.), which may depend on measurement of, e.g., a gnomon shadow or the solar altitude

Lunar eclipse. As Brahmagupta says at the beginning of his fourth chapter, "The knowledge of time is sought by good [astronomers] primarily for the purpose of understanding the syzygies." Indeed, just as the conjunction and opposition of sun and moon are the chief ordinary events of Indian astronomy and the calendar, the solar and lunar eclipses caused when syzygies occur near a node of the moon's orbit are the chief catastrophic events of the celestial cycles, causing various kinds of destruction and benefit depending upon the characteristics of the eclipse. To predict these qualities for a coming eclipse (or to reconstruct them for a past one), the longitudes of sun, moon, and node are calculated for the time of syzygy, as are the moon's latitude at that longitude and the apparent diameters of the luminaries and the earth's shadow (considered proportional to the luminaries' current velocities). From these, the amount of obscuration, duration of eclipse, duration of totality, and direction of impact of the eclipse are calculated. Another ominous characteristic, the color of the eclipsed lunar orb, is said to depend on the amount of obscuration.

Solar eclipse. The calculation of solar eclipses requires quantities similar to those in lunar eclipses, with the additional necessity (since parallax determines the visibility of a solar eclipse) of finding the parallax components due to the size of the earth: longitudinal (which advances or delays the time at which the conjunction appears to occur) and latitudinal (which changes the extent of the apparent obscuration). The fifth chapter, therefore, gives rules for calculating these parallaxes, and for applying their effects to the relevant quantities.

Planetary visibility. The computation of the times when a planet becomes visible or invisible, due to its decreasing or increasing proximity to the sun, is the subject of the sixth chapter. The locations of the rising and setting points on the ecliptic are computed by a geometrical approximation, and compared to the ascendant and to the arc of visibility (minimum elongation from the sun, in time-degrees) for the desired planet. The importance of this computation, as of those in the remaining chapters, is primarily owing to its astrological significance.

Orientation and projection of the moon's crescent. The seventh chapter treats the relative positions of the moon and the sun, and the resulting size and orientation of the moon's illuminated part, which is to be displayed graphically. Calculations are also discussed for arranging sighting-tubes for observation of the moon by the public or a patron.

Lunar shadow. The calculation of the shadow cast by the moon is described; it is essentially equivalent to the corresponding procedures for the shadow of the sun in the third chapter on direction, place, and time, and serves a similar function in timekeeping.

Planetary conjunctions. The definition of "conjunction" in the case of planets is a matter of importance in the ninth chapter; simply computing the time of equality of the planets' longitudes is apparently not enough to identify a conjunction, which requires the planets to be on the same secondary to the ecliptic. As in Western astrology, conjunction is considered to have a profound effect on the influences of the celestial bodies, so it is crucial to identify the event correctly.

Conjunctions of planets and stars. The tenth chapter identifies the positions of some fixed stars (generally the junction-stars of the *naksatra* constellations) in polar coordinates[16] and gives a rule for the computation of their declinations. The conjunction of a planet with a star is computed as for that of two planets.

The arrangement of the topics described above, while fairly representative, is by no means uniform or complete. Often the given topics are presented in a different arrangement (although the order of the first five chapters is more or less standard); often other subjects are included among them, and alternative formulas provided; and often an additional section, with additional chapters on a variety of subjects, is appended. Some of the most frequently encountered supplementary topics are the following.[17]

Instruments. The text may describe the construction and use of various instruments for time-keeping, such as a gnomon or different forms of water-clocks or sundials; for observation, such as a perforated sighting-ring; or for theoretical study, such as an armillary sphere.[18]

Computation of *pātas*. The ominous events known as *pātas*, which depend on the relationship between the declinations of the two luminaries, are predicted in a fashion similar to that for the conjunction of planets.

The sphere. A summary of the technical computations for the celestial and/or terrestrial sphere, often including questions of geography and cosmology, may be included as a separate chapter.

Objections to opposing theories. An author may devote a chapter to contradicting or refuting assertions that differ from his own, whether they originate with other astronomers or in traditional Purāṇic cosmology. Physical considerations, irrelevant to the computational procedures that constitute most of the treatise, are sometimes discussed here, e.g., in demonstrations that the earth does not rotate or that the moon is below the sun.

Problems. A selection of sample problems may be provided to test the reader's knowledge of the procedures explained in the text.

[16] Other texts use ecliptic coordinates or a mixture of the two systems: see David Pingree and Patrick Morrissey, "On the Identification of the *Yogatārās* of the Indian *Nakṣatras*," *Journal for the History of Astronomy* 20 (1989), 99–119.

[17] Examples of such topics can be found in the work of Lalla and the *Sūryasiddhānta* mentioned above, and also in, among others, the tenth-century *Vaṭeśvarasiddhānta* (*Vaṭeśvara-Siddhānta and Gola of Vaṭeśvara*, ed. and trans. K. S. Shukla, 2 vols. (New Delhi: Indian National Science Academy, 1985)), and the twelfth-century *Siddhāntaśiromaṇi* (*The Siddhāṇta śiromani*, ed. Ganapati Deva śāstrī (Varanasi: Chaukhambha Sanskrit Sansthan, 1989)).

[18] Many such instruments are comprehensively discussed in Sreeramula Rajeswara Sarma, *The Archaic and the Exotic: Studies in the History of Indian Astronomical Instruments* (New Delhi: Manohar, 2008).

THE DEVELOPMENT AND INTERRELATIONSHIPS OF RIVAL SCHOOLS

Although the goals and methods of classical Indian astronomy do not diverge very far from the basic model described above, the texts – including almost all forms of astronomical works, and the commentaries on them – are usually clearly classified by their adherence to one or another of the major astronomical schools, or *pakṣas*. These schools are each based on a particular text considered as the foundation of their tradition, and differ from one another primarily in the details of their parameters. Astronomers writing in the tradition of one *pakṣa* may severely criticize adherents of another – usually on the grounds that the rival's theory is inconsistent with sacred texts, or else produces results that do not agree with observation – but many astronomers also write texts in more than one *pakṣa*. Moreover, systems of additive corrections (*bījas*, "seeds") are frequently devised for transforming the results derived with one *pakṣa*'s parameters to the corresponding values for a different *pakṣa*.[19]

The Brāhmapakṣa. The canonical text of this school is the seventh-century *Brāhmasphuṭasiddhānta* ("Corrected astronomical system of Brahmā") outlined above. This *pakṣa*, which flourished especially in the west and northwest regions of India, accepts the definitions of the subdivisions of time, beginning with the *kalpa*, provided in the Purāṇas and in early texts on dharma; it assumes a true conjunction of all the heavenly bodies in longitude and latitude at the commencement of a *kalpa*, and integer numbers of rotations for each of them during that period. The number of civil days per *kalpa* in the Brāhmapakṣa (i.e., the difference between the number of rotations of the fixed stars and those of the sun) implies a year-length of 365;15,30,22,30 days. A sunrise epoch is used; the date of the beginning of the present Kaliyuga is considered to be sunrise (at zero terrestrial longitude and latitude) with the sun at the initial point of the zodiac, 3179 years prior to the beginning of the Śaka era (in 78 CE): i.e., on 18 February 3102 BCE.

The Āryapakṣa. This *pakṣa* was initiated by the *Āryabhaṭīya* of Āryabhata (who claimed to have restored by it the "astronomy of Brahmā," possibly an earlier forerunner of the Brāhmapakṣa?) at the end of the fifth century, 3600 years after the start of the current Kaliyuga. It considers the *kalpa* to contain 1008 *mahāyugas* of 4,320,000 years, whose four component ages are all equal in size (although this latter modification was discarded by some later followers of this school, which was most popular in the southern part of the subcontinent). The planets participate in a mean conjunction (i.e., one not involving most of their apices and nodes) at the

[19] The *paksas* are more fully described in David Pingree, "History of Methematical Astronomy in India," in *Dictionary of Scientific Biography, vol. 15* (Detroit, MI: Charles Scribner's Sons, 1978), 533–633; see also David Pingree, "Bīja-Corrections in Indian Astronomy," *Journal for the History of Astronomy* 27 (1996), 161–72.

end of each age, and thus complete an integer number of rotations during this period. The length of the year is taken to be 365;15,31,15 civil days.

The Ārdharātrikapakṣa. The Ārdharātrika (literally "midnight") *pakṣa* is contemporary with the Āryapaksa, and in fact is based on a lost work of the same author. As its name implies, it takes the beginning of the Kaliyuga to have occurred not at sunrise but at the previous midnight.

The Saurapakṣa. The name of this school is derived from Sūrya, "sun," and it follows the *Sūryasiddhānta* said to have been revealed by the Sun-god, whose most complete surviving version dates from about the eighth century; its influence is best attested in the northern and eastern regions. The Saurapaksa uses the time divisions of the Brāhmapaksa but, like the other two schools, seeks to have a mean conjunction of all planets at the beginning of the Kaliyuga: to make the numbers work out, it assumes that the first 17,064,000 years of the *kalpa* were a period of quiescent creation in which no planetary movement occurred. The epoch time (following the Ārdharātrikapakṣa) is midnight, and the year-length is 365;15,31,31,24 civil days.

The Gaṇeśapakṣa. Based on the influential *karaṇa* or handbook (see next section) called *Grahalāghava* which was composed by Gaṇeśa in the early sixteenth century, this school uses a mix of planetary parameters from the other *paksas*, slightly modified in some cases, to achieve (according to a commentator on Gaṇeśa) the best possible agreement with observed planetary positions at the time of the *Grahalāghava*'s epoch. Its sphere of influence was primarily in Gujarat and the north.

OTHER TYPES OF TEXTS, ASTRONOMICAL AND ASTROLOGICAL

Astronomical handbooks (*karaṇa*). Since it would be highly cumbersome to use the procedures given in a *siddhānta* for every astronomical calculation, there developed side by side with the *siddhānta* literature a class of practical texts called *karaṇas*. A *karaṇa* enables the user to find the same general quantities explained in a *siddhānta*, but in a simplified fashion: the planetary positions are provided for a recent epoch (often the *karaṇa*'s date of composition), and more concise, often approximate formulas (sometimes extremely ingenious) are provided for doing the calculations.

Tables (*koṣṭhaka* or *sāraṇī*). The use in second-millennium Sanskrit astronomy of detailed tables of pre-calculated astronomical function values, rather than the concise algorithms prescribed for computing such values in treatises and handbooks, seems to have been inspired by versified lists of function values in earlier texts and by the example of the Islamic *zīj*. These *koṣṭhakas* are generally intended for the use of calendar-makers, and ideally contain instructions for computing the calendar information from the data in the tables.

Calendars (*pañcāṇga*). The *pañcāṅga*, ephemeral though it is, is in some ways the most important text in Indian astronomy, as it provides the ultimate practical motivation for basic astronomical methods, identifying the moments that determine years, months, days, and their various subdivisions, as well as the stated times for performance of ritual observances, and those that are auspicious and inauspicious.

Other. Sometimes an instructional text is devoted specifically to an auxiliary topic, such as astronomical instruments, geography, or the sphere.[20]

Omens (*saṃhitā*). Celestial omens, as discussed above, seem to be the oldest form of astral prophecy in India; they identify and classify fateful events rather than mathematically predicting them. The genre of *saṃhitā* (literally "collection") also comprises many types of meteorological and other non-astral omens; the purely astronomical ominous phenomena include eclipses, the position of the horns of the lunar crescent, and conjunctions of planets with other planets and with stars. The various effects of such events upon the political, economic, and personal fates of occupants of different regions and social divisions are what the *saṃhitā* texts describe.[21]

Genethlialogy (*horā, jātaka*). This branch of astrology (from *jan*, "to be born") analyzes the influences of an individual's horoscope – the positions of the celestial bodies at the time of birth – upon his or her fate. Simple effects of the sort described in omen literature are heavily complicated here by modifications to the planetary influences depending on their relation to various subdivisions of time (such as the *horā*, "hour") and the zodiac, and to one another's positions. Its Indian developments in turn influenced Western natal astrology.[22]

Catarchic astrology (*muhūrta*) and interrogational astrology (*praśna*). The concept of a horoscope, an assessment of combined planetary influences for a specific moment in time, also inspired the other two major branches of Indian astrology. Texts on *muhūrta*, which takes its name from the time-unit mentioned above and which is ultimately based on Hellenistic catarchic astrology, identify a propitious time for beginning a specified action. The practice of *praśna* (literally "question") reverses this approach to predict a future event based on the horoscope of the time at which the prediction was sought. Two astrological subspecialties, *yātrā* ("military expedition") and *vivāha* ("marriage") subsequently developed

[20] For a survey of these different astronomical genres, see David Pingree, *Jyotiḥśāstra: Astral and Mathematical Literature* (Wiesbaden: Otto Harrassowitz, 1981), pp. 32–55.

[21] The best-known work in this genre is the sixth-century *Bṛhatsaṃhitā* of Varāhamihira (*Varahamihira's Brihat Samhita*, ed. and trans. V. S. Sastri and M. R. Bhat, 2 vols. (Bangalore: V. B. Soobbiah & Sons, 1947)).

[22] David Pingree, *From Astral Omens to Astrology, From Babylon to Bīkāner* (Rome: Istituto italiano per l'Africa e l'Oriente, 1997), pp. 31–8.

from the application of modified forms of *muhūrta* to those particular undertakings.[23]

SOUTH INDIAN ASTRONOMY: THE *VĀKYA* SYSTEM AND THE KERALA SCHOOL

A unique astronomical tradition arising in medieval South India (where, however, standard *siddhānta* astronomy, particularly that of the Āryapakṣa, was also well known) takes a different approach to computing planetary longitudes. In this system, planetary period relations (some identical to Babylonian parameters) for cycles much shorter than the *kalpa* or *yuga* are used. The planet's true longitude is computed for regular intervals within its cycle: these reference values are preserved via an alphanumeric notation in mnemonic sentences (*vākyas*). The astronomer can determine from the given epoch date which of these true longitudes recurs closest to his desired time, and can then find the planet's position at that time by interpolation.[24]

A fourteenth-century adept of the *vākya* tradition, one Mādhava in northern Kerala on the southwestern coast of India, not only produced but also inspired in several generations of students remarkable accomplishments in mathematical astronomy. Working chiefly within the Āryapakṣa canon, they analyzed and revised geometric celestial models as well as trigonometric methods. Infinite series expressions for trigonometric functions, derived by ingenious infinitesimal rationales, and hypotheses of heliocentric orbits for the inner planets Venus and Mercury, were some of their most notable innovations, which however do not seem to have circulated in astronomical knowledge systems elsewhere in India or beyond it until the nineteenth century.[25]

ISLAMIC INFLUENCE IN THE LATE MEDIEVAL PERIOD

Especially in the northern part of India where Islamic cultural influence was most strongly felt, beginning at the start of the second millennium, several innovations from Islamic astronomy entered the Indian tradition. Besides the abovementioned apparent influence of the Islamic *zīj* on the development of the *koṣṭhaka*, these new features included the planispheric astrolabe

[23] Pingree, *Jyotiḥśāstra*, pp. 101–14.

[24] Ibid., pp. 47–51; see also M. S. Sriram, "Vākya System of Astronomy," in Ruggles (ed.), *Handbook*, pp. 1991–2000, and *Computation of True Moon by Mādhava of Saṅgamagrāma*, ed. and trans. K. V. Sarma (Hoshiarpur: Vishveshvaranand Institute, 1973), pp. 17–24.

[25] Much of the voluminous scholarship on this so-called Kerala school is summarized or cited in K. Ramasubramanian and M. S. Sriram, *Tantrasaṅgraha of Nīlakaṇṭha Somayājī* (Delhi and New York: Hindustan Book Agency and Springer, 2011).

(on which the first of several Sanskrit texts was written in the late fourteenth century) and some novel forms of astrology, including the Perso-Arabic techniques that became popular under the name *tājika*.[26] Although Islamic physical cosmology sparked some controversy among astronomers writing in Sanskrit and the corresponding astronomical theory became a focus of serious study at the eighteenth-century court of Jai Singh in Rajasthan, their Aristotelian/Ptolemaic models were not deeply assimilated into the general Indian theory or practice.[27]

[26] These developments are discussed in Yukio Ohashi, "Sanskrit Texts on Astronomical Instruments During the Delhi Sultanate and Mughal Periods," *Studies in History of Medicine and Science* 10–11 (1986–7), 165–81; Pingree, *From Astral Omens to Astrology*, pp. 79–89; and Martin Gansten, "Some Early Authorities Cited by Tājika Authors," *Indo-Iranian Journal* 55 (2012), 307–19.

[27] See, e.g., David Pingree, "Indian Reception of Muslim Versions of Ptolemaic Astronomy," in F. J. Ragep and S. P. Ragep (eds.), *Tradition, Transmission, Transformation* (Leiden: E. J. Brill, 1996), pp. 471–85, and David Pingree, "An Astronomer's Progress," *Proceedings of the American Philosophical Society* 143.1 (1999), 73–85.

26

MATHEMATICS IN EARLY INDIA (1000 BCE–1000 CE)

Clemency Montelle

I glean from the great ocean of the knowledge of numbers a little of its essence, in the manner in which gems are (picked up) from the sea, gold is from the stony rock and the pearl from the oyster shell.

Mahāvīra, *Gaṇitasārasaṅgraha* (mid-ninth century)[1]

OVERVIEW

"A mixture of costly crystals and common pebbles" wrote the tenth century Islamic scholar, Al-Bīrūnī, as he reflected on the early culture of mathematical inquiry in India. Al-Bīrūnī's lapidarian metaphor, echoing that of Mahāvīra (cited in the above epigram) a century before has echoed down throughout history and in ways that he could hardly have anticipated. India's intellectual legacy is staggering and her mathematical tradition contains a complex combination of the consummate and the conventional. Accounting for India's past in the social sciences has always been influenced to some extent by *Indian Exceptionalism*,[2] that is, the conviction that India can be accounted for only on its own special terms, and this sentiment has pervaded modern traditions of social and historical studies with some notable consequences. Indeed, India is contrasted against other early and modern cultures of inquiry in both direct and nuanced ways. Establishing cultural context is an enormous task. It requires consideration of multiple compounding factors: long-established peoples in immense geographical expanse; large dynamic populations; continual interaction with other cultures; polytheistic and multiple religious traditions; complex relations between community and state; the dictating influence of nature on culture;

[1] As translated by M. Raṅgācārya, *The Gaṇitasārasaṅgraha of Mahāvīrācārya with English Translation and Notes* (Madras: Printed by the Superintendent, Government Press, 1912), p. 3.

[2] See, for example, B. Stein, *A History of India* (Boston, MA: Wiley-Blackwell, 1998), pp. 18–19.

the prevalence of oral tradition; and the predominance and popularity of the exegetical tradition, to name a few. All these aspects and more made their mark on the intellectual disciplines, mathematics included. Furthermore, beyond these contextual factors, mathematics proves to be a special case for the historian. Accounting for its history, particularly in India, is to balance the more culturally contingent conceptions of mathematics with enduring, universalistic aspects and epistemological qualities of this discipline more generally. Accordingly, the challenge for scholars when exploring the mathematical tradition in India has been to uncover the many mathematical gems that originated in this culture while keeping close in mind the intellectual circumstances in which this tradition was maintained, as well as the everyday activities and demands that have shaped and sustained mathematical practice and development.

The geographical region of India, particularly in ancient times, encompassed many nations in South Asia which are today considered separate, most notably India, Pakistan, Bangladesh, Nepal, and Sri Lanka. From the outset, the shaping and defining of India was influenced by internal factors and events as well as what lay beyond the subcontinent. India experienced continued contact of varied sorts from all geographical directions. In the period prior to datable written sources, archaeological evidence testifies to the continuous interaction between Indus Valley inhabitants and peoples from western and central Asia. Trade and exchange with settlements along the western coast and Pharaonic Egypt and Mesopotamia can be dated as early as the fifth millennium BCE. Interactions with Southeast Asia date from about the third century BCE, and Chinese records indicate important trade relations from the second century BCE along the east coast. The opening centuries of the Common Era saw ongoing contact with Rome and the Eastern Mediterranean. Centuries later, through invasions in the North, India was affected by the social and political impact of Islam.

The earliest forms of settlement in India can be traced to the Indus and Gangetic plains in the north in about the middle of the third millennium BCE with the first urban sites and the beginnings of the so-called Harappan culture. Large cities, notably Mohenjo-Daro and Harappa, as they were later named by archaeologists, grew and were characterized by strong centralized control, sophistication of infrastructure, and surplus of food production and trade. But from the beginning of the second millennium BCE, archaeological records reveal that this civilization began to wane, and in about the middle of the second millennium BCE peoples from around the Caspian sea, speaking an Indo-European language and calling themselves Arya, entered and settled in the northern reaches of the subcontinent, bringing with them social structures, agrarian skills, and religious hymns and ritual, which were orally transmitted and preserved. They founded and spread culture based on a canon of "hymns" known as the Vedas, and their prominence ensured that the language they employed, Sanskrit, became dominant. The political structure of this early period began as tribal kingdoms that evolved into sixteen "great communities" (*mahajanapadas*) in the northern "Gangetic" plain, which were eventually transformed into ruling lineages of monarchies in about the sixth century BCE. In addition to Vedic culture, the middle of the first millennium also saw the emergence of the two religious traditions of Jainism and Buddhism, and their influence on culture and intellection was considerable.

The late sixth century witnessed the expansion of the Persian Empire as far as the northwestern Gandhara region which was taken over by Alexander the Great several centuries later in the late fourth century. Later in the fourth century, northern India was brought under unified rule by Candragupta Maurya, who founded the Mauryan kingdom, a dynasty which was to last for some 550 years until 230 CE, when another Candragupta founded the Gupta kingdom. This period saw the composition of the two major epics, the *Mahābhārata* and the *Rāmāyaṇa*,[3] as well as the codification of the Sanskrit language by the grammarian Pāṇini in the *Aṣṭādhyāyī*, and probably the *Arthaśāstra*, a detailed and prescriptive political work on statehood and governance. Arguably the most famous of the Maurya rulers was Candragupta's grandson, Aśoka, who reigned from 270 to 230 BCE. Aśoka was well known for his pacifist rhetoric and at one point had almost all of the Indian subcontinent under his control.

During this time, there was repeated and frequent interaction with neighboring foreign cultures which resulted in conflict as well as commercial and intellectual exchange. In the northwest, post-Alexander, Greek and Indian traditions blended. Groups, known as the Śakas, arrived from central Asia in about the first century CE and expanded into western India. The Kuṣānas, groups originally from Mongolian regions, consolidated their command of northern and western India by about the second century CE. Trade links between them as well as regions in Southeast Asia were established and extensive trade continued with the Roman Empire. From the east, Buddhist pilgrimages arrived regularly from China.

By the sixth century CE, the Gupta empire declined, its downfall exacerbated by the destabilizing invasions from groups in central Asia, the Huns. While overland interaction declined with the demise of the empires to the northwest of India, communication by sea flourished, particularly between the east and southwest of India, and her Southeast Asian neighbors. Arab Muslims traded and settled down along the southwestern coast. Despite these continual changes and transformations, by the seventh century there was established both internally and internationally a unified geographical

[3] Dating these works is somewhat contentious. See J. A. B. Van Buitenen, *The Mahābhārata, Volume 1, Book 1: The Book of the Beginning* (Chicago, IL: University of Chicago Press, 1973), who argues that the oldest portions of the Mahābhārata epic are probably no earlier than 400 BCE and the text developed into its canonical form around the early fourth century CE. The Rāmāyaṇa is thought to be slightly earlier than this.

and political entity known as India, the landmass extending from the mountainous regions of the Himalayas to its southern tip. Furthermore, by this time there can be identified a large-scale adherence throughout the entire subcontinent to common political, economic, and cultural standards and institutions.

Therefore, far from being isolated, Indian peoples and her subcontinent were developed by means of complex and continual interchange with the other early cultures of inquiry, and influences were reciprocal. Given continual political turbulence throughout India's early stages and the continual exchange and assimilation of foreign ideas, her intellectual traditions, particularly the astral sciences, have sometimes been dubbed "the recipient and remodeler of foreign ideas."[4] In light of the substantial impact that foreign inspiration had on her mathematical tradition, historians have the delicate task of identifying the origins and inspiration of the Indian exact sciences. To what extent did transmitted ideas affect the Indian tradition, and how were these ideas assimilated? In what ways was mathematical activity sustained and how did it connect with other intellectual disciplines? What sparked growth and development, and how can we rationalize the many diverse strands of activity that count as mathematics, from those that focused exclusively on utilitarian everyday needs to those which were furthered for more abstract contemplation?

Historians face a certain difficulty in establishing the commonalities, boundaries, and unity of mathematical practice in India. Making sense of the plurality of practice that the sources reveal is a challenge. Mathematical inquiry flourished in various social groups on the Indian subcontinent who were characterized by starkly contrasting ideals, customs, and conventions; these groups fostered and furthered mathematical ideas for quite distinct purposes. The range, scope, and purpose of mathematical practice and inquiry are thus far-reaching. Questions arise for historians which are both critical for characterizing the internal character of India's mathematical traditions, as well as for the comparisons that can be made with other active ancient cultures of inquiry.

A careful analysis of texts and their contents over this early period reveals that although they cover many of the same topics – issues such as order of topics, applicability and scope, how subfields relate and interact with each other, and the standards of results – they were all handled differently by different authors. In many respects, mathematics in this early time was non-standardized. Issues of professional integrity, rivalry, novelty, and elitism entailed that its character was determinedly individualistic, purposefully esoteric, and, at times, competitive.

[4] D. Pingree, "History of Mathematical Astronomy in India," in *Dictionary of Scientific Biography*, vol. 15 (Detroit, MI: Charles Scribner's Sons, 1978), 533–633.

What, then, of Al-Bīrūnī's characterization of the mathematics of India being a combination of gems and pebbles? If anything, his estimation of the mathematical scene in India epitomizes the challenges early thinkers faced as they endeavored to make sense of it on their own terms. These challenges are just as relevant to modern scholars as well. Al-Bīrūnī as a foreign scholar was interested in what was distinctive or exceptional in this so-called ocean of knowledge, but those ideas he singled out, in so far as they were deemed by him to be valuable or not, are more a reflection of his own circumstances and the audience he was addressing. His somewhat critical assessment is more a manifestation of his own prejudices rather than the traditions he was evaluating. By contrast, recent thinkers have proposed that exploration of any aspects of Indian traditions needs to begin "from the inside out";[5] investigation must prioritize the internal coherence of a discipline. But this approach too, when applied to mathematics (and more broadly the exact sciences), has its challenges, as mathematical knowledge is frequently championed as universalistic and seemingly transcendent of culture.

To this end, we will explore the notion of the Indian mathematical tradition and the ways in which it was ultimately and outwardly shaped by both the ambient social circumstances and the intellectual demands in which it was carried out. Accordingly, we will highlight features of this mathematical tradition in early times which were maintained and developed at the intersection between oral and literate culture, as well as situate it in the broader intellectual trajectories which delineate the development of mathematics as a human pursuit throughout history.

A CONTEXT FOR MATHEMATICS

Mathematics in early India had fundamental connections with astronomy and astrology, and was developed under the larger discipline of the exact sciences, captured by the Sanskrit term *jyotiḥśāstra*, which encompassed *gaṇita* (mathematics), as well as *saṃhitā* ((celestial) divination) and *horā* (astrology). The term *gaṇita* was used in many senses and its scope developed over time. In early times it was generally restricted to the computational procedures related to astronomy, perhaps most properly considered "mathematical astronomy." These were usually presented as part of an astronomical treatise (*siddhānta*), albeit in a separate section, and given that these sections were never placed at the beginning of such works, it may be inferred that *gaṇita* was not seen as foundational but rather complementary. But over time, the meaning of the term broadened, and it was

[5] See S. Pollock, "Sanskrit Literary Culture from the Inside Out," in S. Pollock (ed.), *Literary Cultures in History: Reconstructions from South Asia* (Berkeley, CA: University of California, 2003), pp. 39–130, who applied this methodological principle to literary traditions.

used to encompass general computational procedures which could apply to astronomy as well as other contexts. By about the eighth century CE, *gaṇita* eventually broke free from its astronomical context (although it always remained an integral part of astronomy) and was also advanced as a discipline independent of the astral sciences.

Mathematician Śrīpati, using the metaphor of a tree, gives an indication of the ways in which the discipline was understood by the eleventh century, as well as some of the specific applications:

> Knowing well that *gaṇita*, comprising planetary computations (*graha-gaṇita*), arithmetic (*pāṭigaṇita*), and algebra (*bījagaṇita*), forms the deep roots of the tree of the science of *jyotiṣa*, that astrology (*horā*) of diverse aspects forms the branches, and that divination (*saṃhitā*) forms the fruit, here is being set out, in brief, post-natal sacrement, naming the child, girdle ceremony, marriage, travel, etc.
>
> Śrīpati *Ratnamālā* (via the commentary of Mahādeva)[6]

The pivotal role of mathematics was frequently acknowledged. The mid-ninth-century mathematician Mahāvīra is emphatic as to the centrality of computation to everyday life:

> In worldly life or Vedic matters, or even in religious practices, whatever be the dealings, everywhere enumeration is essential. Why say much? In the three worlds . . . whatever is to be transacted, that cannot be done without calculation (*gaṇita*).[7]

In many respects, mathematics played an important socio-political role. One of the key elements that united India was its intricate social conventions (rituals and traditions associated with birth, death, and marriage, etc.), the demarcation of the year, and the measuring of time, and these were ultimately regulated by astronomy, of which mathematics was an integral part. Thus, mathematics most clearly manifested itself through the utilitarian demands associated with public organization and social regulation. Beyond this, mathematics in India never connected formally with other intellectual disciplines (apart from the astral sciences), such as physics, logic, or philosophy as it did in other early cultures of inquiry. In particular, mathematics was always considered as distinct from logic, and connected with practical scientific applications only through astronomy and astrology. Overall, mathematics remained a technical subject and did not have the privileged epistemological status that other antique societies bestowed on it. However, the one discipline that was seen as paradigmatic in Indian intellectual culture

6 As translated by K. V. Sarma in B. V. Subbarayappa and K. V. Sarma, *Indian Astronomy: A Source-Book (Based Primarily on Sanskrit Texts)* (Bombay: Nehru Centre, 1985), p. 2.

7 *Gaṇitasārasaṅgraha* 1.9.16.

was grammar,[8] which was considered emblematic of the structure of knowledge systems and philosophy. Plato's exhortation concerning the skills necessary to participate in the Athenian Academy – "let no one enter here who is ignorant of geometry"[9] – is contrasted by the requirement for Indian scholars to be thoroughly trained in and conversant with the science of grammar: "Grammar is the most important of all the sciences."[10] Methodologically speaking, only scant attention was given to issues such as rigorization, standards of practice, or exploring the foundations of mathematics. The requirements too of an oral environment had a significant impact on content, format, and delivery. For these reasons, mathematical activity in India was furthered and developed in ways that contrast markedly with other ancient cultures.

Despite the lack of formal connection to philosophy and related subjects, there was a sense of a connection to the super-mundane. *Jyotiḥśāstra* was a discipline divinely revealed by Gods and *ṛṣis* (ancient sages),[11] mathematics included:

> Those treatises are victorious which are accurate in mathematics and spherics because of (their authors') flashing intelligence (and) which were proclaimed by (the Gods) Sūrya, Soma, Brahmā (and by the *ṛṣis*) Vasiṣṭha, Garga, Atri, Romaka, Pulastya, and Parāśara.
>
> Nityānanda *Sarvasiddhāntarāja*[12]

The attribution of divine authorship gave works their esteem and authority. On the face of it, proficiency in *gaṇita* was not predominantly originality. Rather, attributes that were valued were that one had a prodigious memory; could carry out mental arithmetic quickly and accurately and thus deliver an answer swiftly; could satisfy the request for explanation by a quick example, diagram, or illustration; and could nurture a rounded balance between respect for authority of received learning and a certain inventiveness to devise new procedures to tackle old problems.

Mathematics and its connection to the super-mundane manifests itself more broadly also. For several key religious traditions in India, mathematical

8 See J. F. Staal, "Euclid and Pāṇini," *Philosophy East and West*, 15.2 (1965), 99–116: "Historically speaking, Pāṇini's method has occupied a place comparable to that held by Euclid's method in western thought" (p. 114).

9 For the various ancient attributions of this phrase see the entry *ageōmétrētos* in H. G. Liddell and R. Scott, *A Greek–English Lexicon* (Oxford: Clarendon Press, 1996).

10 From the *Haracaritacintāmaṇi* XXVII 269 *sarvavidyānām mukhyam vyākaraṇam*; as cited in R. Torella, "Examples of the Influence of Sanskrit Grammar on Indian Philosophy," *East and West* 37.1 (1987), 151–64, 152, fn. 6.

11 See, for instance, D. Pingree, "The Logic of Non-Western Science: Mathematical Discoveries in Medieval India," *Daedalus* 132.4 (2003), 45–53.

12 As cited from D. Pingree, "Indian Reception of Muslim Versions of Ptolemaic Astronomy," in F. J. Ragep, S. P. Ragep, S. J. Livesey (eds.), *Tradition, Transmission, Transformation: Proceedings of Two Conferences on Pre-Modern Science held at the University of Oklahoma* (Leiden: Brill, 1996), pp. 471–88, p. 477

inquiry offered a platform for the contemplation of the sacred, including divine ritual, practice, and metaphysical rumination (see sections "Śulbasūtras" and "Jainism") and many profound mathematical concepts and properties were nurtured in these contexts.

LANGUAGE AND TEXTUAL SOURCES

The majority of early mathematical works were composed in Sanskrit. This language and its closely related predecessor, Vedic, are among the oldest members of the Indo-European language family. By the end of the first millennium BCE, Sanskrit (literally: "Perfected (Speech)") had become primarily a learned language. Sanskrit, commonly written in a script called *devanāgarī*, is a heavily inflected language with a large vocabulary base and complex grammar, and is organized and built around strict phonological rules. Despite the development and prevalence of vernaculars in various geographical regions, "Classical" Sanskrit, as it came to be known, remained the primary scholarly language until recently and is still used in some contexts today. Notably, Sanskrit has, for the most part, resisted many of the usual pressures of the passage of time. This remarkable feature is to a large extent due to the contributions of the fourth-century BCE grammarian Pāṇini and his substantial authoritative codification of the Sanskrit language called the *Aṣṭādhyāyī*. This work became central in traditions of learning, and grammar was established as a core component in education and training.

In the earliest times, the corpus of Sanskrit was made up primarily of religious and supporting works; however, as time passed, a rich body of literature was produced, including poetry, epic, drama, technical treatises, mathematics, and astronomy, to name a few. The sources for the study of mathematics are vast. Some surveys have estimated that there exist around thirty million manuscripts today[13]; more conservative estimates have been made at five to seven million[14]. Depending on how one might define what counts as falling under the mathematical sciences, as high as ten per cent of these manuscripts may be related to mathematics in some way.

This abundance of written sources is both a boon and a challenge for historians. Due to environmental circumstances, as well as archival procedures, the access we have to original texts is through copies made by many successive generations of scribes. The oldest Sanskrit manuscripts we have,

[13] See Pingree, "The Logic of Non-Western Science," p. 46; D. Pingree, "Review of Staal, *The Fidelity of Oral Tradition and the Origins of Science*," *Journal of the American Oriental Society* 108.4 (1988), 637–8, 638; and Dominik Wujastyk, "Indian Manuscripts," in Jan-Ulrich Sobisch, Dmitry Bondarev, and Jörg Quenzer (eds.), *Manuscript Cultures: Mapping the Field* (Stuttgart: Steiner Verlag, 2014), pp. 159–82, p. 160.

[14] B. N. Goswamy, *The Word Is Sacred, Sacred Is the Word* (New Delhi: National Mission for Manuscripts, 2007), p. 17.

therefore, are rarely older than a few centuries, due to the fragility of the writing media and the ambient environmental conditions in India. This makes them in many instances, for ancient works, more than a thousand years removed from their original time of composition.[15] Careful textual studies and codicology, though, can give modern scholars in many cases a firm sense of the original text.

FORMAT

From the earliest times in India, the spoken word was esteemed above all else, with the result that almost all disciplines were carried out in an oral environment. Mathematics was no exception, and was sustained and developed in line with the standards of oral learning. Mathematical techniques, computations, and results were produced to be recited and memorized.[16] To facilitate this requirement that content should be committed to memory, mathematical ideas were composed in verse characterized by metrically rhythmical and repetitive patterns. Typically, these verses were made up of a predetermined arrangement of heavy and light syllables, grouped into four "feet" or quarter-verses (*pādas*). This had some interesting consequences for mathematical content. Firstly, long and involved technical expressions have little traction in an oral environment and they were frequently abbreviated or syncopated. Consequently many mathematical verses in Sanskrit are very succinct. Commonly these verses were not intended to be a rich explanatory source, but rather more like a memory-jog, whose missing information was supplied by context. For example, the Sanskrit verse:

> *karṇakṛteḥ koṭikṛtiṃ viśodhya mūlaṃ bhujo bhujasya kṛtiṃ* |
> *prohya padaṃ koṭiḥ koṭibāhukṛtiyutipadam karṇaḥ* ||
>
> Brahmagupta *Brahmasphuṭasiddhānta* 12.24

translates literally to mean:

> Subtracting the square of the upright [of a right-angled triangle] from the square of the hypotenuse, the square-root [of the remainder] is the base [of a right-angled triangle]. Subtracting the square of the base . . . the square-root [of the remainder] is the upright. The square-root of the sum of the squares of the upright and the base is the hypotenuse.

This procedure will be recognized by many as the so-called Pythagorean theorem which relates the squares of the sides (here, the upright and the

[15] Exceptions, of course, do exist, which are generally written on media other than Indian paper. See, for instance, the *Bakhshālī* manuscript (section entitled "Bakhshālī Manuscript").

[16] The extent to which works were actually memorized verbatim can only be guessed at. See, for instance, K. Plofker, "Sanskrit Mathematical Verse," in E. Robson and J. Stedall (eds.), *The Oxford Handbook of the History of Mathematics* (Oxford: Oxford University Press, 2009), pp. 519–36, p. 524.

base) of a right-angled triangle to its hypotenuse. However, translated literally the verse seems incomplete (as indicated by the " . . . " and the need to add prose as indicted by the use of the brackets in the translation). In fact, the " . . . " in the middle can be reconstructed using context: "from the square of the hypotenuse" here completes the mathematical sense. This necessary phrase has simply been omitted due to the lack of syllables remaining in the verse; it is assumed by the author that it can be generated from the context (or reread from the previous sentence).

Secondly, technical terminology was flexible. Given the strict requirements of meter and the premium of space, composers required at their disposal multiple terms that referred to the same concept so that they could select one that would insert into the required metrical pattern. Accordingly, a feature of the Sanskrit mathematical terminology is that there were generally many synonyms for any one concept. For example, the concept "base of a right-angled triangle" could be expressed by *bāhu* (lit. arm, but can also refer to a door post, gnomon shadow, or constellation, etc.) or *bhūja* (lit. hand, but can also mean an elephant's trunk, a branch, or a side of a geometrical figure in general). Likewise, the concept "three" could be expressed by the term *tri* (literally 'three'), *agni* (fires), as well as all synonyms for fires and worlds. As well as many words to express a single concept, single words could have flexible technical meanings depending on context. Thus, the Sanskrit word *pāda* means foot, and by extension could mean "the base of a triangle," but also "the square root," and sometimes even "the result." It was these features that prompted Al-Bīrūnī to conclude that Sanskrit mathematical verse contained "much misty and constrained phraseology merely intended to fill up the meter and serving as a kind of patchwork, and this necessitates a certain amount of verbosity,"[17] but, of course, such an evaluation lacked insight into the intellectual constraints of the oral environment that we have noted.[18] This feature was an ingenious and deliberate way to coax technical content into a predetermined fixed format. The formation of a standard and consistent nomenclature for mathematics was simply not an aspiration, given these constraints.

Despite this veneration of oral delivery and transmission, the written word was integral to intellectual pursuits as well, and by the early centuries in the first millennium BCE, writing was commonly used in India. In the mathematical context, written commentaries were produced to assist with the meaning of mathematical verses and became an integral part of the tradition in and of themselves. Commentaries could do a number of things: they could restate the verse in prose, offering synonyms for various terms;

[17] E. C. Sachau, *Alberuni's India: An Account of the Religion, Philosophy, Literature, Geography, Chronology, Astronomy, Customs, Laws and Astrology of India about A.D. 1030, Volume 1* (London: Routledge, 2001), p.19.

[18] For more features of this oral environment see Plofker, "Sanskrit Mathematical Verse."

parse the grammar and analyze complex grammatical expressions; explain more fully the mathematical content alluded to in the verse; give definitions for technical terms or concepts; and give justifications as to why the rule in question worked, sometimes by way of a worked example. More broadly and less commonly they might compare and contrast other works that dealt with the same subject, excerpting and quoting relevant passages, give some biographical information on the author, occasionally offer an abstract exploration of the underlying meaning, draw a diagram, provide instructions on how to construct the problem, and the like. For this reason, commentaries provide valuable testimony regarding the status and reception of mathematical works in subsequent generations.[19]

NUMERALS AND SYSTEMS OF NUMERATION

One of the most significant innovations developed by Indian thinkers was a base-ten positional, or place-value, number system which (after several additional developments) is now known as the Indo-Arabic decimal place-value system. Unlike previous systems of numeration, any number whatsoever could be depicted by means of only ten distinct symbols or glyphs.[20] Inscriptions in caves and early coins reveal a system of numeration which had distinct glyphs to depict the first numerals (as well as their multiples of tens and hundreds). Now called Brahmi numerals, these symbols seem to have been critical to the insight of a base ten place value system. These glyphs were modified stylistically around the fourth century CE, and then evolved into the *nāgarī* numerals (from the term *devanāgarī*) around the seventh century, and continued to be modified until the eleventh (see Figure 26.1). These numerals were transmitted to the Islamic world and beyond, and had significant consequences for the construction and process of arithmetical algorithms.

One parallel trajectory of numerals and the developing place value notation is attested in the early Jain tradition (see section entitled "Jainism"). It has been hypothesized that as early as the fourth or fifth century BCE the issue of numeration was addressed by followers of this religious tradition. In their literature, there exists a list accounting for the different types of

[19] See, for instance, A. Keller, "On Sanskrit Commentaries Dealing With Mathematics (Fifth–Twelfth Century)," in F. Bretelle-Establet (ed.), *Looking at it from Asia: The Processes that Shaped the Sources of History of Science* (Boston Studies in the Philosophy of Science, volume 265; Dordrecht: Springer, 2010), pp. 211–44.

[20] Contrast, for instance, the system of numeration in the Ancient Near East, which was sexagesimal and only partially place-value, using just two distinct glyphs to represent all numbers in base 60, and also the Ionian system in Greece (see C. Montelle, "Roots, Rocks, and Newton Raphson: Algorithms for Approximating the Square Root of 2 3000 Years Apart," in D. Jardine and A. Shell-Gellasch (eds.), *Mathematical Time Capsules: Historical Modules for the Mathematics Classroom* (MAA Notes 77; Washington, DC: Mathematical Association of America, 2011), pp. 229–50), which was not place-value and based on the 24 letters of the Greek alphabet and required additional notation to represent numbers greater than 10,000.

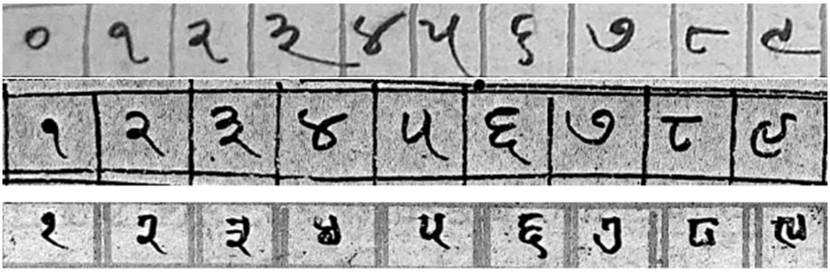

Figure 26.1. Examples of the *nāgarī* numerals from various manuscripts.

characters (*lipi*), which includes "number-characters" (*aṅka-lipi*) and "computation-characters" (*ganita-lipi*).[21] What is significant about this distinction is the separation of systems of numeration for representation and for computation. It has been suggested that the former types of numerals were associated with engraving techniques and the latter with writing and copying. Thus while paleographic evidence provides modern scholars with a concrete means to trace the development of systems of numeration, lexicographical evidence too can offer insight into the practices of early thinkers.

One of the consequences of an oral environment is that long strings of numbers needed to be made memorable; it is near impossible to compose a memorable verse the majority of which is a list of numbers. There were several ingenious ways developed to weave numbers in a memorable way into the metrical environment. These fall into two main strands: systems which create synonyms for number-words; and systems which develop ways to represent numerals through some sort of alpha-numeric coding.

Perhaps the most common and the earliest system,[22] known as the *Bhūtasaṃkhyā* system (sometimes known as object-numerals) in which number words could be substituted by everyday objects and mythological items with corresponding numerical significance. For example, the word "hand" or "eye" could be substituted for the number-word "two"; the "moon" could be used for one; "tastes" for six (sweet, bitter, sour, salt, pungent, astringent); "elephants" or "snakes" (by mythological reference) for eight; and so on. Longer strings of numbers could be produced from compounded smaller numbers which by careful arrangement could conjure up visual or imaginative analogies. So for example, the following verse:

vibudhanetragajāhihutāśanatriguṇavedabhavāraṇabāhavaḥ |

[21] See L. C. Jain, *Exact Sciences from Jaina Sources: Basic Mathematics* (Prakrit Bharti Pushpa-12; Jaipur: Rajasthan Prakrit Bharti Sansthan, 1982), pp. 38ff. This list has been included in the *Prajñapanāsūtra* of Syāmārya (d. ca. 150 BCE).

[22] The earliest occurrence known is a third-century CE astrological text. See Pingree, "History of Mathematical Astronomy," p. 506.

navanikharvamite vṛtivistare paridhimānam idaṃ jagadur budhāḥ ||

literally translates to

> Gods, Eyes, Elephants, Snakes, Fires, Three, Qualities, Vedas, Constellations, Elephants, (and) Arms. The wise said that this is the measure of the circumference when the diameter of a circle is nine hundred billion (10^{11}).
>
> Mādhava[23]

Using the *Bhūtasaṃkhyā* system of numeration, which assigns the following numbers to the above words:

vibudha	Gods	33
netra	Eyes	2
gaja	Elephants	8
ahi	Snakes	8
Hutāśana	Fires	3
Tri	"Three"	3
Guṇa	Qualities	3
veda	Sacred Books	4
Bha	Constellations	27
Vāraṇa	Elephants	8
bāhu	Arms	2

the number "encoded" in this verse is (taking the digits in reverse order as was standard):

$$2\ \ 827\ 433\ 388\ 233$$

and

$$\frac{2827433388233}{900000000000} = 3.14159265359$$

which is a rational approximation to π correct to 11 decimal places when divided out. This verse then provides a quick memorable means to call to mind π when an accurate value is required.

The most prominent system of the alphanumeric kind is the so-called *kaṭapayādi* system which dates to around the middle of the first millennium CE and was popular in south India. According to this system each of the thirty-three consonants of the Sanskrit language is assigned, in a particular order, one of the digits from 0 to 9. Numbers are then represented by stringing together appropriate consonants, with the eventual aim of spelling out a real Sanskrit

[23] As cited from Pingree, "The Logic of Non-Western Science," p. 49.

Table 1: *The numerical equivalences in the kaṭapayādi system*

1	2	3	4	5	6	7	8	9	0
ka	kha	g a	gha	ṅa	ca	cha	ja	jha	ña
ṭa	ṭha	ḍa	ḍha	ṇa	ta	tha	da	dha	na
pa	pha	ba	bha	ma					
ya	ra	la	v a	śa	ṣa	sa	ha		

word or, more ambitiously, a meaningful phrase. Entire texts were composed of just such phrases, known as *vākyas* (literally: sentences).[24]

A single verse was composed to explain how the system worked:

> *nañāvacaśca śūnyāni saṃkhyāḥ kaṭapayādayaḥ*
> *miśre tūpāntyahal saṃkhyā na ca cintyo halasvaraḥ*

> *na, ña* and the vowels represent zero. (The consonants) beginning with *ka, ta, pa*, and *ya* denote, in order, the (nine) digits. In a conjunct consonant,[25] the last of the consonants alone counts. The vowel suffixed to a consonant, too, is to be ignored.
>
> Śaṅkara Varman *Sadranamālā* 3.4[26]

Thus, in the *kaṭapayādi* system, numbers are allocated as is indicated in Table 1. As is clear from the table, the name of the system of numeration comes from the set up of the system: *ka-ṭa-pa-ya* are the four letters which designate the number 1, and *ādi* is a suffix in Sanskrit which means "and so on" to produce the name *kaṭapayādi*. The flexibility and facility of this system meant that scores of important numbers could be memorized with ease. The system was most popular in the south of India, particularly Kerala from perhaps as early as the fourth century CE.

The *kaṭapayādi* system was most commonly employed in *vākya* texts. *Vākya*, literally "sentence," was a phrase which not only encoded a numerical parameter, but also translated into a meaningful Sanskrit sentence. For instance, the following example reads as a pithy aphorism:

> *Śrīguṇamitrā*
> Wealth is a friend of virtues.
>
> *Vākyakaraṇa* 4A.

But it also encodes the numbers 2–55–32. According to the rules and to the correspondences given in Table 1, the number can be retrieved as follows: disregard the *ś*; the *r* is 2; *g* is 3; *ṇ* is 5; *m* is 5; disregard the *t*; the *r* is 2.

[24] K. Mahesh, "Saṅkramavākyas of the *Vākyakarana*," *Indian Journal of History of Science* 49.2 (2014), 157–70.
[25] That is, two consonants appearing next to each other in a word.
[26] As cited in Subbarayappa and Sarma, *Indian Astronomy*, p. 47.

Placing these in a string and reading backwards gives us the desired number, which, given the context, can be interpreted as 2 days, 55 degrees, and 32 minutes, which is an important astronomical additive relating to the time of the transit of the sun from one zodiacal sign to another.[27] Thus, large collections of long strings of technical numbers could be memorized and retained with relative ease.

Another system, which was proposed and used almost exclusively by late fifth-century CE mathematician Āryabhaṭa, was the *varga* system. This system, like the last, was alphanumeric and somewhat idiosyncratic, which perhaps explains the fact that it never became widely used. However, it was used in Āryabhaṭa's famous encoding of his tabulated values of Sines. The name of the system *varga* refers to the technical grammatical term to refer to the grouping of the first twenty-five consonants of the Sanskrit alphabet into five groups of five[28] and the *a-varga* consonants are the remaining eight:

***varga* letters**				
ka	kha	ga	gha	ṅa
1	2	3	4	5
ca	cha	ja	jha	ña
6	7	8	9	10
ṭa	ṭha	ḍa	ḍha	ṇa
11	12	13	14	15
ta	tha	da	dha	na
16	17	18	19	20
pa	pha	ba	bha	ma
21	22	23	24	25

***a-varga* letters**							
ya	ra	la	va	śa	ṣa	sa	ha
30	40	50	60	70	80	90	100

Āryabhaṭa explained the system as follows:

Vargākṣarāṇi varge 'varge 'vargākṣarāṇi kāt ṅamau yaḥ |
svadvinavake svarā nava varge 'varge navāntyavarge vā ||

> The *varga* letters from *k* (to *m* should be written) in the *varga* (places) and the *avarga* letters (from *y* to *h*) in the *avarga* (places). (The numerical value of the initial *avarga* letter) *y* (is equal) to *ṅ* plus *m*. (In the places of) the two nines of zeros, the nine vowels (should be written) in the *varga* and *avarga*

[27] For full details, see Mahesh, "Saṅkramavākyas of the *Vākyakaraṇa*," pp. 160–2.
[28] K. Plofker, *Mathematics in India* (Princeton, NJ: Princeton University Press, 2009), pp. 73–5 also notes the pun, as in a mathematical context *varga* also means "square."

(places). In the *varga* (and *avarga* places) beyond (the places denoted by) the nine vowels too (assumed vowels or other symbols should be written, if necessary).

Āryabhaṭa I *Āryabhaṭīya* I.2[29]

This verse largely described how to lay out the Sanskrit consonants and then assigned each number to them. The system allocated each number in ascending order to each Sanskrit consonant beginning at 1 from *k* until 25 at *m*, and then from *y* onwards (the non-varga consonants), in multiples of 10 beginning from 30 (i.e., *ṅm* = 5 + 25 = 30) to 100. The *varga* and *a-varga* organization was also applied to vowels which gave digits their order of magnitude up to ten to the power of "two-nines" (18). Each vowel could have two interpretations as follows:

The vowels

	a	i	u	ṛ	ḷ	e	ai	o	au
varga	10^0	10^2	10^4	10^6	10^8	10^{10}	10^{12}	10^{14}	10^{16}
a-varga	10^1	10^3	10^5	10^7	10^9	10^{11}	10^{13}	10^{15}	10^{17}

This system was most famously used in Āryabhaṭa's description of sine differences:

makhi-bhakhi-phakhi-dhakhi-ṇakhi-ñakhi-ṇakhi-hasjha-skaki-kiṣga-śghaki-kighva-ghlaki-kigra-hakya-dhāhā-sta-sga-śjha-ṅva-lka-pta-pha-cha

225, 224, 222, 219, 215, 210, 205, 199, 191, 183, 174, 164, 154, 143, 131, 119, 106, 93, 79, 65, 51, 37, 22, 7

Āryabhaṭa *Āryabhaṭīya* I.12[30]

For instance, the term *makhi* encoded the number 225: *m* was 25, the *a* 10^0, so that the syllable *ma* gave 25, and *kh* was 2 and the *i* 10^2, so that the syllable *khi* was 200, hence producing the number 225.

The main advantage of this system was that very large numbers could be expressed quite concisely. This was particularly useful for Āryabhaṭa for the large numbers he used in astronomical computations. However, one of its great drawbacks, which undoubtedly led to its demise, was its inflexibility. Unlike the *kaṭapayādi* system, which had the flexibility of several letters at one's disposal for a single number, it was only sheer coincidence that one might be able to compose an actual meaningful Sanskrit word using this *varga* system. For any number, one usually had a single choice. Thus, the above verse composed by Āryabhaṭa, even to the ears of one familiar with Sanskrit, was but a string of

[29] Based on a translation as given by Subbarayappa and Sarma, *Indian Astronomy*, p. 47.
[30] As cited from K. S. Shukla, *Āryabhaṭīya of Āryabhaṭa with the Commentary of Bhāskara I and Someśvara* (New Delhi: Indian National Science Academy (INSA), 1976), p. 29.

nonsense sounds, awkward clusters of consonants, which must have been both a chore to memorize, and quite disconcerting to listen to.

The creation and maintenance of several different types of schema for representing numbers, many of which coexisted alongside the place-value ten-glyph system for denoting numbers, reveal an interesting priority amongst mathematical practitioners. Of importance was making numbers memorable, and this took precedence in many cases over representing a number in its place-value determined arrangement. It also highlights an important distinction between how numbers were recorded and stored in India, and how they were used in the context of arithmetical manipulation and so forth. More broadly, it raises the deeper philosophical question of the minimal requirements for a functional numerical system. While alpha-numerical, object-association number systems are playful and creative, they bring with them limitations in a technical setting. They are not at all suited to the requirements of actually performing a calculation, and, for the most part, we can only guess at how practitioners actually carried out their numerical computations and arithmetic algorithms in the practical context.

ORIGINS AND EMERGENCE

MATHEMATICS IN VEDIC INDIA

The Vedas are sacred texts and they date to around the middle to late second millennium BCE. These texts were religious and contained hymns and invocations that were to be spoken during rituals. In association with these texts were developed six disciplines that were devoted to supporting and maintaining the traditions associated with the Vedas, the so-called *vedāṅga* (literally "six-limbs") and included pronunciation, grammar, etymology, prosody, ritual practice, and astronomy and calendrics. Mathematics in this context was developed to deal with both the geometric aspects involved in the construction of altars and the computations required to compute the date and time involved with calendrics. Over time though, the discipline matured and expanded. While remaining closely linked with religious demands, mathematics developed in ways that were independent of sacred application, flourished as a product of other worldly applications (including astronomy, astrology, and prosody), and was driven by the curiosity and capabilities of certain key individuals. These supporting texts often presented their content in verses, or in short sentences, which were called *sūtras*, literally "strings" or rules.[31] The significance for mathematics of the concept of *sūtra* is of great importance.

[31] This word originally meant a condensed aphorism, but over time became associated with a "rule" or "algorithm."

The hymns of the *Ṛgveda* (lit. Praise-Vedas), the oldest Vedic compositions, contain references to numbers, which are imbued with sacred significance. Numbers appear as worthy of praise in the later *Yajur Vedas* (lit. Sacrifice-Vedas): "… praise to two, praise to four, praise to six … "[32] and some recensions list this type of praise-pattern of multiples of ten up to a trillion! A key passage from a related text, the *Śatapathabrāhmaṇa* (lit. Brāhamana of One Hundred Paths, ca. 1000 BCE) associates the building of a sacrificial altar with the number 720, the number of bricks associated with the creator god Prajāpati.[33] This passage outlines systematically the number of ways Prajāpati arranges these bricks into various divisions, notably dividing by all the integers from two to thirty which produced integer piles upon division; for those numbers such as seven and eleven, which do not, it is explicitly stated that he did not divide.[34] Thus, in this cosmological context, the notion of divisibility is privileged. These early texts then reveal a clear interest in numerical quantities, both finite and infinite, and their composition, and the ways in which they embody cosmological and numinous significance reveal a noteworthy commitment to mathematical structures.

Related Vedic texts indirectly testify to some intricate mathematical procedures which are found in quite different contexts. For instance, the *Chandaḥsūtra* (lit. Rule(s) of Prosody) is a text on prosody by the author Piṅgala (fl. before 200 BCE), where a rule for computing the number of possible syllable patterns for a given verse is expressed. In the Sanskrit context there were two options for each syllable: light (*laghu*) and heavy (*guru*). The rule Piṅgala gives for a verse of n syllables amounts to the computation of 2^n by doubling and squaring, rather than the slower procedure of successive doubling.[35] This algorithm reveals some creative insights in number theory and in fact anticipates aspects of the ways in which modern computer science tackles this problem today.

ŚULBASŪTRAS

Mathematical activity in the early part of the first millennium BCE arose and flourished as a result of religious observances. Geometrical procedures were developed as a result of the requirements of the construction and measurement of elaborate altars as part of Vedic rituals, the

[32] *Yajurveda* VII 2.11–20, as cited from K. Plofker, "Mathematics in India," in V. J. Katz (ed.), *The Mathematics of Egypt, Mesopotamia, China, India, and Islam: A Sourcebook* (Princeton, NJ: Princeton University Press, 2007), pp. 385–514, p. 386.

[33] E. Robson, *Mesopotamian Mathematics, 2100–1600* BC*: Technical Constants in Bureaucracy and Education* (Oxford, Clarendon Press, 1999) notes that the number 720 is a fundamental metrological unit in the Ancient Near East, known as the "brick sar."

[34] See Plofker, *Mathematics in India*, p. 15, for the full passage.

[35] For details, see Plofker, "Mathematics in India," pp. 393–4.

orientation, shape, and areas of which had to be precise for religious purposes. These collections of short prose-sentences were known as the *Śulbasūtras* (lit. rules of the rope), a term compounded from the word *Śulba* meaning rope or cord, and *sūtra* meaning "rule." The *Śulbasūtras* were likely codified around 800 BCE and their contents outline predominantly construction procedures which, as their name implies, are intended to be effected by a rope and a marker.[36] However, at times, abstract geometrical relations are stated, such as the mathematical relation between the two sides and the hypotenuse of a right-angled triangle, which later came to be known as the Pythagorean theorem (notably some five centuries before Pythagoras) and precise approximations for key values such as π and $\sqrt{2}$.

The work opens with a detailed description of metrical units and their conversions, which is followed by procedures for constructing squares, rectangles, triangles, trapeziums, and so on. Next, more elaborate procedures are spelled out which consider the basic shapes and their transformations from one to the other, and combining any two resulting areas or taking their difference. Of note are the attempts to transform a square into a circle while preserving the area.[37] The geometric procedures have been considered by some historians to be a form of "geometrical algebra,"[38] that is, that the procedures expressed in discourse describe more general algebraic relations (such as quadratic equations) in a geometrical context. Out of these basic geometrical shapes complex forms were to be produced depending on the desired "effect" (including prosperity, success in battle, territory, and divine favor). Accordingly, altars could be fashioned in the form of a bird, an isosceles triangle, a chariot wheel, a rhombus, a pyre, a circular trough, a tortoise, or an isosceles trapezium, to name a few.

One can only speculate as to the origins of the connection between geometry and the sacred in this context. Whether it was the inspiration of the timeless and essential qualities of geometrical form and spatial property that inspired and shaped religious practice or, conversely, the ultimate requirements of religion that spurred the development of this discipline is a subject which requires further exploration.[39]

[36] Compare, say, the Euclidean "straight edge and compass" requirements.

[37] See, for example, the attempt of the Egyptian Ahmes, as well as Greek consideration of the problem.

[38] See, for example, I. Bashmakova and G. Smirnova, *The Beginnings and Evolution of Algebra* (The Mathematical Association of America, Dolciani Mathematical Expositions, Number 23; Washington, DC: Mathematical Association of America, 2000), pp. 163–72.

[39] It is true that throughout history, mathematics, because of its abstract qualities and its association by many thinkers with the absolute, the pure, and the unchanging, has been regarded as a domain in which to comprehend the divine. See, for instance, T. Koestler and L. Bergmans (eds.), *Mathematics and the Divine: A Historical Study* (Amsterdam: Elsevier, 2005), especially pp. 6–10, and K. Plofker, "South Asian Mathematics," *Encyclopaedia Britannica Online* (2011) <www.britannica.com/EBchecked/topic/1238473/South-Asian-mathematics>.

While the rules alluded to procedures without metric (reminiscent of Euclidean-style geometry), ultimately measure was fundamental in the actual construction of altars. A base length was usually taken to be the "man-height," commonly taken as the length of the patron or sacrificer. In this way, the altars blended elements of the mundane and the divine, and the transcendence of mathematics served as the vehicle between them. For those early cultures whose thinkers pondered the relation of "man" to the universe, via humanistic aphorisms, such as the Presocratic Protagorean *homo-mensura* statement "Man is the measure of all things,"[40] early Indian practitioners manifested this sentiment literally, via the metric for their divine altars.

JAINISM

A rich and nuanced understanding of abstract mathematical ideas emerged in the Jain religious context as early as the fourth century BCE. These ideas were not advanced in a strict mathematical context, but, rather, the Jain religious traditions needed to give quantitative substance to their religious tenets including notions of karmic theory, world order, natural process, and broader cosmological ruminations. The indispensability of mathematics for these ambitions was firmly held, as the Jain scholar Mahāvīra elaborates:

> The configuration of living beings therein, the lengths of their lives, their progress, their staying together, that is, in other words, whatever there is in all the three worlds which consist of moving and non-moving beings, cannot exist apart from *gaṇita*.
>
> Mahāvīra[41]

The centrality of mathematics to the Jain tradition is also revealed by the fact that it was one of the four branches of their literature (*karaṇānuyoga*). Domains of applicability for the discipline were clearly established, and in a way that comes close to a philosophical contemplation of the discipline; mathematics could concern the "mundane" (applied, worldly, practical) or the "super-mundane" (metaphysical, ultimate reality, immaterial). Mathematics allowed these scholars to contemplate the observable as much as the non-observable.

Many critical Jain works are no longer extant, and of those that are, very few have been studied. However, we do know of several authors who give us a glimpse into the sophistication and scope of Jain mathematics. One is Vīrasena who wrote the *Dhavalā* in the beginning of the ninth century. Vīrasena was best known as a philosopher and a man of religious

[40] H. Diels and W. Kranz, *Die Fragmente der Vorsokratiker* (Berlin: Weidmann, 1956), 80b1.
[41] As cited from Jain, *Exact Sciences*, p. 9

importance,[42] not as an active mathematician. For this reason it is supposed that much of the material he includes in the *Dhavalā* is extracted from previous works which were composed from between the third and the seventh century CE. Of particular interest in this work is the early expression of place-value alignment, indices and exponentiation, fraction manipulation, the exploration of "logarithmic" type relations, and infinite processes. Eight analytical methods are expounded by Vīrasena to manipulate and develop certain relations.[43] These included *pramāṇa* (measure); *kāraṇa* (reason); *nirukti* (explanation); *vikalpa* (abstraction); *khaṇṛita* (cut); *bhājita* (division); *viralana* (spread); and *apahṛat* (removal).

Another important work was the *Trilokaprajñapti*, a text which deals with Jain cosmology. Mathematics in this account is central in this set up to describe "the order, the number, and the measurements of the islands, and oceans, the heavens, the hellish pits . . . the mountains and rivers of the universe."[44] Not only were the quantities of these features significant for numerical reasons, also too the objects within this cosmology were based on ideals in geometry and principles of symmetry.

Indeed, fundamental notions at the core of their concepts of space and time caused Jain thinkers to grapple with issues such as countability, numerability, and the infinite. Early characterizations of infinity described it as the number of grains of sand on the brinks of all rivers on the earth, or the drops of water in all of the oceans.[45] However, more sophisticated articulations were also advanced. Numbers could be "numerable," "innumerable," or "infinite." Their conception of infinity was further advanced to embrace "infinite in one direction," "infinite in two directions," "infinite in partial extent," "infinite in entire extent," and "eternally infinite"; some texts, such as the *Sthānāṅgasūtra* of 325 BCE, account for ten types of infinity. Recursive algorithms were developed to generate and conceive of these innumerable numbers, in a way that anticipates modern treatments of such subjects.

The religious platform and related cosmology also motivated questions about things such as the smallest unobservable unit. The *Tiloyapaṇṇattī* by Yativṛṣabha, which is dated to roughly the fifth century CE, is primarily a text to detail the cosmos, but it includes mathematical formula, decimal place value notation, and symbolic expressions for quantities dealing with numerate, innumerate, and infinite amounts, which are correlated to one another in complex relations. Concepts, such as the ultimate fraction of space (*pradeṣa*) and the ultimate particle (*parmāṇu*) are explored; the former is the quantity of space inhabited by the latter, which is indivisible, of one

42 Jain, *Exact Sciences*, p. 5.
43 Ibid., pp. 31–3.
44 T. A. Saraswati, *The Journal of the G. J. R. I., Allahabad, vol. 18: Nov. 1961–Aug 1962*, pp. 27–52. As cited from Jain, *Exact Sciences*, p. 7.
45 See the *Kalpasūtra* and the *Navatattva*, as cited from Jain, *Exact Sciences*, p. 12

taste, of one color, of one smell, but with two "touches." In a similar way, an instant is defined (*samaya*) as an instantaneous occurrence (*vikāra*) which is conceived in terms of a larger "flow" (*dravya*) of particles.

In the domain of "mundane mathematics," computation was classified into ten types, as in the *Sthānāṅgasūtra*:[46] "operation" (probably arithmetical); "procedure"; "rope" (geometry such as that found in the *Śulbasūtras*); "heap" (measurement of heaps of grain); "part-classes" (different operations with different types of fractions); "as much as so much" (multiplication or summation of numbers); "square" or "product"; "cube" or "solid"; "square-square" (fourth powers); "*kalpa*" (permutations and combinations). Many of these terms appear in later works on mathematics outside of the Jain context as standard divisions.

Jain religious commitments also directed them to consider various permutations and combinations. For instance, the *Bhagabatī sūtra* (ca. 300 CE) includes the analysis and computation of many different combinations, including the number of distinct philosophical categories which could be generated out of any *n* objects, taken one at a time, two at a time, and so forth.

The impact of Jain mathematical ruminations, occurring in the period between the flourishing of various Vedic mathematical ideas and the appearance of *Siddhānta*-style traditions, which appeared in around the fourth and fifth centuries CE, is little understood, and the varied impact it had more broadly on developing mathematical ideas is yet to be more clearly determined. More generally, we have clear manifestation of connections between mathematics and the divine that are both subtle and direct. But in this context it was largely the connection between arithmetic and the sacred. Was it the mystical powers of numbers and their mathematical relations and the hint of the eternal and unending that originally inspired and elaborated such cosmological and religious contemplation, or was it, rather, fundamental metaphysical contemplations that found substance in mathematics?

MATHEMATICS IN THE CLASSICAL AGE

The founding and flourishing of the Mauryan dynasty in the early fourth century BCE until the decline of the following Gupta dynasty in around 500 CE is often identified as the Indian "classical" period. This era is nearly synchronous with similar eras in empires in the west, and its cessation was due, as it was largely in the west, to disorder and debilitation caused by invasions from central Asia. Mathematical texts of this period are elaborate and long, and are usually parts of larger works on astronomy. Kinematic models which were fundamentally geometric that had been transmitted

[46] Plofker, *Mathematics in India*, pp. 59–60.

from Greece were imbued with existing Indian parameters and cosmology, and the result was a robust hybrid predictive model which was disseminated and developed in the *jyotiḥśāstra* tradition. Mathematical treatises, *gaṇita*, appear in larger works dedicated to astronomy, usually *siddhāntas*.

ĀRYABHAṬA

The earliest surviving *gaṇita* chapter, which is part of a larger astronomical work, is in the *Āryabhaṭīya* by Āryabhaṭa (b. 476 CE). This chapter is a highly abbreviated account of various mathematical topics; in just thirty-three verses, Āryabhaṭa covers basic arithmetic; plane and solid geometry; root and cube root extraction; procedures to compute interest and barter; trigonometry and a versified table of Sines; series; rule-of-three proportions; inverses; gnomon shadows; computation of unknowns; quadratic equations; and linear indeterminate equations.[47] Notably, although it is part of a large work on astronomy, a significant portion of the content of this chapter is relevant to areas outside of astronomy. It is earlier in this work that Āryabhaṭa outlined his system of numeration, the *varga* system (see above, section entitled "Numerals and Systems of Numeration").

One significant respect in which early mathematical astronomy as practiced in India differed from the Greek tradition was the use of the sine function, rather than the chord function that the Greeks had relied on. The sine function and the chord function are essentially equivalent, since the former can be generated from the latter by doubling the arcs and halving the result, via:

$$\frac{\text{crd } 2\theta}{2} = \sin \theta$$

Enumerations of sines, called *jyārdha* (lit. half-chord) were produced which set the standard for trigonometry to come. The terminology for such concepts seemed to have been inspired by the graphical representation which, by analogy, resembled a bow and arrow. In place of the archer's bow, mathematicians saw the arc, for the bow-string, the chord, half of this, the sine, and the arrow, the Versed Sine (see Table 2).

Practitioners selected different values for the radius, R, of their base circle,[48] where these early values corresponded to the modern via:

[47] For a comprehensive study of this work see A. Keller, *Expounding the Mathematical Seed: A Translation of Bhāskara I on the Mathematical Chapter of the Āryabhaṭīya*, 2 vols. (Basel: Birkhäuser, 2006).

[48] As opposed the modern convention which sets it at $R = 1$. Presumably these larger values were intended to generate higher integer accuracy.

Table 2: *Table showing the Sanskrit terms and their mathematical equivalents*

Sanskrit term	Literal Translation	Mathematical equivalent	
cāpa	Bow	Arc	BDC
jyā	Bow-String	Chord	BEC
jyārddha	Half the Bow-String	Sine	BE or EC
śara	Arrow	Versed Sine	ED

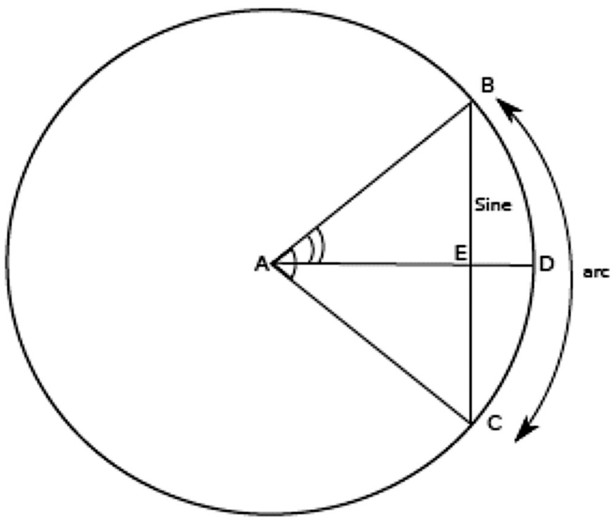

Figure 26.2. Various trigonometric components.

$$\text{Sin}\,\theta = R\sin\theta$$

Some of the selections for R by Indian scholars indicate that transmission from Greek sources may have taken place, such as Varāhamihira's $R = 120$ (which may be derived from the radius of Ptolemy's chord function which was 120 parts), Āryabhaṭa's $R = 3438$ (which results from measuring the circle in $60 \cdot 360 = 21{,}600$ minutes and may be derived from Hipparchus' value), and some which are novel, such as Brahmagupta's $R = 3270$ and $R = 150$. Like all other mathematical material, key trigonometric values, usually for intervals of $3\frac{3}{4}$, were enumerated in verse, and interpolation procedures for generating non-tabulated values accompanied these.

Āryabhaṭa enumerates not the values of the sines themselves, but rather the differences between successive sines, and he does so in the alphanumeric coding system he calls the *varga* system (see above, section entitled "Numerals

Table 3: *The numerical entries corresponding to Āryabhaṭa's verse on Sine differences and their relation to arcs and Sines*

θ	Sin^{θ}	Difference
3;45	225	**225**
7;30	449	**224**
11;15	671	**222**
15	890	**219**
18;45	1105	**215**
22;30	1315	**210**
26;15	1520	**205**
30	1719	**199**
33;45	1910	**191**
37;30	2093	**183**
41;15	2267	**174**
45	2431	164
48;45	2585	**154**
52;30	2728	**143**
56;15	2859	**131**
60	2978	**119**
63;45	3084	**106**
67;30	3177	**93**
71;15	3256	**79**
75	3321	**65**
78;45	3372	**51**
82;30	3409	**37**
86;15	3431	**22**
90	3438	**7**

and Systems of Numeration"). This may reflect the way in which he generated successive sines, by means of a difference method.[49] Another reason to enumerate the sine differences rather than the sines is that from the point of view of memorization, there are less actual digits to remember! While the sines themselves begin at three significant figures and increase up to four, the sine differences start at three significant figures and decrease down to one.

The ways in which these differences relate to the original angles and their sines is shown in Table 3 (Āryabhaṭa's values are shown in bold; recall R = 3438). In this context, the demands of an orality impressed itself on mathematics in a significant way. Given the restrictions of memory, the number of tabulated values that were enumerated was always limited, and, in order to compensate, ingenious interpolation techniques that would minimize the number of "tabulated" values necessary to recall were developed. For instance, Ptolemy's chord table lists a value for each half-degree

[49] For more details and a technical analysis, see, for instance, G. van Brummelen, *The Mathematics of the Heavens and the Earth. The Early History of Trigonometry* (Princeton, NJ: Princeton University Press, 2009), pp. 96–102, or Plofker, *Mathematics in India*, pp. 79–83 and 127–28.

arc to three significant sexagesimal places, 360 tabulated entries in all. By contrast, Āryabhaṭa's table contains twenty-four entries. Intermediary (non-tabulated) values are to be generated by interpolation (which incidently, Āryabhaṭa does not explicate). Many of Āryabhaṭa's successors give versified sine tables which are even more succinct.[50]

BRAHMAGUPTA

Flourishing slightly later than Āryabhaṭa was renowned scholar Brahmagupta (b. 598 CE), who also presented two chapters dedicated to mathematics in his large work on mathematical astronomy, the *Brahmasphuṭasiddhānta*, which contained twenty-six chapters in all. The twelfth chapter in this work covered mathematical basics in sixty-six verses and the other, chapter eighteen, is 101 verses long. What is interesting about Brahmagupta's presentation of mathematical material in two distinct sections is how he divided the content of each. The first chapter treats what Brahmagupta and later commentators identify as arithmetical operations, which include addition; subtraction; multiplication; division; squares; square roots; cubes; cube roots; reduction of fractions; rule-of-three proportions and extensions of these; barter and exchange; and procedures, which he lists as mixture, series, figures of geometry, volumes, piles (the size of stacked quantities), lumber computations, heaps (computing mounds of grain), and shadows of the gnomon. The second mathematical chapter contains descriptions of the so-called pulverizer,[51] as well as various procedures for determining unknowns, including second-degree indeterminate equations. The division of mathematical knowledge in this way reveals the emerging twofold classification of knowledge into *pāṭi-gaṇita* (lit. board-computation) and *bīja-gaṇita* (lit. seed computation) (see Śrīpati's division of this, above section entitled "A Context for Mathematics"). The former can loosely be described as "arithmetic," the "board" being emblematic of the medium on which such arithmetic calculations were carried out, and the second as "algebra," reckoning with unknown quantities (or seeds).[52]

This separation is revealing. That a distinction is made between mathematics that deals with known quantities and their manipulation, on the one hand, and mathematics that deals with unknowns, on the other, reveals that practitioners' attitudes towards mathematics were evolving.

[50] Compare, for instance, Brahmagupta (discussed in the next section) who gives a sine table with only six tabulated values. See *Brahmasphuṭasiddhānta* 25, 16.

[51] A certain procedure for determining unknown quantities using divisors and their remainders in the context of linear indeterminate equations.

[52] The analogy that is pertinent is presumably that one doesn't know what a seed will grow into, hence it is unknown.

By Brahmagupta's time these branches were named respectively as "manifest" and "unmanifest" mathematics. This distinction reveals an interest in a branch which was to later be called "algebra," but which is distanced in meaningful ways from what we would recognize as modern algebra today. This visual orientation (manifest/unmanifest) is fascinating and may hint at deeper cognitive modes of thought surrounding the status and scope of reckoning with unknowns.

Trigonometry is treated in an entirely separate chapter, in chapter 21 in seven verses, and actual sine values are given much earlier in chapter 2. Brahmagupta uses the idiosyncratic radius for his circle $R = 3270$. The origin of this value is still a mystery to historians of mathematics. These verses describe how to determine new sine values geometrically from those already known. Furthermore, that trigonometry was treated in an entirely different section is telling. This separation is done by other practitioners too, and reveals that trigonometry is a subject clearly in the domain of geometry but with direct association and applicability to astronomy rather than a topic considered part of general mathematics.

Brahmagupta, like his contemporaries, was sensitive to the distinction between "accurate" and "approximate" results, and gives computational rules to produce both, for instance when he computes the areas of triangles and quadrilaterals (*Brahmasphuṭasiddhānta* 12.21). Āryabhaṭa too gives a "practical rule" as well as a "rule without remainder" for the volume of a sphere. The discrimination of results on this basis is a sign of a sophisticated and experimental attitude towards computational procedures. It also reveals an approach to mathematics which was moderated by practical considerations, including ease of computation, a priority of simplifying assumptions over analytic completeness, precision of results, and the like.

POST-CLASSICAL MATHEMATICS

Despite the Gupta era coming to an end, distinct forms of economy, culture, and governance that had developed during this period persisted, and were not overturned until the decline of the Indian-Islamic rulers, almost twelve centuries later. Thus the appellation "medieval" to this period is a modern convention, done so out of convenience because it parallels trajectories in western cultures; by contrast, unlike other parts of the world, the social, political, and intellectual circumstances on the Indian subcontinent were characterized by a certain amount of continuity during this period. At the same time, there can be observed a changing structure of mathematical knowledge. Mathematics as a discipline emerged as a pursuit in its own right, and while astronomical treatises still remained an important context for mathematical development, treatises devoted exclusively to *gaṇita* began to appear.

BAKHSHĀLĪ MANUSCRIPT

After the seventh century, mathematical treatises started appearing as independent texts. One of the first of these (which is also the earliest surviving Sanskrit mathematical manuscript) is the so-called Bakhshālī manuscript, a mathematical treatise written on birch-bark.[53] This manuscript has been dated (by various means) to as early as the eighth century and no later than the twelfth, and stylistic features indicate that the work may be as old as the seventh century. This treatise contains passages in both verse and prose and outlines mathematical problems and algorithms. Numerals are represented by nine distinct gylphs, and zero is represented by a small round centered dot. This manuscript is interesting not only because of the selection of content it includes, but also owing to the layout and organization of worked problems and the notation used to do so. The manuscript exhibits a sophistication of mathematical expression and notation. Fractions are written vertically with the numerator on top and the denominator on the bottom, negative quantities were indicated by a small cross after the quantity (probably an abbreviation of the Sanskrit word *ṛṇaṃ* which means "negative" or "subtractive"), and many other words which expressed mathematical operations, such as add, subtract, and so on, were also abbreviated. Furthermore, the manuscript has a symbolic form to represent an unknown, a small dot (which was often confused with zero, although context would reveal which one was which). All of these features, combined with an aligned tabular type structure of presentation, mean that this mode of mathematical presentation can be considered proto-symbolic. Many of these notational conventions and layout conventions were adopted and used in mathematical manuscripts in successive centuries. The content of the manuscript is somewhat eclectic; material is not arranged into any obvious themes or structure. Furthermore, many of the situations are on specialized real-world problems, including payments, conversions, barter, and the composition of alloys.

The Bakhshālī manuscript has a distinctive exposition. Typically a rule (*sūtra*) is followed by a worked example (*udāharaṇa*). For instance, the following passage (in the course of solving a larger problem relating to an arithmetical progression) gives a rule for finding a square root of a number by an iterative technique.[54] In this particular case a square root is found ($20\frac{20}{21}$) to a number (481), but a closer approximation is desired (which turns

[53] See T. Hayashi, *The Bakhshālī Manuscript: An Ancient Indian Mathematical Treatise* (Groningen: Egbert Forsten, 1995).

[54] See ibid., pp. 326–7.

out to be $\frac{424642}{19362}$). The problem is set out, the rule is stated, then the procedure in the rule is followed, and the answer is given. The text reads:

> . . . The square root (obtained before) is 20 $\frac{20}{21}$. This is inaccurate. Therefore, (we recall the square root extraction *sūtra* again):
>
> The divisor for the remainder from the diminution of the non-square number by the square of the first approximation, is twice (the first approximation). Division of half the square of that by the second approximation (is made). Subtraction (of the result from the second approximation gives the third approximation). Less the square.
>
> (The first half of this rule), "the divisor for the remainder . . . is twice" has already been applied, (the result being 20 $\frac{20}{21}$). "The square of that":

21			"the 2nd approx"	21	:divisor
21	441	"half"	2		21

> The remainder (400) should be removed. When one has applied (the rule) . . . reduction to the same denominator is made:

$$\frac{425042}{19362}$$
$$-\frac{400}{19362}$$

Remainder:

$$\frac{424642}{19362}$$

The square root extraction algorithm described here appears to be a geometric equivalent of the modern "Newton-Raphson" method,[55] which was developed many centuries later in a different context. It works by making guestimations at the square root (thinking of this as the side of a geometric square), computing the "area" of a square with that side and comparing it with the area of the original square, and thus refining it by successive iterations. Such a procedure seems to have been used by Babylonian scribes in their four significant place computation of the square root of two found on a cuneiform tablet.[56] It is through this text

[55] In this case, it is based on a specific instance of this method, sometimes known as the Mechanics Rule for computing square roots.
[56] For details, see Montelle, "Roots, Rocks, and Newton Raphson."

that we get a firm sense of how problems were tackled in practice, giving insight into the ways in which practitioners approached problems. The careful use too of alignment in the setting out of the problem in addition to the statement of the rule gives insight into the ways in which a rhetorically expressed theoretical statement of a rule was recast into a more tractable means of solution-finding for a specific case. In the absence of a more symbolic expression, close observance to a generic way of setting out the solution procedure to the problem (which is maintained in several problems of the same type) is the closest the text comes to a general "formula" for the rule.

MAHĀVĪRA

A work called the *Gaṇitasārasaṅgraha* (lit. Epitome of the Essence of Computation) written by a Jain called Mahāvīra in the mid-ninth century is the first known independent mathematical treatise to have been handed down in its entirety. The work is enormous and treats various mathematical topics, in its own order and distribution, in more than 1,100 verses. Mathematics is expressed both in terms of rules and procedures, as well as posed problems, and, from time to time, prose sentences link verse.

The tone of much of Mahāvīra's work is highly playful and imaginative. Mathematical problems are staged in theatrical and creative scenarios, designed to delight, amuse, and capture the attention of the reader. Mathematics here is seen as a public pursuit, an amusing pastime, designed for pleasure and entertainment, and Mahāvīra's work is simultaneously educational and recreational.

As well as accounting for the basic operations of arithmetic, Mahāvīra clarifies operations with zero, as well as the arithmetic of positive and negative quantities. Some of the content and format of his problems are reminiscent of those in other cultures, such as his "purse problems,"[57] which are similar in scope to those seen in Fibonacci.[58]

In the introductory sections of this work, when outlining his conventions and metrology, Mahāvīra tackles the topic of space and time, and their smallest divisions. He portrays these as "infinitesimal" quantities; the smallest conceivable length he states is the size of an "indestructible quantity of substance" or "ultimate atom."[59] The smallest amount of time, an instant, is described in terms of the time it takes for one atom to move past another, and the resulting time unit is defined as an innumerable instant. These considerations are reminiscent of those explored by Jain thinkers and allude

[57] A scenario in which a purse or purses are found, and, given certain stipulations, the amount in the purse is to be determined.
[58] See *Liber Abaci* section 4, chapter 12.
[59] See Plofker, *Mathematics in India*, p. 163.

to the rare instances in India of mathematical rumination addressing deeper metaphysical issues.

CONCLUDING REMARKS

The tradition of mathematics in India is almost as old as literate culture itself, and traces of mathematical thinking can be found in the earliest texts, and even before, in archaeological remains. The challenges of characterizing and evaluating this tradition have impressed themselves on later thinkers in ways that are both direct and nuanced. To be sure, mathematics in India did not evolve completely in isolation, and was influenced by the ideas exchanged through contact with other early civilizations. However, cultural circumstances and constraints on intellectual activity, as well as broader religious commitments, meant that the practice and nature of mathematics remained distinct. Despite the fact that mathematics remained a technical subject, in some contexts its reach extended to the contemplation of sacred, supernal, and deeply metaphysical topics. True to its characterization, it embraced an "ocean" of topics, and unlike other cultures of inquiry in antiquity whose organization and presentation were strictly delimited by canonical texts, it embraced diversity and pluralism. Mathematics served simultaneously as an instrument of religious expression, a fundamental element in the cosmic narrative, the foundation of astrology, and a social regulator. Beyond a field in the service of religion though, mathematics was deemed perfect for concepts and ideas that were otherworldly, and from the earliest times it was intimately connected with ritual and sacred cosmology.

Mathematical practice in India is embedded in its socio-political circumstances and function. It has diverse, pluralistic demands which are nurtured and maintained in very different contexts and pressures. More broadly, the portrayal of multiple approaches to a subject often deemed universalistic directly conveys the ways in which different early peoples responded to and approached similar challenges that were context-dependent. Thus, our survey of but a few episodes from the early history of mathematics in India reveals that even though these works are united by modes of expression, objects of investigation, and methods, other key aspects such as sphere of applicability, organization of content, scope, and pitch differed extensively. On the one hand, this is a pointed reminder that the focus of inquiry is one that is in many respects tenuously unified, but on the other, that a mathematical climate free from fixed standard notions of scope, structure, and practice nevertheless nurtured and produced some formidable and profound achievements.

27

INDIAN MEDICINE AND AYURVEDA

Philipp A. Maas

INTRODUCTION

The cultural, intellectual, and religious heritage of South Asia comprises medical systems that offer themselves as supplements or alternatives to modern bio medicine. Some of these, namely Homoeopathy, Naturopathy, Yoga, Siddha, Sowa-Rigpa, Unani, and Ayurveda, receive promotion from the Ministry of Ayurveda, Yoga and Naturopathy, Unani, Siddha and Homoeopathy (AYUSH) of the Government of India, which defines among its tasks to "upgrade the educational standards of Indian Systems of Medicines . . . [, and to] evolve Pharmacopoeial standards for Indian Systems of Medicine and Homoeopathy drugs."[1]

This agenda betrays a conceptional influence of modern bio medicine and its theoretical foundations on the self-conception of AYUSH as a representative of Indian medicine in so far as it can be read as an implicit response to a critique, according to which complementary and alternative medicines (CAMs) are of unequal educational standards, and apply non-standardized medicinal drugs, the efficacy of which was never proved with standardized tests, etc. In taking up this criticism, the Ministry of AYUSH shows a far-reaching readiness to accept modern bio medicine as the scale against which CAMs are to be measured, possibly in order to position itself in the ongoing debate of whether the CAMs of South Asia are based on, or comply with, "science" in a modern academic understanding of the word, or whether they should be labeled as "pseudoscience." A final conclusion of this discussion, which has to be led from a multi-disciplinary perspective in which practicing scientists, practitioners of CAMs, philosophers of science, social anthropologists, and Indologists may participate, is not to be expected for the near future, and even a summary of the main arguments of the different camps falls outside the scope of the present chapter.

[1] http://ayush.gov.in/about-us/background (accessed on 11 April 2017).

Although the Indian Ministry formulated its program vaguely, it leaves no doubt that it regarded all the medical systems listed above as "Indian medicine," presumably because these CAMs are nowadays widely practiced in India and have a written textual tradition. It could be argued, however, that this categorization is too narrow. It excludes, for example, Indian folk medicine, shamanism, astrology, and faith healing, because these are partly based on oral traditions.

It could also be argued that the above outlined categorization of "Indian medicine" is too wide, because not all of the medical systems practiced today in India were developed in South Asia, or because they are not practiced in all parts of modern India.

Homoeopathy, to start with, was invented by the German physician C. F. Samuel Hahnemann at the end of the eighteenth century. It was effectively introduced in South Asia by the Transylvanian physician Johann Martin Honigberger, who worked at the court of Ranjit Singh in Lahore from 1829 to 1833. From this time onwards, homoeopathy met with great success in South Asia, and it is nowadays widely practiced in India.[2]

Naturopathy, which was also developed in Europe, but about a hundred years later than homoeopathy, became prominent in modern South Asia not only through the translation of the foundational works of Louis Kuhne[3] and others into English, but most of all through Mahatma Gandhi's strong advocacy of Nature Cure.[4] The similarity of a number of the basic conceptions of naturopathy with conceptions of modern Yoga led to both systems becoming "completely integrated, at least from the vantage point of practitioners of Nature Cure and the government of India, if not, by any means, all practitioners of Yoga."[5]

As Mark Singleton has shown recently, modern yoga developed from a blend of the views of European bodybuilding and gymnastic movements with Indian nationalism and political Hinduism as well as from obscure indigenous yoga traditions.[6] These political conceptions oscillated between a fascination for modern western ideas and a self-affirming appraisal of the Hindu religious and philosophical traditions, among which the philosophical ideas of Advaita Vedānta played a prominent role. The association of yoga with modern medical science, for which there is evidence as early as

[2] See Raekha Prasad, "Homoeopathy Booming in India," *The Lancet* 370.9600 (November 2007), 1679–80.

[3] See, for example, Louis Kuhne, *Die neue Heilwissenschaft oder die Lehre von der Einheit aller Krankheiten und deren darauf begründete einheitliche, arzneilose und operationslose Heilung. Ein Lehrbuch und Ratgeber für Gesunde und Kranke* (2nd enlarged edn; Leipzig: Kuhne, 1891).

[4] See Joseph S. Alter, *Yoga in Modern India. The Body Between Science and Philosophy* (Princeton, NJ: Princeton University Press, 2004), pp. 109 f.

[5] Ibid., p. 110.

[6] Mark Singleton, *Yoga Body. The Origins of Modern Posture Practice* (Oxford: Oxford University Press, 2010).

1889,[7] can be seen as an expression of the self-affirmation of Indian intellectuals against the ruling British colonial power, which tended to present the quickly developing bio medicine as a sign of the general superiority of British or western culture.

Siddha medicine is a popular medical system in the south Indian sate of Tamil Nadu.[8] On a theoretical level, it combines tantric religious ideas with basic conceptions of alchemy and classical Ayurveda (see section "Classical Ayurveda," below). Its textual basis consists of a textual corpus composed in the Tamil language, of which the majority of works exist only in unpublished palm-leaf manuscripts, and even the primary sources of Siddha medicine that have been published in print so far are not yet sufficiently researched. It appears, however, that the oldest strata of Tamil Siddha literature were composed in the sixteenth century CE. The tradition itself claims, however, to be of a much earlier date.[9]

Sowa-Rigpa (gSo ba rig pa), or Tibetan medicine,[10] is in its various forms widely practiced in Central Asian and Himalayan regions, some of which are nowadays part of the Republic of India. As a consequence of the migration of Tibetan refugees to India and other parts of the world, gSo ba rig pa, which in Tibetan means "authoritative knowledge of healing," also became popular in regions of India that are located south of the Himalayas and in globalized societies outside Central and South Asia. The theoretical foundations of Tibetan medicine developed from the seventh century CE onwards from a combination of indigenous Tibetan medical knowledge with classical Chinese and ayurvedic medicine (see section "Classical Ayurveda," below) as well as with various Buddhist and alchemical sources.

Unanani Tibb is prominent throughout South Asia. It was introduced to the Indian subcontinent in the Middle Ages (from the early eleventh century onwards) in the context of Islamic invasions.[11] The very name of this medical system, which means "Medicine of the Greek," indicates that this CAM developed from translations of the Greek works of the Hippocratic School as well as of the Latin works of Galen into the literary languages of Islam. In the course of its history, Unanani Tibb was further developed by a number of medical authorities, among whom Ibn Sina figured most prominently.

[7] Ibid., p. 50

[8] See Richard S. Weiss, *Recipes for Immortality. Medicine, Religion, and Community in South India* (Oxford: Oxford University Press, 2009).

[9] See Hartmut Scharfe, "The Doctrine of the Three Humors in Traditional Indian Medicine and the Alleged Antiquity of Tamil Siddha Medicine," *Journal of the American Oriental Society* 119.4 (1999), pp. 609–29.

[10] For a detailed overview of various aspects of Tibetan Medicine see Fernand Meyer, "Theory and Practice of Tibetan Medicine," in Anthony Aris and Jan van Alphen (eds.), *Oriental Medicine. An Illustrated Guide to the Asian Arts of Healing* (London: Serindia Publications, 1995), pp. 109–41.

[11] For a more detailed overview of Unani Medicine in South Asia see Claudia Liebeskind, "Unani Medicine of the Subcontinent," in Aris and van Alphen (eds.), *Oriental Medicine*, pp. 39–65.

Any argument that the before mentioned CAMs are not actually "Indian Medicine," because they are not of South Asian origin, or because they are only practiced regionally, would result from an essentialization of the attribute "Indian" that can hardly be justified on a theoretical level. Any definition of "Indian medicine" should do justice to the cultural, intellectual, and religious plurality of Indian medical theories, beliefs, and practices by being as comprehensive as possible.

Among the medical systems of South Asia there is, however, a single CAM with a long (pre-)history in South Asia that can be largely reconstructed from written sources. This is the medical system designated with the Sanskrit word *āyurveda*, i.e. "knowledge of longevity." It is this CAM, as well as its historical predecessors and developments, on which the following sections of the present chapter will focus.

PREHISTORIC MEDICINE IN SOUTH ASIA

In the absence of written records, all information about the pre-history of South Asia has to be gathered from archaeological findings, the earliest of which are over two million years old. This archaeological evidence suggests a long period of nomadic settlement in the northwestern part of South Asia.[12] The most ancient indication for permanent settlements in this area are the remains of mud-brick houses that occur together with indications of agriculture at ca. 6500 BCE in Mehrgarh in modern Pakistan. The archaeological findings from Mehrgarh do not allow for any conclusions on whether or not the religion of Mehrgarh included the conception of religious healing. However, the people of Mehrgarh apparently practiced an early form of dentistry already 9,000 to 7,500 years ago.[13] This suggests that, probably, other medical practices were also employed in this early phase of South Asian medicine.

The following stage of South Asian cultural history is represented by an early high culture, the Indus Valley civilization, which developed in the region of today's Pakistan and western India from about 3000 BCE and reached a cultural peak between 2150 and 1750 BCE. From ca. 1900 BCE onward, this civilization declined. The reasons for this development were presumably internal factors, reinforced by environmental changes that led to a shift of the course of rivers.[14]

As in the case of the prehistoric settlement of South Asia, nothing definite can be said about the religion of the Indus Valley civilization due to the

[12] Jonathan M. Kenoyer, *Ancient Cities of the Indus Valley Civilization* (Karachi and Islamabad: Oxford University Press; American Institute of Pakistan Studies, 1998), p. 33.

[13] See A. Coppa, A. Cucina, D. W. Frayer, C. Jarrige, J.-F. Jarrige, G. Quivron, M. Rossi, M. Vidale, and R. Macchiarelli, "Early Neolithic Tradition of Dentistry," *Nature* 440.6 (April 2006), 755–6.

[14] Kenoyer, *Ancient Cities*, p. 173.

absence of intelligible written sources. In the archaeological records, numerous seals depicting human beings or anthropomorphic deities stand out. Quite a number of artifacts – among them many steatite seals – bear symbols similar to a script. All attempts to decipher these symbols consistently have failed so far, and it has been suggested that they do not constitute a script that was ever meant to record a natural language.[15] There is also virtually no information available on medical beliefs, theories, and practices. However, some of the bronze razors, pins, and pincers that were found "must have," according to Kenoyer, "been the tools of a barber or a physician."[16]

This meager archaeological evidence for medical beliefs, theories, and practices in pre-historic South Asia hardly justifies a treatment of this phase of South Asian cultural history within a historical overview on Ayurveda, the medical concepts of which originate from the intellectual environment of a much later time.[17] Although this can hardly be disputed with historical arguments, we find the anachronistic claim in some currents of modern Ayurveda that Ayurveda originated in the peak period of the Indus Valley civilization. The reason for this claim is the equation of antiquity with authenticity, on which some modern forms of Ayurveda draw to create acceptance for their CAM in the globalized world.[18]

VEDIC MEDICINE

The next phase of South Asian cultural history began from about 1750 BCE, when nomad tribes speaking the Indo-European language of Vedic immigrated in successive currents to the northwestern part of the Indian subcontinent. These tribes, as well as acculturated groups, shared some civilizing accomplishments, a common language, and a set of common religious beliefs. The Vedic religion was a polytheism, in which personified powers of nature and ethical principles played an important role. Among these, the twin gods named Aśvin-s are particularly connected with providing remedies in distressing situations of life, and, accordingly, they also function as physicians of humans and the gods.[19]

The most important religious practices documented in the early Vedic literature are sacrificial rituals. These were performed to praise and to feed the gods as a sign of gratitude for their support, and in order to make them favorably disposed toward one's own clan. In the course of time, sacrificing

[15] Steve Farmer, Richard Sproat, and Michael Witzel, "The Collapse of the Indus-Script Thesis: The Myth of a Literate Harappan Civilization," *Electronic Journal of Vedic Studies* 11.2 (2004), 19–57.
[16] Kenoyer, *Ancient Cities*, p. 128.
[17] See also Jean Filliozat, *The Classical Doctrine of Indian Medicine: Its Origins and Its Greek Parallels* (Delhi: Munshiram Manoharlal, 1964), p. 187.
[18] Kenneth G. Zysk, "New Age Ayurveda or What Happens to Indian Medicine When It Comes To America," *Traditional South Asian Medicine* 6 (2001), 10–26, p. 23.
[19] See Filliozat, *The Classical Doctrine*, pp. 86–91.

became an increasingly complicated matter that had to be performed by specialists, the Brāhmaṇa priests. From the middle Vedic period of ca. 1200–800 BCE onwards, sacrifices were seen as quasi-mechanical tools with which the sponsor of a sacrifice could accomplish for himself desirable results like victory in battle, wealth in procreation, cattle and horses, well-being in this world, and the attainment of heaven after death.

The Vedic literature consists basically of four collections of texts called *Ṛg-*, *Sāma-*, *Yajur-*, and *Atharvaveda*, which respectively mean "knowledge (*veda*) of the sacred hymns (*ṛc*), melodies (*sāman*), sacrificial formulas (*yajus*), and spells (*atharvan*)."[20] Each of these collections comprises the three different text-types of (1) *saṃhitā*-s, containing mostly metrical hymns for use in sacrificial ceremonies (dateable to ca. 1750–1200 BCE), (2) *brāhmaṇa*-s, consisting mainly of interpretations of the sacrificial mechanics (datable to ca. 1200–850 BCE), and (3) *āraṇyaka*-s and *upaniṣad*-s that are either quite similar in content to the *brāhmaṇa*-s, or they contain religious-philosophical speculations (datable to ca. 850–500 BCE, with many works being several centuries later).

The textual material that is pertinent to Vedic medicine is mainly contained in the *saṃhitā*-s of the *Atharvaveda* and, to a much lesser extent, of the *Ṛgveda*.[21] From these texts, it appears that the anatomical knowledge of Vedic India resulted basically from chance observation during horse and human sacrifices, which led to the composition of lists of parts of the bodies of horse and of man being preserved in the *brāhmaṇa*-texts.[22] The actual medical practice of Vedic India can be characterized as being essentially of a magico-religious nature. This means that

> [c]auses of disease are not attributed to physiological functions, but rather to external beings or forces of a demonic nature, who enter the body of the victim and produce sickness. The removal of such malevolent entities usually involved an elaborate ritual ... The principal figure in the rite was the healer (*bhiṣáj*) ...[23]

Within rituals, the Vedic healers employed medicinal plants and other substances. These were classified according to their habitat and their morphological features and were either locally collected or acquired by trade. For their medical application in ritual contexts, the substances were either

[20] For more details, see Stephanie W. Jamison and Michael Witzel, "Vedic Hinduism," in Arvind Sharma (ed.), *The Study of Hinduism* (Columbia, SC: University of South Carolina Press, 2003), pp. 65–113.

[21] Kenneth G. Zysk, *Religious Healing in the Veda. With Translations and Annotations of Medical Hymns from the "Ṛgveda" and the "Atharvaveda" and Renderings from the Corresponding Ritual Texts* (Philadelphia, PA: American Philosophical Society, 1985), p. 7. Ibid., p. 5.

[22] Ibid., p. 7.

[23] Ibid., pp. 7 f.

processed into medicines that were to be drunk in a solution or they were fashioned into amulets or talismans.[24]

In spite of the similarity in names, Ayurveda is not a successor of Vedic medicine. This becomes evident from the fact that basic theoretical concepts of ayurvedic medicine are not mentioned in Vedic literature. Moreover, the early mythological depictions of the Aśvin-s as twin physician-gods indicate that the medical profession was not highly valued in the Vedic milieu. On the contrary, physicians were regarded as ritually impure and excluded from Brahmanical rituals, because of their occupation with ritually impure substances. Therefore medicine was not practiced widely, if at all, among the members of the three higher classes of the Vedic societies.[25]

THE MEDICINE OF THE *ŚRAMAṆA*-MOVEMENTS AND EARLY BUDDHISM

A religious complex that was different from and largely independent of the Vedic religion developed at the time of the second urbanization of South Asia around 500 BCE in the eastern part of the Gangetic plain. This complex consisted of the so-called *śramaṇa*- or ascetic religions of Greater Magadha, which were the ancestors of the Ājīvikism, Jainism, and Buddhism. The early *śramaṇa*-religions shared a number of common conceptions that were alien to the Vedic religion, such as, for example, the notion of cyclical time and the idea of karmic retribution of actions happening in different realms of rebirth.

This intellectual and religious environment is also the home of medical theories and practices that developed into Ayurveda,[26] as can be concluded from the fact that early Buddhist works reflect medical conceptions similar to those of Ayurveda, whereas the Vedic text corpus does not reveal comparable ideas. For example, the sermon of the Buddha on the cultivation of mindfulness (*Satipaṭṭhānasuttam*) of the Pali-canon contains the following description of the human organism in the context of a meditation meant to prevent Buddhist monks from identifying with their own bodies:

> Again, O monks, a monk contemplates on this same body . . . as being up to the skin full of many kinds of ugly impurity: "In this body occur hair of the head, hair of the body, nails, teeth, skin, muscle flesh, sinew, bone, bone marrow, the kidney, heart, liver, pleura, spleen, lungs, bowels, mesentery, stomach, excrement, bile, phlegm, pus,

[24] On the usage of herbs for Vedic healing, see ibid., pp. 96–9.

[25] See Kenneth G. Zysk, *Asceticism and Healing in Ancient India. Medicine in the Buddhist Monastery* (Delhi: Motilal Banarsidass, 1998), pp. 22–4 and Patrick Olivelle, "The Medical Profession in Ancient India: Its Social, Religious, and Legal Status," *eJournal of Indian Medicine* 9 (2017), 1–21.

[26] See also Johannes Bronkhorst, *Greater Magadha. Studies in the Culture of Early India* (Leiden and Boston, MA: Brill, 2007), pp. 56–60.

> blood, sweat, fat, tears, grease, saliva, snot, serous fluid, and urine" . . . Again, O monks, a monk reviews this same body . . . as thus consisting of elements: "In this body there are the earth element, the water element, the fire element, and the air element." Just as though a skilled butcher or his apprentice had killed a cow and was seated at the crossroad with it cut into pieces; so too, a monk reviews this same body . . . as consisting of elements . . .[27]

This passage reveals that – possibly as early as in the fourth century BCE – its author, who for believing Buddhists was the Buddha himself, had a quite advanced anatomical knowledge and that his attitude towards corpses was unaffected by worries about ritual pollution. In his final analysis, the author viewed the human body as consisting simply of the four elements of matter. In this way, he exhibits the *weltanschauung* of *śramaṇa*-physicians, which is, according to Zysk, empirico-rational. This means that unlike the Vedic practitioners, the *śramaṇa*-physicians emphasized direct observations, systematized the acquired data, and analyzed them in a rational way that lead to the development of theories about the nature of health and the causes of diseases.[28]

Moreover, in a sermon from the collection of "Connected Discourses of the Buddha," again from the Pali-canon,[29] the Buddha answers the question of the non-Buddhist renouncer Sīvaka of whether the view of some renouncers is true, according to which all pleasurable, painful, or neutral human experiences are instances of karmic retribution in the following way:

> Sīvaka, some feelings indeed arise as being caused by bile. Sīvaka, one can know by oneself that here some feelings indeed arise as being caused by bile. Also the world regards it as a truth that here some feelings arise as being caused by bile. In this regard, Sīvaka, the non-Brahmanical and Brahmanical renouncers who proclaim and who believe that whatever a human feels, whether it is pleasure or pain or neutral, is caused by the [ethical value of] actions, these renouncers contradict what they themselves have understood, and they contradict what is considered as a truth in the world. Therefore I say that the view of these non-Brahmanical and Brahmanical renouncers is wrong.

Then, the Buddha repeats the same wording for seven other possible causes of the different kinds of feelings: phlegm, wind, a combination of the three before mentioned substances, the changes of the seasons of the year,

[27] Bhikkhu Ñāṇamoḷi and Bhikkhu Bodhi, *The Middle Length Discourses of the Buddha: A New Translation of the Majjhima Nikaya* (Boston, MA: Wisdom Publications, 1995), pp. 147 f., slightly modified.

[28] See Zysk, *Asceticism and Healing*, p. 29.

[29] M. Leon Feer (ed.), *Saṃyuttanikāya. Part 4, Saḷāyatana-Vagga* (London: The Pali Text Society, 1894), pp. 230 f.

unsuitable care,[30] acts of violence, and, finally, karmic retribution. Accordingly, karma is just one factor out of eight that cause well-being or otherwise of humans. Most of the first seven factors can be influenced by human beings with suitable knowledge, as for example physicians. It was presumably this attitude, according to which human suffering and disease are not exclusively the outcome of former ethically objectionable actions on the side of the suffering individual that supported the development of medical knowledge in Buddhist circles.

Bile, phlegm, and wind, the first three causes for human sensations, which occurred also in the previously cited list of bodily constituents, play a prominent role also in the etiology of classical Ayurveda. There, the same words designate the corruptions or humors (*doṣa*) that determine individually or collectively the basic constitution and the degree of health of human beings. As was shown by Hartmut Scharfe, the mentioning of these substances in early Buddhist literature does not, however, justify the conclusion that classical ayurvedic theories were already current at the time of the Buddha.[31] In fact, it almost took one thousand years before a multitude of similar but partly conflicting theories developed into the more or less standardized corpus of conceptions that characterizes classical Ayurveda from the works of Vāgbhaṭa (seventh century CE) onwards (see section "Classical Ayurveda," p. 541).

The sources for tracing this development are not too rich with regard to the earlier phases. Besides passages from early Buddhist canons, there is the possibly oldest completely transmitted medical work in Sanskrit, the sixteenth chapter of the Mahayana Buddhist *Suvarṇaprabhāsasūtra* ("Sūtra of Golden Radiance") that was translated into Chinese between 416 and 421 CE.[32] Ancient fragments of medical texts are the Qizil fragment (written ca. 200 CE on leather) and the Bower manuscript (written ca. 525 CE on birchbark).[33]

CLASSICAL AYURVEDA

The main sources of Ayurveda in general are large compendia written in Sanskrit. The study of these works is still in its infancy, because neither critical editions nor annotated scholarly translations that could serve as the

[30] This interpretation of the two last mentioned Pali terms follows Richard F. Gombrich, *What the Buddha Thought* (London and Oakville, CT: Equinox, 2009), p. 20.

[31] Scharfe, "The Doctrine of the Three Humors," p. 615.

[32] See Johannes Nobel, *Ein alter medizinischer Sanskrit-Text und seine Deutung* (Baltimore, MD: American Oriental Society, 1951).

[33] Lore Sander, "Origin and Date of the Bower Manuscript. A New Approach," in Marianne Yaldiz and Wibke Lobo (eds.), *Investigating Indian Art* (Berlin: Staatliches Museum Preußischer Kulturbesitz, 1987), pp. 313–23, p. 321b.

basis for further studies have yet been published.[34] A monumental research tool on many works and aspects of Ayurveda and their treatment in secondary literature is Gerrit Jan Meulenbeld's *History of Indian Medical Literature*,[35] whereas Dominik Wujastyk's *The Roots of Ayurveda* contains a recommendable introduction to the topic and a selection of nicely readable translations from original sources.[36]

The sources of classical Ayurveda are mainly six text collections in the Sanskrit language (and the rich tradition of commentaries on these works) that bear titles referring to the names of their respective compilator-authors, namely the Compendium of Caraka (*Carakasaṃhitā*); the Compendium of Suśruta (*Suśrutasaṃhitā*, which was composed a short time after the *Carakasaṃhitā*); the Compendium of Bhela or Bheḍa (*Bhela-* or *Bheḍasaṃhitā*, datable to the time span between ca. 400 and 750 CE);[37] the Compendium of Kaśyapa or Kāśyapa (*Kāśyapasaṃhitā*, ca. fourth to sixth century CE);[38] the Compendium being the Heart of the Eightfold Science [of Ayurveda] (*Aṣṭāṅgahṛdayasaṃhitā*) of Vāgbhaṭa;[39] and the Summary of the the Eightfold Science (*Aṣṭāṅgasaṃgraha*),[40] which is also attributed to an author named Vāgbhaṭa.[41]

Of these, the Compendium of Suśruta is famous for its section on surgery,[42] which depicts this branch of medicine in a far more advanced and professionalized stage of development than previous as well as later ayurvedic sources. Apparently, surgery "ceased to be part of the professional practice of traditional physicians" and "migrated to practitioners of the 'barber-surgeon' type"[43] shortly after the composition of the Suśruta's work, probably in the second century CE.

The most influential work in the history of ayurvedic medicine is Vāgabaṭa's "Heart of the Eightfold Science," which is a summary of previous works, such as the compendia of Caraka, Suśruta, and Bhela, that was composed at the beginning of the seventh century. Vāgabaṭa successfully created a standardized form of Ayurveda in his well-organized and struc-

[34] A critical edition and annotated English translation of a part of the oldest of the treatises, the *Carakasaṃhitā*, is currently under preparation by a research team at the Univerity of Vienna, which is directed by Karin Preisendanz. See www.istb.univie.ac.at/caraka/ (accessed on 11 May 2018).

[35] Gerrit Jan Meulenbeld, *A History of Indian Medical Literature*, 3 vols. (in 5 parts) (Groningen: Forsten, 1999–2002).

[36] Dominik Wujastyk, *The Roots of Ayurveda. Selections from Sanskrit Medical Writings. Translated with an Introduction and Notes* (3rd edn; London and New York: Penguin Books, 2003).

[37] For the date of Bhela's compendium, see Meulenbeld, *Indian Medical Literature*, vol. 2.A, 22–4.

[38] See ibid., pp. 25–41.

[39] See Meulenbeld, *Indian Medical Literature*, vol. 1.A, 391–474.

[40] See ibid., pp. 475–594.

[41] See ibid., pp. 595–686.

[42] For different dates assigned to this work, see the *Suśratasaṃhitā* pp. 342–4.

[43] Wujastyk, *The Roots of Ayurveda*, p. 66.

tured metrical composition.[44] His work was – and still is – memorized by medical students all over South Asia, especially in the modern State of Kerala, and it exists in written form not only in numerous printed editions but also in thousands of unpublished manuscripts. The influence of this monumental work even extended beyond the boundaries of South Asia to Tibet, where the Tibetan translation of the *Aṣṭāṅgahṛdayasaṃhitā* from the time of 1013–55 contributed to the development of Tibetan medicine (gSo ba rig pa, on which see section "Introduction", above).[45]

The oldest of the classical ayurvedic compendia is probably that of Caraka, which is usually dated to a time span of between 100 BCE and 200 CE,[46] but assuming a date of composition around the year 50 CE may be the best educated guess. Caraka's work was revised and supplemented by a redactor named Dṛḍhabala at some time between 300 and 500 CE.[47] In the course of its transmission in manuscripts, this version of Caraka's work, consisting of 120 chapters (*adhyāya*) in eight books (*sthāna*), developed into the multiple versions that are today current in manuscripts and printed editions.[48] The eight books of the *Carakasaṃhitā* are entitled as follows: the book of stanzas, or the book on the essentials of medical knowledge (*Śloka-* or *Sūtrasthāna*); on diagnosis (*Nidānasthāna*); on precise judgment (*Vimānasthāna*); on what is related to the body (*Śārīrasthāna*); on the omens for death (*Indriyasthāna*); on therapy (*Cikitsāsthāna*); on pharmacy (*Kalpasthāna*); and on medical success (*Siddhisthāna*).[49]

The number of books in the *Carakasaṃhitā* agrees with the number of areas of medical knowledge, although the medical branches differ thematically. They are: internal medicine; medical treatment of the supraclavicular region; the extraction of foreign bodies (such as arrows); treatment of intoxication; demonology; treatment of women during pregnancy and thereafter as well as pediatrics; life extension; and aphrodisiacs.[50]

At the time when the *Carakasaṃhitā* was composed, medical practitioners had affiliated themselves with early Hinduism to such a degree that members of the first three classes of society – i.e. Brāhmaṇa-s, the warrior nobility (*kṣatriya*-s), and the class of free working men (*vaiśya*-s) – were admitted to

[44] On the theoretical foundation of classical Ayurveda from Vāgbhaṭa onwards, see Julius Jolly, *Medicin* (Strassbourg: Trübner, 1901), pp. 39–42.

[45] See Claus Vogel, *Vāgbhaṭa's Aṣṭāṅgahṛdayasaṃhitā. The First Five Chapters of its Tibetan Version* (Mainz: Deutsche Morgenländische Gesellschaft and Wiesbaden: Steiner (in Kommission), 1965), p. 21.

[46] According to Meulenbeld, *Indian Medical Literature*, vol. 1.A, 114.

[47] According to ibid., p. 141.

[48] For a more detailed account of the structure of the *Carakasaṃhitā* see ibid., pp. 93 f. On the later textual history of the *Carakasaṃhitā* see Philipp A. Maas, "On What Became of the Carakasaṃhitā after Dṛḍhabala's Revision," *eJournal of Indian Medicine* 3.1 (2010), 1–22.

[49] See Meulenbeld, *Indian Medical Literature*, vol. 1.A, 7–92.

[50] According to *Carakasaṃhitā* Sūtrasthāna 30.28 in Jādavji Trikamjī Ācārya (ed.), *Caraka Saṃhitā by Agniveśa. Revised by Caraka and Dṛḍhabala. With the Āyurveda-Dīpikā Commentary of Cakrapāṇidatta* (repr. of the Bombay 1941 edn; Varanasi: Krishnadas Academy, 2000), p. 189a. See also Meulenbeld, *Indian Medical Literature*, vol. 1.A, 26.

practice Ayurveda.[51] This affiliation was successful to such a degree that the memory of Ayurveda's origin in the milieu of the *śramaṇa*-religions was completely lost in the medical tradition.

Additional strategies for securing acceptance in a society committed to Brahmanical norms were, for example, using Sanskrit as the medium of codifying medical knowledge, modeling the initiation into medical studentship in accordance with Vedic rituals,[52] and establishing a relation between Ayurveda and the *Atharvaveda* (see section "Vedic Medicine"). Moreover, the early medical authorities traced the origin of ayurvedic knowledge in origination myths to late Vedic and Vedic gods. According to the account of the *Carakasaṃhitā*, the sequence of the transmission of ayurvedic medicine began with the all-knowing god Brahmā, who imparted it via the gods Prajāpati, the Aśvin-s, and Indra to the human seer Bhāradvāja, who then instructed Ātreya Punarvasu. This seer taught a group of six medical authorities, each of whom composed his own medical treatise. One of these works is entitled "The authoritative teaching of Agniveśa" (*Agniveśatantra*). This work, of which nothing else is known, was later – and this appears to be historically reliable information – revised by a redactor named Caraka into the *Carakasaṃhitā*.

According to this mythological narrative, Ayurveda is of divine origin. This reflects the basic attitude of the early classical medical authorities towards their medical knowledge as being in principle perfect. Ayurveda, just like virtually all other systems of authoritative knowledge (*śāstra*) in premodern South Asia, is regarded as being beyond any need or capacity of improvement by means of "the discovery of what has never been known before." All that the authorities can aspire to is a "recovery of what was known in full in the past."[53] However, this search for the recovery of primordial perfection left room for developments that from a historical perspective appear as discoveries and innovations in ayurvedic theory and practice throughout its history.

Medical innovations and discoveries benefited from the fact that Ayurvedic medicine is to a considerable degree based on observations and committed to rationality. Direct observation (*parikṣā*), reasoning (*yukti*), and inference (*anumāna*) were considered important means for diagnosis

[51] See *Carakasaṃhitā* Sūtrasthāna 30.29, p. 189b.

[52] See Karin Preisendanz, "The Initiation of the Medical Student in Early Classical Āyurveda. Caraka's Treatment in Context," in Birgit Kellner et al. (eds.), *Pramāṇakīrtiḥ. Papers Dedicated to Ernst Steinkellner on the Occasion of his 70th Birthday* (Vienna: Arbeitskreis für Tibetische und Buddhistische Studien, Universität Wien, 2007), vol. 2, 629–68, p. 634 and p. 649, where Preisendanz announced that she will deal with the relationship between ayurvedic and brahmanical rituals of initiation in a forthcoming article entitled "Medicine and Brahminical Orthodoxy in Ancient India: On Some Ritual Elements in the *Carakasaṃhitā*."

[53] Sheldon Pollock, "The Theory of Practice and the Practice of Theory in Indian Intellectual History," *Journal of the American Oriental Society* 105.3: *Indological Studies Dedicated to Daniel H. H. Ingalls* (1985), 499–519, pp. 512 ff.

and for the acquisition of valid knowledge in general. Additional means for the transmission and generation of medical knowledge were oral instructions from teachers to pupils as well as debates among fully educated physicians.[54]

Rationality also played an important role in medical treatment. As outlined in the eighth chapter of the *Carakasaṃhitā* Vimānasthāna, treatment was based on a complicated process of reasoning, in which the physician had to draw into consideration ten topical complexes consisting, among other things, of the qualities of the practicing physician himself, the medicine, disease, the geographical region, time – in general, with special reference to the season of the year, and with reference to the condition of the patient – and finally the patient himself from the perspectives of, for example, his or her natural constitution, vigor, size, affinities, character, bodily fitness, and age.[55]

Afflictions were either treated with ritual or with non-ritual means. Ritual or religious methods of healing consisted of, for example, giving gifts, invocations of blessings, food offerings, auspicious ceremonies, fire oblations, self-restrictions, penances, fasts, and the application of mantras.[56] This faith in the efficiency of medical rituals was based on the belief that in some cases diseases may result from destiny (*daiva*) or from ethically bad actions (*karman*) that the diseased person has committed either in a previous birth or earlier in his or her present life.

Non-ritual treatments were designed to clean body and mind, or to pacify a corruption. These therapies partly involved the application of medicinal substances. Treatments without medical substance consisted of, for example, pointing out dangers to the patient, or in surprising, making forget, agitate, delight, threaten, or strike him or her, or putting them to sleep, whereas there were diverse methods of non-religious medical treatment by means of medicinal substances. Among these, the "five therapies" (*pañcakarma*), i.e. emesis, purgation, two kinds of enema, and the evacuation of the head, figure prominently not only in the *Carakasaṃhitā*[57] but also, partly modified, throughout the later history of Ayurveda.

The ayurvedic therapies mentioned and described in the *Carakasaṃhitā* are theoretically grounded on a variety of partly supplementary and partly contradictory medical theories that were current around the time of the composition and compilation of this work. One account of such a theory occurs in the first book of the *Carakasaṃhitā*:

[54] On the education of a medical student according to the *Carakasaṃhitā*, see Dagmar Wujastyk, *Well-Mannered Medicine: Medical Ethics and Etiquette in Classical Ayurveda* (New York: Oxford Universtiy Press, 2012), pp. 68–109.

[55] See *Carakasaṃhitā* Vimānasthāna 8.68–151 (pp. 272b–286a) as thematically analyzed in Preisendanz, "The Initiation," pp. 659–60.

[56] See *Carakasaṃhitā* Sūtrasthāna 30.21, p. 186b translated into English.

[57] See *Carakasaṃhitā* Vimānasthāna 8.87, p. 275a.

> In this regard food becomes an essence, called "pure matter," as well as waste, called "impure matter." Sweat, urine, feces, wind, bile and phlegm, impure matter arising from the ears, eyes, nose, mouth and the pores of the skin and parts such as the hair of one's head, the beard, the hair of one's body, the nails, etc., thrives from waste, whereas chyle, blood, muscle flesh, fat, bone, marrow, semen and strength (*ojas*) develop from the food essence . . . When they are thriving from the [food] essence and from impure matter, all of these bodily constituents – called "impure matter" and "pure matter" – conform to their individual measure in accordance with age and body. Thus, when [food] essence and impure matter keep their individual measure, they maintain the suitable ratio (*sāmya*) of constituents belonging to a body [which can thus be regarded as] having constituents in a suitable ratio (i.e. to be healthy).[58]

According to this theory, digestion transforms food into two substances called "pure matter" and "impure matter," respectively. These two substances are further transformed into bodily waste products and bodily constituents. Health is the result of a suitable ratio of bodily constituents and waste products, whereas an unsuitable ratio causes disease. The three waste products wind, phlegm, and bile are most important among the listed constituents, because quite a number of passages of the *Carakasaṃhitā* (as well as of later ayurvedic literature) stress their ratio as the decisive factor for health and disease. In the context of their etiological potential, these elements are frequently termed "corruptions" (*doṣa*-s), i.e. pathogenetic substances or humors.

The physician can treat the ratio of the humors in the human body, because the amount of humors in the human body depends to some degree upon the tastes (*rasa*) of substances that the patient consumes. *Carakasaṃhitā* Vimānasthāna 8.1.4–8 provides a general outline of the relationship between tastes and humors:

> 4. First of all, there are six flavors: (1) sweet, (2) sour, (3) salt, (4) pungent, (5) bitter, and (6) astringent. Used properly, they support the body, but used wrongly, they agitate the humors. 5. The humors, for their part, are three: (1) wind, (2) bile, and (3) phlegm. In their natural state, they benefit the body. If, however, they are modified, they torment the body with manifold diseases. 6. In this regard, three flavors generate, and three flavors pacify a single humor in the following way. The pungent, bitter, and astringent flavors generate wind, but the sweet, sour, and salt flavors pacify it. The pungent, sour, and salt flavors generate bile, but the sweet, the bitter and the astringent

[58] *Carakasaṃhitā* Sūtrasthāna 28.4, as translated in Philipp A. Maas, "The Concepts of the Human Body and Disease in Classical Yoga and Āyurveda," *Wiener Zeitschrift für die Kunde Südasiens* 51 (2007/2008), 125–62, p. 136.

> flavors pacify it. The sweet, sour, and salt flavors generate phlegm, but the pungent, bitter, and astringent flavors pacify it. 7. However, if flavors and humors are in combination, flavors increase humors with identical or largely identical properties. Flavors having opposite properties or largely opposite properties pacify, if they are regularly consumed. Because of this relation, it is taught that uncombined there are six flavors and three humors. 8. The number of options for their combination is infinite, because there are infinite options.[59]

The theory underlying this relationship of tastes and humors is an early South Asian philosophy of nature, which holds material entities including medicinal substances to be modifications of the five gross elements space/ ether, air, fire, water, and earth. The six flavors inherent in medicinal substances derive from these elements. A high amount of water causes the sweet taste, whereas a predominance of fire and earth produce a sour flavor, etc. The flavors, which are properties of the substances, are metaphorically said to possess certain qualities or properties, namely the ten pairs of being heavy or light, cold or hot, unctuous or dry, sluggish or sharp, immobile or flowing, soft or harsh, clear or smeary, smooth or rough, gross or fine, viscid or liquid.[60] In this regard they resemble the three humors that also are modifications of the gross elements and have similar properties. However, the theory of the gross elements as the basis of the flavors

> suffers from a lack of explanatory force with regard to the properties and actions of the tastes ... It is especially hard to understand the relationship postulated between substances of a particular taste and specific disorders.[61]

The difficulties of understanding ayurvedic pharmacology as a consistent theory are aggravated by the fact that the efficiency of any medicinal substances depends not only on its flavors but also on its post digestive flavors (*vipāka*), its potencies (*vīrya*), and its specific action (*prabhāva*).[62]

In contrast to the flavors, which are six, the post-digestive flavors (*vipāka*) are only three, i.e. sweet, sour, and pungent. The reduction of number from

[59] *Carakasaṃhitā* Vimānasthāna 8.1.4–8, pp. 231a–232a, translated into English. Unlike "the number of options for combination," the number of combinations is not infinite but exactly sixty-three. On this combinatory problem see Dominik Wujastyk, "The Combinatorics of Tastes and Humours in Classical Indian Medicine and Mathematics," *Journal of Indian Philosophy* 28 (2000), 479–95.

[60] See Gerrit Jan Meulenbeld, "Reflections on the Basic Concepts of Indian Pharmacology," in G. Jan Meulenbeld and Dominik Wujastyk (eds.), *Studies on Indian Medical History: Papers Presented at the International Workshop on the Study of Indian Medicine Held at the Wellcome Institute for the History of Medicine, 2–4 September 1985* (Groningen: Egbert Forsten, 1987), pp. 1–18, p. 8, n. 19, who, however, only lists eight pairs.

[61] Ibid., pp. 6 f.

[62] The following account of post-digestive flavors (*vipāka*), potencies (*vīrya*), and specific actions (*prabhāva*) is based on Meulenbeld, "Basic Concepts of Indian Pharmacology," pp. 5–17.

six to three happens during the digestive transformation when salt becomes sweet, and bitter and astringent become pungent. This theory is apparently based on two observations: first, on the fact that some substances produce medical effects that cannot be explained on account of their flavors; and second, the fact that food changes its properties in the course of digestion, as can be directly experienced in the case of vomited substances. One of the problems resulting from the combination of the two theories of flavors before and after digestion is that it leaves the question unanswered of how salt, bitter, and astringent flavors can be medically efficient at all, if they are in any case transformed into sweet and pungent flavors.

The theory of potencies (*vīrya*) maintains that medicinal substances – independently of their flavors – are endowed with special properties that, according to the dominant view already at Caraka's time,[63] may consist of eight of the twenty abovementioned properties, namely of either being soft or sharp, heavy or light, unctuous or dry, and hot or cold. The potencies are held to resist the digestive fire, because they possess a special strength, so that they may even dominate the efficiency of the flavors and the post digestive flavors. The theory does not explain, however, why potencies consist only of a part of the so-called properties of flavors, and how the flavor-properties that do not survive the digestive fires are medically efficient.

The final pharmacological concept that is met with in the *Carakasaṃhitā* and later works of Ayurveda is that of specific action (*prabhāva*). This concept is evoked in cases in which medicinal substances produce effects that are unpredictable on account of their flavors, post-digestive flavors, and potencies. In other words, the specific action, which is seen as the result of the nature of the respective substance, serves as a joker to be drawn in order to explain otherwise inexplicable pharmacological effects.

The efficiencies of the four pharmacological factors described above differ from each other.[64] In cases of equal power, the specific action outweighs the three factors of potency, post-digestive flavor, and flavor. Moreover, potency is stronger than both post-digestive flavor and flavor, and, finally, the post-digestive flavor is more efficient than flavor.

A large amount of the *materia medica* mentioned in the *Carakasaṃhitā* consists of various parts of a variety of plant substance. The botanical identification of these plants is a difficult task requiring expertise in the diverse fields of knowledge of Ayurveda and its regional varieties, botany, pharmacology, Sanskrit philology, and the cultural and medical history of South Asia. Although sometimes the identification of plants mentioned by Caraka and later authors appears to be an easy task, there remain many problematic cases.

[63] See *Carakasaṃhitā* Sūtrasthāna 26.64, p. 147b.
[64] See *Carakasaṃhitā* Sūtrasthāna 26.72c–73b, pp. 148b–149a.

Besides plants and their various parts, Caraka prescribed the use of mineral substances and metals.[65] Moreover, his pharmacopeia contains animal products such as the milk, blood, urine, and meat of sheep, goats, cows, and other animals as well as, in certain cases, alcoholic beverages. This is remarkable, because the Brahmanical dietary prescriptions prohibit the consumption of meat and alcohol in virtually all cases. However, since in the view of Ayurveda health is the ultimate condition for achieving any aim in human life, the ayurvedic authorities considered medical prescriptions more important than following social and religious norms (*dharma*).[66]

MEDIEVAL AND EARLY MODERN AYURVEDA

The medieval and early modern period of Ayurveda is characterized by a large literary production of original works and commentaries that mirror exegetical efforts as well as conceptional innovations. From the nineteenth century onwards, three medieval works came to be designated as "the shorter triad" (*laghutaryī*), namely the *Rogaviniścaya* or *Mādhavanidāna* (ca. eighth century), the *Śārṅgadharasaṃhitā* (fourteenth century), and the *Bhāvaprakaśa* (at some time between 1550 and 1590) of Bhāvamiśra. In this way, these compendia were juxtaposed to "the longer triad" (*bṛhattrayī*) of classical works consisting of the before mentioned Compendia of Caraka and Śuśruta, as well as of Vāgbhaṭa's "Heart of the Eightfold Science."

The *Mādhavanidāna* is largely a compilation of passages from earlier works belonging to the classical period of Ayurveda that deals with etiology, prodromes, symptomatology, therapeutic diagnosis, and pathogenesis. It stands out for its innovative arrangement of topics related to disease that were previously dealt with in various passages scattered throughout the literature. More specifically, Mādhava invented a new scheme for the classification of diseases that became the standard for many later works on Ayurveda, in the context of which he mentioned and described a number of diseases that had not been known or recognized before.[67]

The *Śārṅgadharasaṃhitā* is remarkable not only because of its clarity and the well-structured arrangement of topics, but also because of its innovativeness. For example, Śārṅgadhara simplified ayurvedic theories by reducing the number of medically relevant qualities to five, and the number of potencies in medicinal substances from eight to two. In addition, he introduced pulse diagnosis as a new means to determine the nature of diseases.

[65] On inorganic substances in the *Carakasaṃhitā*, see Meulenbeld, *Indian Medical Literature*, vol. 1.A, 104 f.
[66] On the problem of meat eating in Ayurveda see Dominik Wujastyk, "Medicine and Dharma," *Journal of Indian Philosophy* 32.5 (2004), 831–42.
[67] For more details, see Meulenbeld, *Indian Medical Literature*, vol. 2.A, 61–77.

Śārṅgadhara's pharmacopeia included opium, and mercury for internal use, and his recipes frequently contain cannabis.[68]

The *Bhāvaprakaśa* contains, among many other innovative features, the earliest ayurvedic description of syphilis. Bhāvamiśra classified this disease into three types that he said resulted from contact with foreigners from Western countries. His treatment of syphilis draws mainly on mercurial drugs.[69]

MODERN AND GLOBAL AYURVEDA

The encounter of traditional South Asian medicine with modern bio medicine from the time of the British colonization onwards led to major and unprecedented challenges to Ayurveda. Ayurveda became politically, commercially, and conceptually dominated by modern bio medicine, which called the very validity of ayurvedic medical theories, practices, and courses of medical education into question. This process triggered the development of the two new kinds of Ayurveda that in recent academic writing are designated as "Modern" and "Global Ayurveda."[70]

"Global Ayurveda" refers to the diverse phenomena of ayurvedic medicine that were originally developed outside South Asia, such as, for example, New Age Ayurveda,[71] Ayurveda as mind-body medicine, Maharishi Ayur-Ved, and the modern continuation of traditional Ayurveda in urban settings. These spin-offs of a traditional South Asian medical system have in recent years become popular expressions of lifestyle also in the urban societies of India.

"Modern Ayurveda" is geographically located in South Asia. It is characterized by its adaptation to the standards of modern bio medicine with regard to the institutionalization of medical education, and the standardization of medical practice and pharmacology (cf. the agenda of the Indian Ministry of Health cited in section "Introduction"). Moreover, it deemphasizes (or even eliminates) the magical and religious aspects of Ayurveda and aims at establishing Ayurveda as an empirical science in the modern Western sense, although the claim that Ayurveda has been a strictly empirical science in the modern Western meaning of the term throughout its history is hard to maintain on the basis of an evaluation of ayurvedic Sanskrit sources.[72]

[68] For more details, see ibid., pp. 196–207.
[69] For more details, see ibid., pp. 239–47.
[70] See, also for the following part of this section, Frederick M. Smith and Dagmar Wujastyk, "Introduction," in Dagmar Wujastyk and Frederick M. Smith (eds.), *Modern and Global Ayurveda: Pluralism and Paradigms* (New York: SUNY Press, 2008), pp. 1–28, and the individual contributions to the volume.
[71] See Zysk, "New Age Ayurveda," pp. 10–26.
[72] See Steven Engler, "'Science' vs. 'Religion' in Classical Ayurveda," *Numen* 50.4 (2003), 416–63.

Part V

CHINA

28

MATHEMATICAL KNOWLEDGE AND PRACTICES FROM EARLY IMPERIAL CHINA UNTIL THE TANG DYNASTY

Karine Chemla [1]

The mathematical documents attesting to mathematical activity in ancient China, that is, during the time span between about the establishment of the Chinese empire in 221 BCE, by the Qin ruling house, and the Tang dynasty (618–907 CE), are of two very different types. Some were books handed down through the written tradition from later periods, whereas others are writings in manuscript form, mainly found in recent decades through archaeological excavations or bought on the antiquities market. This chapter argues that the extant sources pose two clear limits to our knowledge of the history of mathematics in ancient China. First, setting aside rare exceptions, they do not allow historians to discuss with certainty mathematical activity in the Chinese territory *before* the Qin dynasty (221 BCE–206 BCE). Second, examining more closely the nature of these sources suggests they only document mathematical practices and bodies of knowledge related to state institutions, and not beyond. Granting these restrictions, we will investigate the specific contexts in which our sources evince that mathematical activity was carried out.[2] We will also examine both the attested bodies of knowledge and the attested practices of mathematics, emphasizing the differences among them that depend on their specific context, and the mixture of theory and actual practice to which they testify.

[1] The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013) / ERC Grant agreement n. 269804. I have pleasure in thanking the editors for their generous help in the preparation of this chapter.

[2] Note, however, that I do not discuss mathematical activity in the context of the astral sciences, since it is dealt with in chapter 30 of this volume.

SOURCES HANDED DOWN THROUGH THE WRITTEN TRADITION: THE OFFICIAL SHAPING OF A TRADITION

A common fate links the mathematical books written in Chinese before the eighth century and handed down through the written tradition to a key figure – Li Chunfeng 李淳風 (602–70 CE) – and to a time period – the early decades of the Tang dynasty. A few words on this scholar and this time period will shed light on the history of the selective preservation of ancient mathematical documents.

Li Chunfeng,[3] who, from 627 onwards, enjoyed a position as official in the imperial bureaucracy of the Tang dynasty, is well known for his activities in the astral sciences and in the official historiography. Emperor Gaozu 高祖 (r. 618–26) and then Emperor Taizong 太宗 (r. 626–49) launched projects to compile official histories for previous dynasties, following a tradition that had been established in the Han dynasty (206 BCE-220 CE). In this tradition, for which two official histories composed during the Han had set a model, a standard history was expected to include monographs devoted to technical matters important for the ruling house. From 641 on, specialists were appointed to compose monographs for the official histories compiled at the Tang court. Li Chunfeng was one of them, and in this context he authored several technical monographs for the *History of the Sui dynasty* (581–617) and the *History of the Jin dynasty* (265–420), in which he dealt with the history of measurement units and standards, mathematics, astral sciences, and harmonics.[4] Among other sources, he relied on mathematical writings, which he quoted, and material artifacts he could examine in the court's holdings. We will return below, in the section entitled "The Management of Grains as Revealing a Link between Mathematical Writings and Administrative Regulations," to these monographs.

In the early decades of the Tang dynasty, equally great emphasis was placed on the canonical literature. State institutions launched the composition of new editions of Confucian canons, with selected ancient commentaries.[5] Sub-commentaries (commentaries on commentaries) on the selected canons and commentaries were also composed at the same time. These measures aimed at disseminating standard texts and interpretations. An edict, issued in 653, ordered *The True Meaning of the Five Canons* 五經正義 to be distributed throughout the empire. Moreover, an official education system, mainly based at a court institution, the "State Academy

[3] Documents for his biography are gathered in Li Yan 李儼，中國古代數學史料 *(Documents on Mathematics in Ancient China)* (Shanghai: Shanghai kexue jishu chubanshe, 1954), pp. 83–4, 87–8.

[4] On the compilation of official histories in Li Chunfeng's time and, more generally, historiography at the Tang court, see David McMullen, *State and Scholars in T'ang China* (Cambridge Studies in Chinese History, Literature, and Institutions; Cambridge: Cambridge University Press, 1988), pp. 159–73; Denis Twitchett, *The Writing of Official History under the T'ang* (Cambridge Studies in Chinese History, Literature, and Institutions; Cambridge: Cambridge University Press, 1992).

[5] McMullen, *State and Scholars in T'ang China*, pp. 67–89.

Directorate 國子監," was reestablished, to prepare students for examinations that opened the way to a career in the officialdom. This canonical literature played a central role in the curriculum of the main schools.

This context is meaningful if we seek to understand the process through which mathematical writings of the past were handed down through the written tradition. Indeed, canonical literature in mathematics simultaneously underwent a similar process. Li Chunfeng's biography in the *Old History of the Tang Dynasty* (an official history completed in 945) attests to a perception at the time that there were "canons" in mathematics, and that the reasonings of "the ten mathematical canons" of the past were in disorder. The Emperor likewise ordered that commentaries on these "ten canons" be prepared, and he appointed Li Chunfeng to fulfill this task with other officials. For some of these writings, the editors selected early commentaries. They prepared editions for all canons and commentaries, as well as a subcommentary on the set.[6] Li Chunfeng presented the resulting anthology, *Ten Canons of Mathematics* (算經十書), to the throne in 656. A "School of Mathematics 算學" had been established in the context of the State Academy Directorate. In 656, the Emperor likewise ordered that these canons be used for education in this institution. Various sources, including the *Digest of Tang Law* (唐六典), completed in 738, and the *New History of the Tang Dynasty* (compiled in 1060 under Ouyang Xiu 歐陽修's supervision), document the administrative organization of the "School of Mathematics," its curricula, and the related examinations. They show that the curricula relied mainly on the ten canons, grouped in two sets corresponding to two programs of seven years each. Further, they mention that two additional writings were also used in both programs.[7] This last fact implies that actors drew a distinction between canonical writings and textbooks.[8]

6 Features of their editorial work are analyzed in Karine Chemla, "Ancient Writings, Modern Conceptions of Authorship. Reflections on Some Historical Processes That Shaped the Oldest Extant Mathematical Sources from Ancient China," in Markus Asper (ed.), *Writing Science. Medical and Mathematical Authorship in Ancient Greece* (Berlin: de Gruyter, 2013), pp. 63–82.

7 On mathematical education in this context, and the curriculum based on the *Ten Canons*, see Siu Man-Keung and Alexei Volkov, "Official Curriculum in Traditional Chinese Mathematics: How Did Candidates Pass the Examinations?," *Historia Scientiarum* 9 (1999), 85–99; Alexei Volkov, "Argumentation for State Examinations: Demonstration in Traditional Chinese and Vietnamese Mathematics," in Karine Chemla (ed.), *The History of Mathematical Proof in Ancient Traditions* (Cambridge: Cambridge University Press, 2012), pp. 509–51; Alexei Volkov, "Mathematics Education in East-and Southeast Asia," in Alexander Karp and Gert Schubring (eds.), *Handbook on the History of Mathematics Education* (New York: Springer, 2014), pp. 55–72, 79–82. The latter gives a complete bibliography. Primary sources are gathered in Li Yan李儼, "唐丶宋丶元丶明數學教育制度 (The System of Mathematical Education in the Tang, Song, Yuan, and Ming Dynasties)," *Science* 科学 17.10 (1933), 1545–65. Republished in: 李儼錢寶琮科學史全集 *(Li Yan's and Qian Baocong's Complete Works in the History of Science)*, 10 vols. (Shenyang: Liaoning jiaoyu chubanshe, 1998), vol. 8, pp. 223–66.

8 On this distinction, see Karine Chemla, *Classic and Commentary: An Outlook Based on Mathematical Sources* (Preprint/Max-Planck-Institut für Wissenschaftsgeschichte, vol. 344; Berlin: Max-Planck-Institut für Wissenschaftsgeschichte, 2008), pp. 41–4.

The impact of these Tang state editorial enterprises and institutions on the preservation of mathematical writings of the past is pivotal. Firstly, the only mathematical writings composed in China before the eighth century that were handed down were either canons in this sense (this holds true for eight of them) or one of the auxiliary texts used as textbooks in the School of Mathematics, *Records on the Procedures of Numbering Left Behind for Posterity* (*Shushu jiyi* 數術記遺, hereafter abbreviated to *Records*) attributed to Xu Yue 徐岳 (fl. ca. 220). The only exception to this rule confirms it: an eighth-century book, the *Mathematical Canon by Xiahou Yang* (*Xiahou Yang suanj-ing* 夏侯陽算經) was handed down with the anthology, only because it was later mistakenly taken to be one of the canons that had been lost. If no other book of the past survived through the written tradition, two canons were nevertheless lost in this process: the book just mentioned, and the *Zhui shu* 綴術, whose authorship is attributed to Zu Chongzhi 祖沖之 (429–500).[9]

Moreover, all extant editions of the surviving canons bear in one way or another the mark of the editorial work carried out under Li Chunfeng's supervision, except one, to which we will return. This Tang edition of canons and selected commentaries, together with its sub-commentary, thus eclipsed all earlier editions. Through the documents handed down, historians cannot reach beyond the period of the Tang edition.

The transmission of these texts up to the present day is due to the fact that imperial institutions regularly launched important editorial projects that included mathematical writings in their scope. In this context, the canonical literature in its Tang form, and to a lesser extent the only surviving auxiliary textbook used in the Tang official education system, were given pride of place. The only (partly) extant editions of these texts that were carried out prior to the fifteenth century all derive from these official editorial enterprises. They include the edition of the Tang canons supplemented by another textbook, which Bao Huanzhi published in the thirteenth century, and the editions of some canons inserted in the encyclopedic project *Grand Classic of the Yongle Period*, commissioned by the Emperor Yongle (r. 1403–25) and carried out between 1403 and 1408.

We thus understand in what sense Li Chunfeng was a key figure, and the early decades of the Tang dynasty a key moment, for the history of the mathematical writings that survived through the written tradition. We also see that these writings had been selected by circles close to the court and were used in state institutions. This process of selecting canons and ancient commentaries conceivably has a long history. As we have shown, the mention of "ten canons" predates the work carried out under Li Chunfeng's supervision. Alexei Volkov has suggested that state institutions provided mathematical education perhaps as early as the Northern Wei dynasty (386–534). He has

[9] On Zu's scientific works, see Yan Dunjie 嚴敦傑, 祖沖之科學著作校釋 *(Annotated Collation of Zu Chongzhi's Scientific Works)* (Shenyang: Liaoning jiaoyu chubanshe, 2000).

further convincingly argued that the numerous mathematical writings that bibliographical treatises in official histories associate with the name of Zhen Luan 甄鸞 (ca. 570) might possibly indicate that some of the same "canons," together with the same auxiliary textbooks, were used for a state-run mathematical education during the Northern Zhou dynasty (557–81).[10]

THE CANONICAL CORPUS IN MATHEMATICS

As a background for our discussion, let us say a few words about the mathematical canons that, according to the testimony of official histories, formed the curricula in the Tang School of Mathematics.[11] Following Volkov's insight, I list the canons in the order given in the passages of the *Digest of Tang Law* and the *New History of the Tang Dynasty* that arguably describe the curricula. However, rather than the titles used in these documents, I list below the slightly different titles attested in received editions. These canons included:

1. *Mathematical Canon by Master Sun* (*Sunzi suanjing* 孙子算經). Qian Baocong argues that the book was composed around 400 CE, but its received version displays hints of later, Tang, changes. Its second and third chapters are composed of problems and procedures, whereas its first chapter introduces sequences of measurement units and tables. Most importantly, the book describes a system to represent numbers with counting rods on a calculating surface and related basic algorithms for common arithmetical operations.[12]
2. *Mathematical Canon of the Five Administrative Departments* (*Wucao suanjing* 五曹算經). Qian Baocong attributes its composition to Zhen Luan (ca. 570). However, the issue is still disputed. The book contains elementary procedures, given in the framework of problems and useful for local government officials. The titles of its chapters derive from various administrative departments, including those in charge of the management of croplands, granaries, and financial issues. Procedures are sometimes simpler than those in other canons and bring simpler types of numbers into play.[13]

[10] Volkov, "Mathematics Education in East- and Southeast Asia," pp. 58–9.

[11] What follows mainly relies on the critical edition of eight canons, the *Mathematical Classic by Xiahou Yang* and the *Records on the Procedures of Numbering Left Behind for Posterity* in Qian Baocong 錢寶琮, 算經十書（錢寶琮校點）*(Critical Punctuated Edition of* The Ten Classics of Mathematics*)*, 2 vols. (Beijing: Zhonghua shuju, 1963). Another critical edition was published more recently: Guo Shuchun 郭書春 and Liu Dun 劉鈍,算經十書 *(Ten Mathematical Classics).* 郭書春⊠劉鈍 點校 *(Punctuated Critical Edition by Guo Shuchun and Liu Dun)*, 2 vols. (Shenyang: Liaoning jiaoyu chubanshe, 1998; reprint, Taipei: Jiuzhang chubanzhe, 2001).

[12] For an English translation, see Lam Lay Yong and Ang Tian Se, *Fleeting Footsteps: Tracing the Conception of Arithmetic and Algebra in Ancient China* (revised edn; River Edge, NJ: World Scientific, 2004).

[13] The argument is made for the chapter on croplands in Chen Wei 陳巍 and Zou Dahai 鄒大海, "中古算書中的田地面積算法與土地制度——以"五曹算經"'田曹'卷為中心的考察 (Mathematical Procedures for the Areas of Croplands in Medieval Chinese Mathematical Books

3. *The Nine Chapters on Mathematical Procedures (Jiuzhang suanshu* 九章算術, hereafter abbreviated to *The Nine Chapters*). Like Qian Baocong, I argue that the text of the canon displayed in the ancient editions dates from the first century CE. The Tang edition included a commentary on *The Nine Chapters* that Liu Hui had completed in 263.[14]

4. *Mathematical Canon of the Sea Island* (*Haidao suanjing* 海島算經), by the same Liu Hui. Liu Hui's commentary on *The Nine Chapters* was introduced by a preface, in which he argued the Canon failed to cover all categories of mathematical problems. Liu Hui added to *The Nine Chapters* a section of text devoted to measuring lengths at a distance, with the intention of filling the gap indicated in the preface. The Tang editors eliminated this section from the edition of *The Nine Chapters* and turned it into a separate canon, ascribed to Liu Hui.

5. *Mathematical Canon by Zhang Qiujian* (*Zhang Qiujian suanjing* 張丘建算經). Qian Baocong argues this canon was composed in the second half of the fifth century. Its shape and topics generally follow those of *The Nine Chapters*. Besides a mention of Li Chunfeng's sub-commentary, the earliest extant edition, printed by Bao Huanzhi in the thirteenth century, indicates that Liu Xiaosun (in the sixth century) added details on the execution of the procedures (*xi cao* 細草). It also signals the inclusion of a commentary by Zhen Luan, whose text was apparently not handed down.

6. *Mathematical Canon by Xiahou Yang*, whose original text is lost.

7. *Mathematical Canon of the Gnomon of the Zhou [Dynasty]* (*Zhoubi suanjing* 周髀算經, hereafter abbreviated to *The Gnomon of the Zhou*). This book, whose completion is dated by Qian Baocong to ca. 100 BCE, expounds mathematical knowledge useful for calendar making, within the framework of a specific cosmographical theory ("Heaven like a chariot-cover (*Gaitian*)").[15] The Tang editors selected two commentaries for their edition of the canon: Zhao Shuang's commentary, completed in the third century, and Zhen Luan's sub-commentary, whose mistakes were pointed out by Li Chunfeng's sub-commentary.

8. *Mathematical Procedures for the Five Canons* (*Wujing suanshu* 五經算術). Qian also attributes the authorship of the book to Zhen Luan. The canon quotes excerpts of the five Confucian canons and ancient commentaries on them that yield numerical assessments. Zhen Luan adds comments describing mathematical procedures accounting for the correctness of the commentators' statements.

9. *Zhui shu*, mentioned above and now lost.

10. *Mathematical Canon continuing the Ancients* (*Qigu suanjing* 緝古算經), written by Wang Xiaotong in the first half of the seventh century. Except for a first problem, concerning an astronomical conjunction, the canon gathers problems

and the Land System – Examination Centered on the Chapter "Department of Croplands" in the *Mathematical Canon of the Five Administrative Departments*)," *Ziran kexueshi yanjiu* 自然科學史研究 28.4 (2009), 426–36.

[14] For a critical edition and a French translation, see Karine Chemla and Guo Shuchun, *Les Neuf Chapitres. Le Classique Mathématique de la Chine Ancienne et ses Commentaires* (Paris: Dunod, 2004).

[15] For an English translation of the Canon, see Christopher Cullen, *Astronomy and Mathematics in Ancient China: The Zhou Bi Suan Jing* (Cambridge: Cambridge University Press, 1996).

dealing with volumes or right-angled triangles, and solved by cubic or biquadratic equations.

Except for *Mathematical Canon continuing the Ancients*, all the extant ancient editions of the canons bear the mark of Li Chunfeng's editing and sometimes his sub-commentary. However, understandably, as an auxiliary textbook and not a canon, the received text of *Records* bears no mark of Li Chunfeng's intervention. The book describes in a discursive way various numeration systems and instruments for counting, some of which were thought to have been prototypes of the abacus. It was appended to *The Ten Canons*, and printed with this anthology, apparently for the first time in the edition Bao Huanzhi completed in 1213. This edition, the earliest extant one for (part of) this corpus, also constitutes the earliest example of printed mathematical texts worldwide.

Canons 1 to 8 were used in one of the seven-year programs of the Tang educational institution, whereas canons 9 and 10 were used in the other. The various sources attesting to the list of ten canons used in the Tang School of Mathematics differ in the order given. However, every time the curricula are described in detail, with the number of years per text, the order used is the one given above. This, and other facts, support Volkov's thesis that this order was in fact neither one of importance nor a chronological one, but the one followed in each curriculum.[16]

If we set aside Treatises 7 and 8, probably the last ones studied in the first program, all the treatises present mathematical knowledge in the form of problems and procedures (*shu* 術, in modern terms "algorithms"). Some of them also include numerical tables.

In conclusion, our survey of this first type of document reveals a key fact, which other pieces of evidence support: these treatises shed light *only* on mathematical practices and knowledge in ancient China meaningful in the context of specific state institutions. This does not mean, however, that mathematical activity was carried out uniformly in state institutions. As Zhu Yiwen has recently shown, the writings of other scholars, active in the same years in commenting on Confucian classics in relation to the Tang court, testify to different mathematical practices and knowledge.[17] The diversity of mathematical practices in state institutions awaits further research. However, through the documents handed down available so far, historians cannot meaningfully reach beyond them.

[16] Volkov, "Mathematics Education in East- and Southeast Asia," pp. 515–18.

[17] See Zhu Yiwen 朱一文, "Another Culture of Computation from 7th Century China," in K. Chemla, A. Keller, and C. Proust (eds.), *Cultures of Computation and Quantification in the Ancient World* (Dordrecht: Springer, forthcoming), and Zhu Yiwen 朱一文, "Different Cultures of Computation in Seventh Century China from the Viewpoint of Square Root Extraction," *Historia Mathematica* 43 (2016), 2–25.

SOURCES YIELDED BY ARCHAEOLOGICAL EXCAVATIONS: MATHEMATICAL MANUSCRIPTS

The twentieth century witnessed a revolution in the sources available to historians. Archaeology and then the antiquities market have yielded new types of sources. They shed new light in particular on mathematical activity in ancient China. At the present day, they constitute by far the earliest extant documents available, some of them being more than fourteen centuries older than the earliest document available before, that is, Bao Huanzhi's 1213 printed edition of the canons, mentioned earlier. In contrast to the canons, whose content we do not know through documents contemporary with their composition but through later editions carried out in the context of official projects, the writings to which the recently discovered manuscripts testify were not handed down through the written tradition. Moreover, the new documents show mathematical manuscripts in the form they had in the hands of their last users. Sometimes, the archaeological context also gives information on the social milieu in which these documents were produced and used. Roughly speaking, we can distinguish between three different kinds of context in which these sources were discovered.

First, in 1900, mathematical manuscripts written on paper were found among the tens of thousands of documents hidden in a chamber, adjacent to a cave-temple at Dunhuang (Central Asia), which had been sealed around 1000.[18] Numerical tables, in particular multiplication tables, and statements of systems of measurement units abound in these documents, the remaining content being reminiscent of parts of some canons. The stacks of manuscripts in the chamber give us no evidence of the context in which they were used. However, an inscription on the back of a manuscript seems to indicate that the document had been used in a learning environment.

Second, documents attesting to Han administrative activity, mainly written on wooden and bamboo slips once tied together by cords, or on wooden tablets, were also found in the northwestern part of China. Some of these documents show mathematical operations in the context of practice, for instance in accounting books recording quantities of grain received by granaries and granted either as rations or as salaries.[19] Samples of

[18] These mathematical documents are edited in Li Yan 李儼, "唐代算學史 (History of Mathematics During the Tang Dynasty)," in Li Yan 李儼 (ed.), 中國算學史論叢 *(Collection of Articles on the History of Mathematics in China)* (Taipei: Zhengzhong shuju 正中書局, 1954), pp. 26–99; republished in: 李儼錢寶琮科學史全集 *(Li Yan's and Qian Baocong's Complete Works in the History of Science)*, vol. 8, pp. 382–422. For an analysis, see Ulrich Libbrecht, "Mathematical Manuscripts from the Dunhuang Caves," in Zhang Mengwen, Li Guohao, Cao Tianqin, and Hu Daojing (eds.), *Explorations in the History of Science and Technology in China* (Shanghai: Shanghai Chinese Classics Publishing House, 1982), pp. 203–29.

[19] See account books translated in Michael Loewe (ed.), *Records of Han Administration*, 2 vols. (University of Cambridge Oriental Publications, 11–12; London: Cambridge University Press, 1967).

multiplication tables discovered in this context also testify to the learning and use of mathematical knowledge in officials' circles.[20]

The same conclusion seems to hold true for manuscripts more recently excavated from tombs, the third context of discovery of new manuscripts. In recent decades, funerary archaeology has significantly developed in China, and tombs sealed in the first decades of the Chinese empire have been explored. The libraries found along with other objects in some tombs sometimes contain mathematical manuscripts, which are relatively lengthy compared with the documents mentioned above. Several manuscripts of this type have been discovered to date, and more will probably surface. We expect that these new sources will significantly change our understanding of mathematical practices and knowledge in early imperial China. I will focus here on the two manuscripts whose texts have already been published.

In the winter of 1983/4, a first mathematical manuscript was uncovered at Zhangjiashan (Hubei province) in Tomb 247, which had been sealed ca. 186 BCE. The set of 190 bamboo slips on which the document was written bears a title, inscribed on the reverse of a slip: *Writings on Mathematical Procedures* (*Suanshu shu* 筭數書, hereafter abbreviated to *Writings*). The tomb occupant was an official working at a lower (local) administrative level of the imperial bureaucracy. Among the books he was taking to the afterworld, archaeologists also found a set of administrative regulations enacted in the second year of Empress Lü (186 BCE), *Statutes and Edicts of Year 2 (of Empress Lü)* (*Ernian lüling* 二年律令).[21]

The second significant mathematical manuscript, titled *Mathematics* (*Shu* 數), was discovered, along with other documents, in a set of more than 1300 bamboo slips bought in two lots, in 2007 and 2008, on the Hong Kong antiquities market. Its editors argue that *Mathematics*, written on more than 219 slips, dates from the Qin dynasty and was composed no later than 212 BCE.[22] These manuscripts being the product of illegal excavations, their

[20] On early multiplication tables, see Li Yan 李儼, 中國古代數學史料 *(Documents on Mathematics in Ancient China)*, pp. 14–18.

[21] For the documents found in the tomb, see 張家山二四七號漢墓竹簡整理小組 Group of editors of the bamboo strips from the Han tomb 247 at Zhangjiashan, 張家山漢墓竹簡（二四七號墓） *Bamboo Slips from a Han Tomb at Zhangjiashan (Tomb Number 247)* (Beijing: Wenwu chubanshe 文物出版社, 2001). For an annotated edition of *Writings on Mathematical Procedures*, see Peng Hao 彭浩, 張家山漢簡《算數書》注釋 *(Commentary on "Writings on Mathematical Procedures," a Document on Bamboo Strips Dating from the Han and Discovered at Zhangjiashan)* (Beijing: 科學出版社 (Science Press), 2001). Translations into English include: Christopher Cullen, *The Suan Shu Shu* 筭數書 *"Writings on Reckoning": A Translation of a Chinese Mathematical Collection of the Second Century BC, with Explanatory Commentary* (Needham Research Institute Working Papers; Cambridge: Needham Research Institute, 2004); Joseph W. Dauben, "算數書. Suan Shu Shu (a Book on Numbers and Computations). English Translation with Commentary," *Archive for History of Exact Sciences* 62 (2008), 91–178.

[22] The text of the manuscript is analyzed in, respectively, Xiao Can 肖燦, "嶽麓書院藏秦簡《數》研究 (Research on the Qin Strips 'Mathematics' Kept at the Academy Yuelu)" (PhD thesis, Hunan University 湖南大學, 2010); Zhu Hanmin 朱漢民 and Chen Songchang 陳松長主編 (gen. eds.), 嶽麓書院藏秦簡（貳） *(Qin Bamboo Slips Kept at the Academy Yuelu (2))* (Shanghai 上海: Shanghai cishu chubanshe 上海辭書出版社, 2011).

funerary context is irremediably lost. However, they were probably excavated from a single tomb, and, interestingly enough, administrative regulations were also found in the same set with the mathematical manuscript in this case too. This fact again indicates the plausible relation between the owner of the slips and the bureaucracy. We thus reach the same conclusion, whichever type of mathematical documents we examine: the sources that we have so far that document mathematical activity in ancient China were all produced or used in close connection with the imperial bureaucracy.

The same fact also confirms the link between the use of mathematical writings and the practice of administrative regulations that *Writings* had already suggested. Peng Hao and Zou Dahai have further shown that both manuscripts contain many problems and data reminiscent of administrative regulations, such as those discovered at Shuihudi for the Qin dynasty, *Qin Statutes in Eighteen Domains* (秦律十八種), and *Statutes and Edicts of Year 2* for the Han dynasty.[23] This holds true in particular regarding grains. We will now focus on this topic to reveal how intimate the link between these mathematical writings and the practice of regulation was at the time.

THE MANAGEMENT OF GRAINS AS REVEALING A LINK BETWEEN MATHEMATICAL WRITINGS AND ADMINISTRATIVE REGULATIONS

In the first decades of the Chinese empire, grains were an essential product for the state economy. The croplands owned by the state were allotted to people who had to pay taxes in return mainly in grain. Salaries were generally paid to officials in grain. Soldiers and corvée laborers also regularly received food rations in grain. The management of grain was thus a key issue, and some clauses in the administrative regulations dealt with it. *Writings* and *Mathematics*, in common with canons such as *The Nine Chapters*, contain many data, procedures, and problems concerning grains, which perfectly match the information provided in statutes and other administrative texts. More importantly, *Writings* quotes a regulation about grains for which a parallel occurs in the section on granaries in *Qin Statutes in Eighteen*

[23] Peng Hao 彭浩, 張家山漢簡《算數書》注釋 *(Commentary on "Writings on Mathematical Procedures," a Document on Bamboo Strips Dating from the Han and Discovered at Zhangjiashan)*, pp. 4–12; Peng Hao 彭浩, "Salary of Government Officials and National Tax Income as Seen in Qin and Han Dynasty Bamboo Strips – with a Focus on Mathematical Documents," in Cécile Michel and Karine Chemla (eds.), *Mathematics and Administration in the Ancient World* (forthcoming); Zou Dahai 鄒大海, "睡虎地秦簡與先秦數學 (The Qin Slips from Shuihudi and Pre-Qin Mathematics)," 考古 *(Archeology)* 6 (2005), 537–45; Zou Dahai, "Shuihudi Bamboo Strips of the Qin Dynasty and Mathematics in Pre-Qin Period," *Frontiers of History in China* 2.4 (2007), 632–54. The Qin statutes and edicts found at Shuihudi (Hubei province) are analyzed and translated in: Anthony F. P. Hulsewé, *Remnants of Ch'in Law* (Sinica Leidensia; Leiden: Brill, 1985).

Domains.[24] This regulation lists various grains, and various officially certified states of these grains (plant just harvested; unhusked or husked grain; and, for each grain, the certified degrees of fineness). It uses different types of measurement units (weight, capacity) to express quantities of grain. Notably, the regulation uses the same name, *dan* 石, to designate the highest measurement unit for weights and capacities. One of the purposes of the regulation is to define quantitatively how the various officially approved states of a given grain relate to each other.

Interestingly, chapter 2 in the first-century canon *The Nine Chapters* contains similar information in a tabular format. The list of grains considered is the same. However, the set of states of these grains mentioned is much longer. The nature of the numerical values expressing the relationship between the grains and these various states also differs, since in *The Nine Chapters* they are integers without measurement units (½ occurs sometimes). However, the ratios between the numerical values attached to the states of grain in both the regulation quoted in *Writings* and *The Nine Chapters* are for the most part identical. In *The Nine Chapters* quantities of grains are never expressed in weight measurement units, but only (at least apparently) in capacity measurement units. In this context, the name used for the highest capacity unit (*hu* 斛) now differs from the one for the weight system, which is still *dan*. This key change suggests that *The Nine Chapters* was compiled after Wang Mang's (r. 9–23) reform of measurement units.

Chapter 5 of *The Nine Chapters* also describes measuring vessels specifically designed for grains. These vessels embody, each for a type and a state of grain, a unit, also named *hu*, specifically associated with that grain and that state. *The Nine Chapters* defines these *hu* using volume measurement units. The fact that the *hu* used for specific grains have different volumes implies that in this other context, the unit cannot possibly express a capacity measurement unit, as it does in the other chapters. What do these *hu* mean?

Interestingly, the clue to solving this puzzle is found in a technical monograph that Li Chunfeng composed for the *History of the Sui Dynasty*. His discussion of the history of measurement units in this context relies on standards and documents of the past. In particular, he quotes Confucian canons and officially sanctioned commentaries on them. He also relies on mathematical writings such as *Mathematical Canon by Master Sun*, and *The Nine Chapters* with Liu Hui's commentary. This fact indicates that he considers these writings to be reliable enough to document an official history

[24] See in particular Peng Hao 彭浩, "秦和西漢早期簡牘中的糧食計量 (Measurement of Grains According to Excavated Documents of the Qin and the Early Western Han Time Periods)," 出土文獻研究 *(Research on Excavated Documents)* 10 (2012), 194–204; Zou Dahai 鄒大海, "關於'算數書'，秦律和上古糧米計量單位的幾個問題 (On Some Problems Regarding *Writings on Mathematical Procedures*, Qin Regulations and the Ancient Measuring Units for Cereals)," 內蒙古師範大學學報（自然科學漢文版）*Journal of Inner Mongolia Normal University (Natural Science – Chinese Edition)* 38.5 (2009), 508–15.

of measurement units. In his eyes, they are thus intimately related to state institutions.

Grains and the specific vessels described in *The Nine Chapters* are at the center of his discussion of capacity measurement units. This fact is in line with a long tradition in the technical monographs, in which grains are essential in the definition of official measurement units. The key point is that Li Chunfeng makes explicit that the vessels mentioned in *The Nine Chapters* define for grains a measurement unit of *value* also named *hu*, each vessel embodying the unit of value for the related grain.[25]

In fact, the table of grains contained in *The Nine Chapters* associates numerical values to each type and state of grain in such a way that the ratio between them is equal to the ratio between the volumes of the related vessels. This table thus defines the relationship between the volumes expressing the measurement unit of value attached to each type and state of grain. A significant part of the mathematical procedures also appears to relate to the management of these values. Historical writings and *The Nine Chapters* hence complement each other to document fully this feature of the management of grains at the time.

Similarly, the text of the Qin regulation quoted in *Writings* can be interpreted thanks to evidence provided in mathematical manuscripts; this evidence also complements the administrative documents in an essential way. Interpretation reveals that in the statute on grains, *dan* does not only designate a measurement unit for weight and capacity, but also for value. The Qin statute thereby appears to state the amount of each type and state of grain having the value of 1 *dan*. Further, the definition is based on a specific way of structuring the set of grains, which can be elucidated thanks to mathematical manuscripts. Mathematical documents thus highlight another key feature of the management of grain. *The Nine Chapters* exhibits exactly the same shaping of the system of grains.

In conclusion, mathematical and administrative sources complement each other to clarify the measurement units and the structure of the set of grains, on which administrative practice relied. These facts confirm the intimate connection between the two sets of documents, which the contents of tombs also imply. These conclusions also reveal continuity in the administrative regulations of grains throughout the first centuries of the Chinese empire, despite changes in the measurement units such as those evoked above. They likewise show the continuity between the two types of mathematical documents that we have distinguished above. Mathematical manuscripts, as well as the canons handed down, contain the same type of administrative data regarding grains, reflect the same management, and

[25] Karine Chemla, "Constructing Value with Instruments Versus Constructing Equivalence with Mathematics. Measuring Grains According to Early Chinese Mathematical Sources," in John Papadopoulos and Gary Urton (eds.), *The Construction of Value in the Ancient World* (Los Angeles, CA: Cotsen Institute of Archaeology, 2012), pp. 459–74, 536–95.

contain similar mathematical tools, essential to handle the calculation of equivalent quantities of different grains.[26]

Manuscripts thus help us sharpen our conclusion regarding the intimate connection between our sources on mathematical activity in ancient China and state institutions. Mathematics appears more precisely to have been useful in the definition and the implementation of administrative regulations. We also begin to perceive the close relationship between the recently excavated manuscripts and the canons. Further continuity, but also differences, will emerge when we now turn to the mathematical practices and bodies of knowledge to which manuscripts testify.

LOOKING AT MATHEMATICAL MANUSCRIPTS AS MATERIAL OBJECTS

What other changes have manuscripts brought to our knowledge of the history of mathematics in ancient China? Research on this topic is developing fast; hence the synthesis offered here is only provisional.

In the manuscripts, as in most of the canons, mathematical knowledge takes the form of mathematical problems and procedures for solving them. As a rule, however, the extant mathematical manuscripts contain more numerical tables, or knowledge presented in tabular format, than the canons do. Tables can be defined as lists of parallel clauses associating numerical values with each other, like a multiplication table, whose clauses associate in turn the two numbers multiplied and their result.

In the manuscripts, tables are signaled by specific types of text, for which we can identify two forms. In some manuscripts, the inscription of a table uses registers. A set of contiguous slips (placed vertically) is divided into horizontal registers (for instance, upper, central, lower) separated by horizontal blank spaces running across the set. In this case, each of the clauses composing a table is inscribed in one of the cells thereby formed in the column of writing that a slip records.[27] In other manuscripts, for example *Writings*, punctuation marks are used to separate clauses, which are then written continuously on the space of contiguous slips. The repetition of punctuation marks signals that the text is a table. We thus see that, in general, manuscripts abound in tables providing tools for computations, for example the usual multiplication table, tables of conversion between units, or tables specifying multiplications between powers of ten, between

[26] For the results about grains, see Karine Chemla and Ma Biao 馬彪, "How Do the Earliest Known Mathematical Writings Highlight the State's Management of Grains in Early Imperial China?," *Archive for History of Exact Sciences* 69 (2015), 1–53.

[27] An example is analyzed in Karine Chemla and Ma Biao, "Interpreting a Newly Discovered Mathematical Document Written at the Beginning of the Han Dynasty in China (before 157 BCE) and Excavated from Tomb M77 at Shuihudi 睡虎地," *SCIAMVS* 12 (2011), 159–91.

fractions, and between quantities formulated with length and surface measurement units. Tables are also used to record data or procedures, in particular related to grains.

This first feature echoes specifically two books, *Mathematical Canon by Master Sun* and *Mathematical Canon by Xiahou Yang*, which likewise contain more tables than the others. After each clause of a table, *Mathematical Canon by Master Sun* further appends computations based on the values occurring in it. This feature equally characterizes numerical tables found in Dunhuang and other, older, manuscripts. This reveals a first element of continuity. How can we interpret it?

As was mentioned above, Volkov has argued *Mathematical Canon by Master Sun* probably constituted the first canon studied in one of the curricula of the Tang School of mathematics. Strikingly, this fact echoes recent discoveries about the manuscript entitled *Writings*, which I summarize.

Daniel Morgan has shown that at least two hands alternated in the material production of the manuscript. He has further established that in one section of *Writings*, composed of several paragraphs, the two main figures who wrote the document alternated with one another, on the same slips. Interestingly, in each paragraph the upper part written by the first hand contains a coherent segment of a table, whereas the lower part written by the second hand writes the result of a computation based on the values occurring in the clauses recorded in the upper part. This section echoes the tables with appended computations from Dunhuang manuscripts and *Mathematical Canon by Master Sun*. Its fine structure and material features suggest that the section, and indeed the whole document, was produced in a context of mathematical education. Accordingly, different sets of clues relate *Writings*, like *Mathematical Canon by Master Sun*, to an educational context. We have drawn from this hypothesis several consequences.[28]

First, the type of tables, in which clauses, or a set of clauses, are followed by computations, might stem from a practice of teaching. Material features of the table in *Writings* possibly reflect how it was used in educational practice. If this were the case, the conclusion would yield a criterion to identify mathematical documents – or parts of them – produced in relation to mathematical education. It might give clues to determine stages in this

[28] The facts and provisional conclusions presented here derive from joint research with Daniel Morgan. For detailed arguments, see Mo Zihan 墨子涵 (Daniel Morgan), and Lin Lina 林力娜 (Karine Chemla), "也有輪著寫的：張家山漢簡《筭數書》寫手與篇序初探 (There Is Also Writing in Turns: Initial Investigation of the Hands and Compilational Order of the Han Bamboo Manuscript *Suan Shu Shu* (*Writings on Mathematical Procedures*) from Zhangjiashan)," *Jianbo* 簡帛 12 (2015), 235–52. A revised and augmented English translation has appeared: Daniel Morgan 墨子涵, Daniel Patrick, and Karine Chemla 林力娜, "Writing in Turns: An Analysis of Scribal Hands in the Bamboo Manuscript *Suan shu shu* 筭數書 (*Writings on Mathematical Procedures*) from Zhangjiashan tomb no. 247," *Silk and Bamboo* 1.1 (2018), 152–90. Other publications are in preparation.

training. In particular, perhaps *Writings*, like *Mathematical Canon by Master Sun*, belonged to an elementary stage. However, we must be careful: knowledge shaped for educational purposes might very well have been superseded in that function by other practices, while that knowledge was recycled for other purposes in other mathematical activities.

Second, the material features just mentioned suggest that *Writings* was *not* a "treatise," but rather a number of documents deriving from training in mathematics. This conclusion sheds light on the material relationship between some mathematical documents and an educational context, to which other facts mentioned above testify. It implies that manuscripts record texts strikingly different in nature from the canons. The conclusion also illuminates how documents produced in such a context were used at this time: the tomb occupant probably kept these documents, or used them throughout his life, and later even took them to the afterworld.

Other types of numerical tables found among manuscripts support this hypothesis. They are made using the same physical materials as that used for writing. However, they seem to have been *objects* used for computation. The earliest known example is a table kept at Qinghua University (Beijing), but it is likewise the product of illegal excavations. Its editors, who gave it the title "Calculating table" (*suanbiao* 算表), believe that it dates from even before the establishment of the Chinese empire.[29] The table is composed by a set of twenty-one bamboo slips that are longer and wider than those commonly used for writing. The slips form a double-entry table, whose cells are marked by red horizontal lines drawn across the set. The lines divide slips into cells that record results of multiplication between operands, one of which is on top of the slip, while the other lies on the slip placed to the rightmost position. On the right-hand side and on top, the set shaped by the slips also has holes, in which evidence of silk threads remains, indicating that threads were attached to the holes and helped the user to get the results. The editors argue that the object might have been used to perform multiplications, and perhaps also divisions, between numbers smaller than one hundred and also possibly having the fraction ½; this type of value echoes those used in procedures computing equivalences between grains. As before, one might assume this object was produced in a context of mathematical education and kept through a practitioner's life.

This table as an object echoes one of the Chinese manuscripts from Dunhuang (Pelliot 2490), which records a table written down in 952 and sharing features with this "Calculating table." Made with paper of rather

[29] See photos and analyses in 清華大學出土文獻研究與保護中心 (Qinghua University Center for Research and Protection of Excavated Documents) and Li Xueqin 李學勤 (eds.), 清華大學藏戰國竹簡 *(*肆*)* 上冊、下冊 *(Bamboo Slips from the Warring States Period Kept at Qinghua University. Part 4, First and Second Volumes)*, 2 vols. (Shanghai: Zhongxi shuju 中西書局, 2013); Li Junming 李均明 and Feng Lisheng 馮立昇, "清华简《算表》概述 (Overview of the Slips 'Numerical Table' Kept at Qinghua University)," *Wenwu* 文物, 8 (2013), 73–5.

large size, it is composed of pages attached to a wooden pole. Its sheets are divided into cells marked by red lines, and in the cells, area measurement units are used to record the results of multiplication between lengths, placed virtually on top and on the right hand side (except that some top entries are marked on protruding paper strips). The structural and material similarity between the Qin table and the Tang table is thus striking. The Dunhuang table might actually also be interpreted as a calculating tool, produced with paper, rather than a writing *stricto sensu*. This would point out another similarity in the use of material support for writing to make tools for practice. Finally, the corrected mistakes the Dunhuang table displays might indicate the user made it himself or herself, perhaps in an educational context.

The above conclusions about similarities in practice between manuscripts and canons, and between manuscripts from different time periods, could be reached only through an examination of manuscripts as material objects. The text of the newly found manuscripts also testify to mathematical knowledge and practices for time periods for which we had hitherto no evidence handed down. What changes do manuscripts bring to the history of ancient mathematics in this respect?

CONTINUITIES AND CHANGES IN MATHEMATICAL KNOWLEDGE AND PRACTICES AS EVIDENCED BY MANUSCRIPTS, CANONS, AND COMMENTARIES

The various tables mentioned above are not the only calculating tools to which manuscripts testify. The texts of procedures recorded in *Writings* and *Mathematics* yield additional clues. They both use the verb "place," sometimes several times in succession, or in expressions like "place on the left," to refer to computations that leave no trace in the writing. The verb clearly implies that there existed material representations of quantities, and that computations took place outside the text. *Writings* also mentions the use of counting rods (*suan* 筭) for this. However, no evidence allows us as yet to determine further features of these computations during this period. Moreover, in both *Writings* and *Mathematics*, the operation of division appears to be an issue as central as it was difficult, and it is the object of many procedures.

Canons equally testify to the use of counting rods for computations carried out outside the text. However, the use of counting rods and the execution of arithmetical operations to which they attest apparently differ from what the manuscripts document.[30] The canon placed at the beginning

30 For arguments on these issues, see Karine Chemla, "Shedding Some Light on a Possible Origin of a Concept of Fractions in China," *Sudhoffs Archiv*, 97.2 (2013), 174–98; Karine Chemla, "Observing

of the first curriculum, *Mathematical Canon by Master Sun*, describes in its first pages a use of counting rods to represent integers, and procedures for multiplication (*cheng* 乘) and division (*chu* 除) with them. They all indicate place-value decimal features for both the notation and the arithmetic. Arguably, the Tang editors considered these items of knowledge to be valid for all classics (and the curricula presented some homogeneity in this respect). The evidence on these topics found in the other canons supports this assumption, except for executions of division in some sections of *The Gnomon of the Zhou*. This latter canon, which is apparently (at least in some parts) the oldest writing among the ten canons, has sections that arguably share in this respect practices with the manuscripts rather than with the other components of the curriculum. Interestingly, the ancient commentaries selected for the Tang edition of the canon make this point explicit, systematically contrasting these executions of division with the division *chu*. This remark confirms what other pieces of evidence show: the Tang editors did not alter the text of the canons or commentaries attested in the documents they used.

From these assumptions, important conclusions follow. A decimal place-value notation for integers appeared at about the beginning of the Common Era in China, in parallel with new ways of executing division. Documents like *The Nine Chapters* attest to it. In this canon, the emergence of these new elements of knowledge can be correlated with several theoretical developments linked to this division. Among them, we find the development of new ways of computing square roots, the emergence of cube root extraction, the introduction of quadratic equations, and algorithms for the solution of systems of linear equations. Later canons attest to research carried out on root extraction and algebraic equations, reformulating and expanding what *The Nine Chapters* present. They also record the emergence of new but related interests, particularly in indeterminate analysis.

Beyond continuity in the use of counting rods, the discovery of manuscripts has thus shed light on theoretical work carried out around the beginning of the Common Era and devoted to a number system and arithmetical operations. This work shaped a new theoretical structure based on division. The emergence of the new division can also be correlated with a new practice of computation, performed on a surface on which counting rods were displayed.

This culture of computation is different from what the mathematical manuscripts document. It is also as a whole different from what we have evidence for in commentaries on Confucian canons, composed in other sectors of the bureaucracy in the first half of the seventh century. Since these

Mathematical Practices as a Key to Mining Our Sources and Conducting Conceptual History," in Léna Soler et al. (eds.), *Science after the Practice Turn in Philosophy, History, and the Social Studies of Science* (New York: Routledge, 2014), pp. 238–68.

commentaries relied heavily on older ones, probably their alternative culture of computation is older. However, since in particular the culture displayed in Confucian commentaries apparently uses no material tool for computation, this culture too differs from what is documented in manuscripts. Parts of Confucian canons and ancient commentaries elicited the exposition of these mathematical developments. In *Mathematical Procedures for the Five Canons*, Zhen Luan quotes the very same passages of the Confucian corpus, and accounts for them in terms of the mathematical culture dominant in the mathematical canons.[31] In conclusion, although all our documents attest to mathematical practices *only* in officials' circles, they are by no means uniform.

Like canons and commentaries, the two manuscripts manifest an interest in having exact results for division, in relation to the property that the inverse operation restores (*fu* 復) the original dividend. This requirement probably accounts for the introduction of fractions in the results of division: old divisions like new divisions yielded results as an integer possibly added to a fraction. Fractions are understood as "parts" and handled as a pair of numbers (numerator and denominator). Manuscripts and canons share similar procedures to execute key arithmetical operations on fractions. In addition, they share many mathematical topics and bodies of knowledge, including the rule of three, presented in relation to grains; procedures for unequal sharing, the area of figures, and the volume of solids; and rules of false double position.

These remarks highlight the strong continuity between manuscripts and the received literature. However, mathematical canons and manuscripts differ in some respects. Although knowledge related to the right-angled triangle and the so-called "Pythagorean theorem" is not wholly absent from manuscripts, it remains marginal. It is, however, central in *The Gnomon of the Zhou*, and prominent in other canons such as *The Nine Chapters* – which devotes chapter 9 to it. It may be that the topic was mainly significant in the astral sciences. This difference might thus signal a distance between the mathematics evidenced in the manuscripts and that attached to this other domain of activity. It arguably also reveals that canons like *The Nine Chapters* present a synthesis of mathematical knowledge attached to different spheres of activity. This is another insight suggested by the discovery of new manuscripts.

The introductory part of *The Gnomon of the Zhou* deals with an algorithmic version of the "Pythagorean theorem." Arguably it alludes to an argument establishing its correctness, and referring to diagrams that are also material objects outside the text and manipulated throughout the

[31] This paragraph relies on Zhu Yiwen 朱一文, "Another Culture of Computation from 7th Century China," and Zhu Yiwen 朱一文, "Different Cultures of Computation in Seventh Century China from the Viewpoint of Square Root Extraction."

argument.[32] This remark echoes practices evidenced by commentaries on *The Gnomon of the Zhou* and, even more importantly, *The Nine Chapters*. Commentators deal with proofs of the correctness of algorithms. To fulfill this task in the context of geometrical topics, they also bring figures and blocks into play as tools of interpretation. Likewise, figures, like blocks, were apparently material objects. The mathematical practice to which canons and commentaries bear witness is thus characterized by the fact that in addition to writings that contain *only* discursive parts, practitioners manipulated several types of material objects: counting rods used on a surface; figures; and blocks.

These conclusions rely on the explicit information that commentaries give on actual mathematical practices, to which the manuscripts, like the canons, only allude. They also emphasize that commentaries provide evidence for theoretical developments. These remarks invite us to turn to this evidence from commentaries. In addition to the Tang editors' sub-commentary, we have only a few older commentaries that they selected in their composition of the anthology of ten canons. In the first six canons of the first curriculum, only *The Nine Chapters* is accompanied by Liu Hui's commentary (and by Liu Hui's *Mathematical Canon on the Sea Island*) – we pass over the unclear case of *Mathematical Canon by Zhang Qiujian*. The last two canons taught, *The Gnomon of the Zhou* and *Mathematical Procedures for the Five Canons*, are also associated with commentaries – provided we consider Zhen Luan as in fact commenting upon Confucian scriptures. The situation is unclear in the other curriculum. Interestingly, the passages of sub-commentary composed under Li Chunfeng's supervision that seem less damaged also relate to these four canons.

A commentary like the one attributed to Liu Hui contains the earliest easily recognizable theoretical developments about mathematics written in Chinese. Liu Hui's proofs for the procedures in *The Nine Chapters* testify to a practice of mathematical proof whose goal is to establish the correctness of algorithms.[33] Arguably, the commentator thereby aims to identify the most general operations common to the greatest number possible of algorithms. The commentator further introduces theoretical concepts and philosophical developments attesting to a reflection on mathematics.

[32] The point is disputed. For a discussion and a bibliography, see Karine Chemla, "Geometrical Figures and Generality in Ancient China and Beyond: Liu Hui and Zhao Shuang, Plato and Thabit Ibn Qurra," *Science in Context* 18.1 (2005), 123–66.

[33] For syntheses of research on the issue, see Li Jimin 李繼閔, 東方數學典籍——《九章算術》及其劉徽注研究 *(Research on the Oriental Mathematical Classic* The Nine Chapters on Mathematical Procedures *and on Its Commentary by Liu Hui)* (Xi'an西安: 陝西人民教育出版社 Shaanxi renmin jiaoyu chubanshe, 1990); Guo Shuchun 郭書春, 古代世界數學泰斗劉徽 *(Liu Hui, a Leading Figure of Ancient World Mathematics)* (first edn; Jinan: Shandong kexue jishu chubanshe, 1992); Chemla and Guo Shuchun, *Les Neuf Chapitres*.

Commentaries also give evidence regarding how an ancient reader engaged with, for instance, *The Nine Chapters*. The observation of Liu Hui's and Li Chunfeng's reading of that canon also helps us grasp older theoretical features of mathematical knowledge and practices, concerning which the canon only gives clues. For instance, Liu Hui manifests his understanding that a particular problem and the procedure attached to it form a general statement, even if it is not abstract. More generally, he gives us essential hints about his practice of mathematical problems. Clues indicate that his practice of problems and generality is similar to that of the authors of *The Nine Chapters*.[34]

The commentator also refers to a practice of "abstraction" in the canon. Analysis suggests that he uses this term to refer to a specific feature of the text of *The Nine Chapters*. Regularly, the canon gives two texts of procedures for a given problem, which prescribe the same set of actions. The two texts are formulated differently, one (for the "upper level procedure") referring to the actions by means of theoretical terms that recur in the commentators' proof of the correctness of the other (for the "lower level procedure"). This feature of *The Nine Chapters* highlights that texts of procedure are not always merely direct prescriptions of actions. The commentator thereby gives us clues regarding how theoretical statements are formulated in the canon.

In this case, he ties the practice of abstraction with a specific practice of proof, in ways that are quite specific to this tradition. Interestingly, *Writings* shares this feature with *The Nine Chapters*, and for the same operations, whereas apparently the phenomenon does not occur in *Mathematics*. This remark suggests that despite commonalities, the archaeologically recovered manuscripts might possibly be of a different nature, a point awaiting further research. It also suggests a specific tie between *Writings* and *The Nine Chapters*. Noteworthy, however, is the fact that by contrast to "lower-level procedures," "higher-level procedures" are different in *Writings* and *The Nine Chapters*, showing that work has been carried out precisely on this theoretical dimension.

In conclusion, this account has made clear why the Chinese documents historians of mathematics have so far identified as potential sources for their inquiry do not seem to shed light beyond elements of practices and knowledge available in state institutions. It remains to be explored whether new sources of documentation could be found that would allow us to extend more widely our understanding of mathematical activities in the other social spheres that in China certainly fostered an interest in mathematics.

Moreover, we have shown that in spite of this limitation, manuscripts like canons manifest that mathematical knowledge and practice were not uniform across state institutions.

[34] Karine Chemla, "On Mathematical Problems as Historically Determined Artifacts. Reflections Inspired by Sources from Ancient China," *Historia Mathematica* 36.3 (2009), 213–46.

Finally, we see that despite their close relation to actual practice in specific sectors of the bureaucracy, notably the practice of administrative regulation and of astral sciences, and also their relation to mathematics education, some of these documents testify to an interest in theoretical work in mathematics. Contrary to an all too widespread dogma, theory is not in principle antagonistic to practical usefulness.

29

MEDICINE AND HEALING IN HAN CHINA

Vivienne Lo

The history of medicine in China takes us from antiquity through medieval religious organizations and along the Silk Roads, and then into the modern Chinese diaspora in all its cultural dimensions. This chapter focuses on the many approaches to the body that contributed to medical knowledge and practice in ancient and early medieval China. The period referred to by "ancient" spans the rule of the Shang people (traditional dates: 1766–1122 BCE) and the earliest extant written records through the political unification of a geographic entity identifiable as China (221 BCE) to the Eastern Han period (220 CE) and the failure of the early bureaucratic administration. The subsequent period of political fragmentation saw the rise of the rule of a northern aristocracy, and the growth of Buddhism and its alternative cultural and economic institutions. These were the centuries before maritime trade took over from the land routes, before the rise of a cash economy, and before the decisive impact of the invention of printing. This period, from the fall of the Han ruling house to the end of the cosmopolitan era of the Tang in the early tenth century, is generally thought of as distinctive for these reasons and is referred to as "medieval" in Chinese history.

Increasingly, historians of Chinese medicine, searching for culturally appropriate means to understand their subjects, contest fundamental assumptions about history, namely the constraints imposed by Christian, Marxist, or colonial models of time and interpretation when applied to the thousands of extant pre-Communist (to 1949) medical works listed in the 1991 *National Chinese Medicine Union Catalogue*. Some historians, rejecting the Whiggish pursuit of facts and progress developed in traditional histories of the sciences, look to Chinese ritualists, farmers, bureaucrats, and philosophers for alternative ways of dividing up biological time, framed by the seasons and the calendar, by the movements of stars and planets, or by successive generations with ancestors held directly responsible for ill health, or even incarnations in the case of Buddhist notions of healing.[1]

We have to question the tyranny of conventional historiography that is dominated by a narrative of progress from a medieval period (fifth–fifteenth century) through the Renaissance towards the eighteenth-century Enlightenment and contrasts it with Asian models characterized as cyclical, based on dynastic and regnal periods, cycles of rebirth, and seasonal time. Such histories may not stand up to a broader consideration of the contingencies of time both East and West.[2] We only have to factor in the millenarian beliefs in early Daoist practice for an example of linear time in Chinese thought or, conversely, the calendrical cycles of the medieval European world. Chinese medicine, in its classical form, has enduringly linked health and illness to the seasons and the movements of the heavenly bodies.

In the half-century since Joseph Needham began his project to write a history of science, technology, and medicine in China, seen in its fullest social and intellectual context, the approaches of social and cultural historians have provided new tools to unlock the many dimensions of more popular (i.e. common and socially pervasive) or religious healing practices.[3] The transitions between social and cultural history, the linguistic bias of the latter, and its emphasis on understanding local conditions of practice and the complex of underlying systems and traditions have proved auspicious for rich developments in the field.[4] A "field" was, however, always rather a flat thing when used to describe the History of Chinese Medicine – and too abstract. T. J. Hinrichs pointed out a decade ago, in "New Geographies of Chinese Medicine," that Chinese maps, or "charts," were inclined to fill in social spaces – temples, villages, and schools – in relief.[5] In the case of medicine this realization has meant a greater concentration on the social status of healers, the medical marketplace, and ritual ways of treating the body that align it with the spirit world, the stars, planets, and religious and almanac literature.

[1] Victoria E. Bonnel and Lynn Hunt, *Beyond the Cultural Turn: New Directions in the Study of Society and Culture* (Berkeley, CA: University of California Press, 1999), p. 4.

[2] Jack Goody, *The Theft of History* (Cambridge: Cambridge University Press, 2006) pp. 13–25.

[3] Lu Gwei-Djen and Joseph Needham, *Celestial Lancets. A History and Rationale of Acupuncture and Moxa* (Needham Research Institute Series; Cambridge: Cambridge University Press, 1980).

[4] W. H. Sewell, "The Concept(s) of Culture," in Bonnel and Hunt (eds.), *Beyond the Cultural Turn*, pp. 46–7. For examples of competing forms of health care in Ming novels see Christopher Cullen, "Patients and Healers in Late Imperial China: Evidence from the *Jingpingmei*," *History of Science* 31 (1993), 126–32. For records of medicine in government administration see T. J. Hinrichs, "The Medical Transforming of Governance and Southern Chinese Customs in Song Dynasty China (970–1279 CE)" (PhD thesis, Harvard University, 2003); Asaf Goldschmit, "The Song Discontinuity: Rapid Innovation in Northern Song Dynasty Medicine," *Asian Medicine* 1.1 (2005), 53–90.

[5] T. J. Hinrichs, "New Geographies of Chinese Medicine," *Osiris* 13 (1998), 287–325.

Excavated records recovered from archaeological sites of the Shang dynasty testify to very early divinatory techniques for identifying the cause and progress of illness, which is attributed to the malevolence of spirit ancestors.[6] Yet while modern forms of "traditional Chinese medicine" bear the marked vestiges of the astro-calendrical divinatory traditions, concerted attempts have been made by doctors and scientists in the twentieth century to eradicate its most obviously religious aspects.

Society in early China was interpenetrated by spirit presences, including gods, nature spirits, and deceased ancestors. The *wu* 巫 – diviners, mediums, shamans, or specialists in ritual, both male and female – played an important part in healing. They were employed at court well before the imperial age to avert demonic influences, resolve inauspicious events, and perform the work of communicating with the invisible realm of spirits. Their work involved performing exorcisms and sacrifices at the correct times in the annual and seasonal cycle, and interceding in case of natural disasters such as floods and drought. At funeral ceremonies they summoned up the spirits of the departed. They issued proclamations to expel illness and its causes, and used effigies and talismans to intervene in the course of disease. Female *wu* executed ritual songs, dances, and prayers, and participated in healing ceremonies alongside priests and medical practitioners of various kinds.[7]

The beginning of imperial China is dated to 221 BCE, when the military machine of Qinshi Huangdi 秦始皇帝 (259–210 BCE), first emperor from the state of Qin, put an end to centuries of disunity during the Warring States period of the Zhou dynasty (1045–256 BCE), and established the short-lived Qin dynasty (221–206 BCE). With brutal efficiency, the Qin regime molded a collection of small feudal kingdoms into a highly centralized state, broadly corresponding in geographic terms to what we know as China today. Qin Shi Huangdi achieved control of diverse states through the replacement of their various hereditary posts with a comprehensively centralized and bureaucratized government, and strict rule through detailed written legal codes. This centralization of power also entailed large-scale projects that aimed at standardization, not only in administration but across many aspects of life – currency, wheel and track sizes, and scripts – with varying levels of success. In his attempt to monopolize knowledge, the emperor famously ordered the burning of the kind of books treasured by scholars (although the extent to which he was successful is disputed), many

6 David N. Keightley, "Shamanism, Death, and the Ancestors: Religious Mediation in Neolithic and Shang China (ca. 5000–1000 BC)," *Asiatische Studien/Études Asiatiques* 52.3 (1998), 763–828.

7 Lothar von Falkenhausen, "Reflections on the Political Role of Spirit Mediums in Early China: The Wu Officials in the Zhouli," *Early China* 20 (1995), 279–300; Keightley, "Shamanism, Death, and the Ancestors," pp. 763–831; Michael J. Puett, *To Become a God: Cosmology, Sacrifice, and Self-Divinization in Early China* (Cambridge, MA: Harvard University Press, 2002), chapters 1–2; Donald J. Harper, *Early Chinese Medical Literature: The Mawangdui Medical Manuscripts* (London: Kegan Paul International, 1998), pp. 148–83; Paul U. Unschuld, *Medicine in China: A History of Ideas* (Berkeley, CA: University of California Press, 1985), pp. 17–50.

of whom he distrusted. He reprieved those writings, however, that he deemed to be of essential, practical value, including books on medicine – a measure of the high regard in which he held medicine, and of his concern for his own longevity. In China, medical practitioners were often literate, and their knowledge and practice can be reconstructed both from their own writings and from the written records of scholarly and religious traditions allied to medicine. Through two thousand years of empire, the authority and competence of the Chinese state were constantly embodied in a multitude of texts generated by the organs of government at every level, and medical practice was enmeshed in this bureaucratic process. Access to the upper echelons of the civil service was obtained via a succession of competitive examinations essentially testing mastery of the Confucian canons. An analogous hierarchy existed in scholarly medicine: increasingly, social status depended on the possession, knowledge, and authorship of written texts.

The Han Dynasty, which came to power shortly after Qinshi Huangdi's death, was a crucial watershed in which many of the distinctive aspects of imperial culture were consciously forged in attempts to distance the ruling house from the factionalism of the late Warring States and the realpolitik of the Qin rule. Scholars and bureaucrats imagined a prosperous, ordered, unified state with a centralized administration run by career civil servants. This was to operate according to ethics of merit, duty, and respect as embodied in rituals attributed to the Zhou, retrospectively idealized as a golden era. In fact by the end of the period, the Han had all the cultural diversity of a population of 58 million spread over an empire that stretched well into Central Asia, commanding the eastern end of the Silk Roads and trade links that reached as far as the Roman Empire.

The imperial project to create a unified sense of tradition drew on the authority of the sage rulers of a golden age at the dawn of Chinese civilization. To this end, the myth makers and history writers of the Han retold the stories of the lost golden age for their own times. In traditional Chinese accounts, the origins of medicine and the claim to authoritative wisdom of the medical classics were ascribed to the revelations of sages and cultural heroes. In the Han versions, the task of civilizing and domesticating a savage world fell to the five Sage Emperors, each of whom corresponded to one of the five directions: north, south, east, west, and center.[8] Two of them, the Yellow Emperor (Huangdi 黃帝) and the Red Emperor (Yandi 炎帝), also known as The Divine Farmer (Shennong 神農), are intimately associated with medicine and healing.

The legend of the Divine Farmer enshrines the empirical spirit of Chinese medicine, and the concomitant belief that knowledge of the virtues of drugs and food had to be obtained through trial and error. The Divine Farmer's

[8] Anthony Christie, *Chinese Mythology* (London: Hamlyn, 1968), pp. 84–91.

main role was to rescue human beings from a state of savagery, in which they fed on the raw flesh of the animals they hunted, drank their blood, and dressed in their skins, and to lead them towards an agrarian utopia. Famously, he tasted all living plants to ascertain their properties; and according to later accounts, he struck all the plants with a magical whip to make them yield up their essential flavors and smells. He subsequently classified the plants, and distinguished those that were safe and suitable for consumption and medicinal use. Testifying to the importance of this tradition of empirical testing, his name occurs in the titles of many famous *materia medica* texts, starting with *Shennong bencao jing* 神農本草經 (Divine Farmer's *materia medica*, ca. first century CE).

These cultural heroes are credited with formulating various ideas that are central to Chinese views of the world, the body, and human society. They include interlocked theories of cosmogenesis and statecraft structured around the polarity of Yin 陰 and Yang 陽 and the *wuxing* 五行 "Five Agents": the same basic principles that provide the framework for classical Chinese medical thought. The best-known patron of medicine is undoubtedly the Yellow Emperor. He is particularly associated with knowledge of the cosmic patterns believed to be inherent in all things in heaven and earth: he was considered responsible for establishing laws, punishments, and the calendar in accordance with these cosmic patterns, and he played a role in divination and in dividing up the seasons. These attributes link him with the specialized medical arts of understanding the body's relationship with the cycles and phases of nature, and accurately predicting the progress and outcome of disease.

A sense of transition is also found in changing attitudes towards medicine and healing. Han practitioners inherited and sustained traditions of healthcare that fused household remedies, skin-deep surgery, emergency medicine, demonic and spirit healing, therapeutic exercise, and sexual and breath cultivation. Some passed on their knowledge in rituals that enhanced the prestige of secret knowledge and ancient lore. Throughout the four hundred years of the Han period parts of these healing traditions became the building blocks for the new medicine framed in terms of the rubrics under which all phenomena came to be categorized: *qi* (the essential stuff that powers the universe); Yin; Yang; and the Five Agents (*wu xing*).[9] Knowledge about the new medicine was recorded in a growing body of texts written on silk and bamboo strips.

[9] The variety of translations of *xing*, which I translate as "agent" here, provides a clue to the range of meanings referred to by the term. The more popular translation "elements" is a rather European materialistic rendering of the powers of each *xing* "wood, fire, earth, metal, and water"; "phase" refers to their status as divisions of temporal cycles within a calendrical and seasonal context, and "agency" to their potency in dynamic interaction as an explanatory model for change in the phenomenal world.

The high proportion of texts about the body excavated at Mawangdui 馬王堆 tomb 3 (Changsha guo, present-day Hunan; closed 168 BCE, excavated in 1973), Zhangjiashan 張家山 tomb 247 (Nanjun, present-day Hubei; closed 186 BCE, excavated 1983–4), Shuihudi (Yunmeng, Hubei, ca. 217 BCE, excavated in 1975), and Wuwei (Gansu, first century CE, excavated in 1959) demonstrates that the healing arts were at the heart of scholarly attention at the dawn of the empire.[10] Throughout most of the Han period, medical texts carried a high status, and mere possession of a manuscript could enhance one's personal power and influence. This applied not just to the scholars and physicians who wrote, compiled, and used them – not necessarily all different people – and to the scribes who copied them, but also to noblemen who sponsored and collected the texts. It seems that early medical practitioners of the Former Han (second and first centuries) initially acquired knowledge through oral tradition: teachers would pass on their knowledge to their selected disciples by word of mouth. But committing medical knowledge to writing was a respected and well-established practice during the Han period, encouraged by the imperial court, which collected technical writing.

Medical manuscripts tend to tell very different stories from canonical works preserved in print. They are easier to situate socially and culturally and reveal more diverse forms of healing. Their physical grouping provides an indication of the contemporary classification of knowledge of the healing arts. For instance, the Mawangdui manuscripts recording the earliest extant theories of physiology were buried together with treatises on exercise, on breathing and sexual techniques, on herbs, on skin-deep surgery, and on magical procedures. Recent research into the kind of literature categorized together with remedy collections and standard works on medicine has begun to build a deeper and richer view of the healing arts and medical innovation in Chinese society.[11] It has brought into focus the intricately linked worlds of diviner and physician and their shared iatromantic culture of numerological calculation, astrology, and exorcism.[12]

[10] Seven of the thirty or so manuscripts buried in Mawangdui M 3 are devoted to the healing arts. There are thirty-six titles listed in the abbreviated catalog of the imperial library *Han shu* 漢書 [Book of the Han] (*HS* 30.1776–80) under *fang ji* "Remedies and Techniques," the contemporary classification that includes medical writing among many other practices.

[11] Michel Strickmann, *Chinese Magical Medicine*, ed. Bernard Faure (Stanford, CA: Stanford University Press, 2002); Sakade Yoshinobu 坂出祥伸, "Tounai no Jujutsu Chiryō ni tsuite – [Senkin Yokuhō] 'Kinkyō' wo Chūshin toshite 唐代の呪術治療について一『千金翼方』「禁経」を中心として," in *Ōkubo Takao Kyōjyu Taikan Kinen Ronshū – Karagokoro toha Nanika* 大久保隆郎教授退官紀念論集一漢意とは何か (Tokyo: Tōho shoten, 2001), pp. 433–58.

[12] Donald J. Harper, "Iatromancy, Diagnosis, and Prognosis in Early Chinese Medicine," in Elisabeth Hsu (ed.), *Innovation in Chinese Medicine* (Cambridge: Cambridge University Press, 2001), pp. 99–120; Lisa Raphals, "Divination and Medicine in China and Greece: A Comparative Perspective on the Baoshan Illness Divinations," *East Asian Science, Technology and Medicine*, 23 (2005), 26–53; Vivienne Lo, "*Huangdi hama jing* (Yellow Emperor's Toad Canon)," *Asia Major* 14.2 (2001), 61–100; Li Jianmin 李建民, *Sisheng zhi yu: Zhou-Qin-Han maixue zhi yuanliu* 死生之域：

The manuscripts contain, for example, the earliest extant set of Yin/Yang correspondences as they are recorded in a philosophic treatise excavated from the Mawangdui tomb. Set out as two lists, the qualities concerned with Yin and Yang provide a simple opportunity to explore what is and what is not universal about core divisions in our human tendency to think in opposites: down/up; outer/inner; night/day; warm/cold; female/male; autumn/spring; winter/summer; younger/older; inaction/action; earth/heaven; host/guest; silence/speech; receiving/giving; mourning/having a child; common or base/noble.[13] Where the initial pairs might seem to be more "natural" there is an increasing cultural specificity as the list progresses until the oppositions of "host/guest" or "mourning/having a child" become arresting enough to make the modern reader question the kind of social and cultural environment that produced them.

The Mawangdui manuscripts speak of spirit healing, the persistence of the belief that the body served as a dwelling place for spirits, and how good health involved avoiding the wrath of ancestors and the malevolence of demons. They include household manuals presenting remedies for a variety of illnesses ranging from hemorrhoids to convulsions. The latter are described as having the characteristics of horses, sheep, and snakes, perhaps referring to the noises emitted during convulsions or the quality of the involuntary movements. Charms, exorcism, many kinds of heat treatments, and basic surgery appear among the proposed cures. The tomb also contained the earliest specimens of Chinese medical herbs yet found, including magnolia, Chinese prickly ash, cassia bark, and wild ginger – all of which are used by Chinese herbalists today.

Three of the Mawangdui texts echo – in both style and content – treatises contained in the received canons of acupuncture and moxibustion, that is, a range of heating and cauterization with various materials, generally using *artemisia vulgaris* (mugwort); that is, we can understand both in terms of channels (*mai*) and the emerging importance of the essential physiological material (*qi*). But the way that they differ from the received canon is also significant. Here the emphasis is placed squarely on moxibustion, not needle or stone therapy. The channels (*mai*) described do not follow in the linked pathways of the received tradition, nor do they connect to the internal organs, or describe any loci for needling, but run as separate lines on the surface of the body and into the limbs. Also absent in the texts is any reference to a circulation of *qi* or a system of correspondences like the Five

周秦漢脈學之源流 (Taipei: Zhongyang yanjiuyuan lishi yuyan yanjiusuo, 2000); Donald J. Harper, "Dunhuang Iatromantic Manuscripts: P. 2856 R° and P. 2675 V°," in Vivienne Lo and Christopher Cullen (eds.), *Medieval Chinese Medicine: The Dunhuang Medical Manuscripts* (London: Routledge Curzon, 2005), pp. 134–64; Li Ling, *Zhongguo fangshu kao* 中國方術考 (Beijing: Dongfang, 2000).

[13] Mawangdui Hanmu Boshu Zhengli Xiaozu 馬王堆漢墓帛書正理小組 (ed.), *Mawangdui hanmu boshu* 馬王堆漢墓帛書 (Beijing: Wenwu, 1984).

Agents theory, even though by the second century BCE this was well developed in a ritual context.

With the addition of *Maishu* 脈書 (Writings on Mai) from Zhangjiashan, four early Han tomb texts map some eleven *mai* on the body, a contrast with the twelve or fourteen of later Han descriptions. Oddly the texts make little mention of *qi*, except in the context of breathing disorders or wind trapped in the belly. But there is one important exception: a passage from *Maishu* where the text speaks of piercing the *mai* with a stone to influence irregular movements of *qi*.[14]

If we consider the treatment proposed here to be acupuncture, it is the earliest extant reference to it. By contrast, the widespread use of moxibustion to "draw the *qi*" in Han times is clearly attested in the tomb texts.[15] Compared to needling, heat treatment tended to offer a more accessible and cheaper alternative, and the most popular and widespread technique used was *jiu* 灸, translated as moxibustion. It is likely that moxibustion derived from the practice of manipulating spirits around the body. In ancient times, sweet-smelling mugwort (*ai* 艾; *artemesia vulgaris*) – the main herb used by acupuncturists for moxibustion today – was believed to be efficacious in both driving out malevolent spirits and attracting benign ones, and it was burnt on or over the body for this purpose. By early Han times, its more refined and targeted application to restore weakness and treat pain through the medium of the physiological conception of body was seen as a way of manipulating *qi* in a manner that probably provided a model for later acupuncture formulations.

There is also an important artifact that corroborates the Han tendency to structure the body in lineal tracts such as is expressed in the tomb texts – at least along the Yangzi valley. One tomb of a high-ranking military commander excavated at Shuangbao shan 雙包山 (Guanghan commandery; in present-day Mianyang 綿陽, Sichuan), dating to no later than 118 BCE, has yielded a black lacquer figurine with red lines painted on its torso and vertically along the length of the limbs (Figure 29.1).

The conventional analysis is that it illustrates the medical theories found in the manuscripts with which it shares some similarities, notably an absence of acupuncture loci. It is 28.1 centimeters high and was wrapped in red fabric and placed in the "outer chamber" tomb (*guo* 槨). The figurine bears ten red lacquer lines, nine of which lead from the extremities to the head, while one follows the spine to the bridge of the nose.[16] These lines may have served as

[14] *Zhangjiashan* (*Maishu*), nos. 57–8. Zhangjiashan ersiqihao hanmu zhujian zhengli xiaozu 张家山二四七号汉墓竹简整理小组 (ed.), *Zhangjiashan hanmu zhujian* 张家山汉墓竹简 (Beijing: Wenwu chubanshe, 2006).

[15] *Mawangdui* 4 (*Tianxia zhidao tan*), nos 28–30. Mawangdui Hanmu Boshu Zhengli Xiaozu 馬王堆漢墓帛書正理小組 (ed.), *Mawangdui hanmu boshu* 馬王堆漢墓帛書.

[16] Vivienne Lo and Zhiguo He 何志國, "The Channels: A Preliminary Examination of a Lacquered Figurine from the Western Han Period," *Early China* 21 (1996), 81–123.

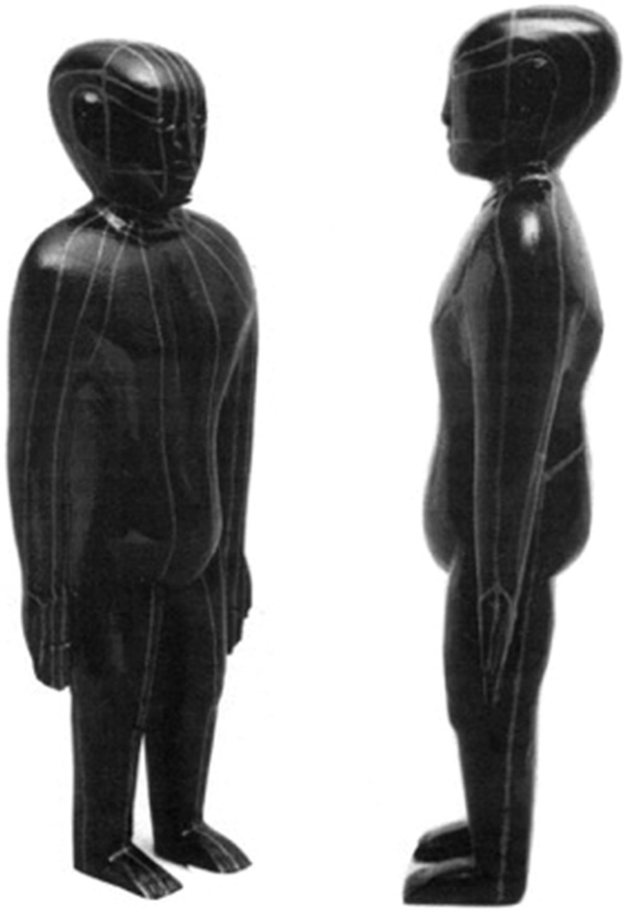

Figure 29.1. Black lacquer figurine excavated from a tomb at Shuangbao shan 雙包山.

a map of the *mai* channels, to be used as a medical teaching device or visual guide, in the manner of much later acupuncture models. However, that interpretation is anachronistic. The figurine is more likely to have been designed as an effigy that could assist the deceased in channeling vitality in the afterlife.[17] With the lines clustering around the sense organs, it also seems to provide new eyes and ears for the corpse.

Why are there only ten channels on this figurine? Mawangdui and Zhangjiashan tomb texts always show eleven *mai* channels in the charts. By the late second and early third centuries CE, the medical canons had

[17] Ibid. For a detailed discussion of this item, see Zhiguo He 何志國, "*Xi Han renti* jingmai qidiao kao 西漢人體經脈漆雕考," *Daziran tansuo* 大自然探索 3 (1995), 116–20.

settled on twelve circulation tracts associated with the organs and bowels, with two more associated with the extremes of Yin and Yang respectively. These, along with a number of subsidiary tracts and channels, were to become the basis of classical acupuncture. But it seems that for centuries there was no consensus on this central aspect of the new medicine.[18] It is possible that the number eleven was derived from the numerical system that associated the number six to heaven and five to earth, as recorded in the *Guo yu* 國語 and *Zuo zhuan* 左轉; eleven is also the number of the lunar phases of Yin and Yang, according to the early ritual calendars called "Monthly Ordinances."[19] In later medical canons, eleven appears as the five internal organ systems and six bowels, but none of the excavated texts make this connection, nor do they show the *mai* linking with the inner body.

It is interesting to note that the Mianyang figurine has no genitalia. In their general approach to body systems, Han physicians did not tend to distinguish between male and female: the physical differences were seen, rather, as equivalents. For instance, men and women both have *jing* 精, the essential form of *qi*: for men this manifests in the semen, for women this is in their reproductive fluids. In the twenty-five case histories recorded in the *Historical Records* biography of the physician Chunyu Yi (active from about 170 to 150 BCE) neither gender differences nor even their essential Yin and Yang nature are given any particular emphasis, either in the description of ailments or in their treatment.

Yet when it comes to the important health issue of sexual cultivation, gender difference becomes a salient issue. The tomb texts show that the function and stimulation of the sexual organs were the subject of considerable scrutiny. A chart found at Mawangdui depicts the female genitals, including the "red pearl" (clitoris) and the "wheat teeth" (pubic hair), and offers instruction on how to enhance female sexual pleasure (Figure 29.2). For the male partner, this was an opportunity to absorb life-enhancing Yin, as its availability increased with the pleasure experienced by the woman.[20]

Yin and Yang and *qi* are all brought to bear on the subject of a woman's sexual satisfaction and male potency. The Mawangdui text given the modern title of *He Yin Yang* 合陰陽 (Harmonizing Yin and Yang) describes the female orgasm in terms of *qi* flooding through the body bringing a brightness of the spirit, achievable through the successful cultivation of specific techniques that result in the reception (*jie* 接) of Yin. According to the same source, both successful breath cultivation and sexual cultivation produce the following effects:

[18] The names of the eleven *mai*, as well as their descriptions, vary somewhat.

[19] *Zuo zhuan* 41 (Zhao Gong 1), 26b. *Guo yu* 3 ("Zhou yu" C), 98.

[20] *Mawangdui* 4 (Tianxia zhi dao tan), nos. 12–67, especially 39. For further information, see Li Ling, Zhongguo fangshu zheng kao 中國方術正考, p. 315, and illustration plate 7. *Lingshu* 2 (9 "Zhongshi").

Figure 29.2. The earliest extant diagram of the vulva, Mawangdui tomb 3, closed 168 BCE. Photographed from Ma Jixing, *Zhongguo chutu guyi shu Zhongguo chutu gu yishu kaoshi yu yanjiu* 中國出土古醫書考釋與研究, 3 vols. (Studies and Textual Analysis of Ancient Medical Texts Excavated in China; Shanghai: Shanghai kexue jishu chubanshe, 2015), vol. 3, 486.

> The *qi* arrives, blood and *qi* flow freely, the ears and eyes are keen and bright, the skin gleams, the voice is clear, and the back, thighs, and buttocks are sturdy, so that one "attains a brilliance of the spirits" (*tong shen ming* 通神明).[21]

In these writings we find some of the earliest ideas about influencing the balance of Yin and Yang in the inner body. The texts are often written in beautiful, but rather obscure language, with techniques written in euphemisms that are now hard to interpret. Fortunately, lest we imagine that sexual cultivation was an entirely esoteric matter, practical remedies survive which tell us that achieving a "brilliance of the spirits" was inseparable from technologies of pleasure. Aphrodisiacs, for example, play a large part in recipe literature. For "increasing craving" in a woman one might use refined "... *fuling* (pine truffle) and discard the dregs. Use the liquid to fatten a suckling pig. Feed it to the woman. It makes her increase in sweetness and makes her inside become fine" or make suppositories of dried powdered cow horn, ginger, and cinnamon soaked in vinegar, or "Have the woman insert it [choice beef or venison] herself deep inside her prohibited part."[22] These remedy texts concerned with sexual stimulation are kept in close proximity with those that describe the techniques for harmonizing Yin and Yang and the proper control of *qi*. This involved both careful observation of the external signs of arousal and micromanagement of the inner sensory world through bodily discipline.

As China's healing arts documented aesthetic experience of how it felt inside to be well and strong, of experiences of pain, passion, and pleasure, it began to medicalize the sensory world. And it is in the language and theories of this culture of animating the inner body that we find a core innovation in early Chinese healing arts – one that survives to confound simple articulations of difference between mind, emotion, and body. The semantic circuits invoked by *qi* unite just these changing states of the inner sensory world.[23] They echo the aesthetics of an ancient time where the boundaries between these experiences were less distinct.[24]

Variations that occur in the medical concepts of *qi* are partially explained by the different practices through which the ideas developed. In some

[21] *Mawangdui* 4 (He Yin Yang), no. 133. See for example the conversations between Huangdi and Cao Ao, and then Rong Cheng; *Mawangdui* 4 (Shi wen), nos. 15–41.

[22] Harper, *Early Chinese Medical Literature*, p. 336. See also Donald J. Harper, "Ancient Medieval Chinese Recipes for Aphrodisiacs and Philters," *Asian Medicine* 1.1 (2005), 91–100; see also Vivienne Lo, "Pleasure, Prohibition and Pain: Food and Medicine in China," in R. Sterckx (ed.), *Of Tripod and Palate: Food, Politics, and Religion in Traditional China* (London: Palgrave Macmillan, 2005), pp. 163–86.

[23] Thomas Ots, "The Silenced Body – The Expressive Leib: On the Dialectic of Mind and Life in Chinese Cathartic Healing," in T. J. Csordas (ed.), *Embodiment and Experience: The Existential Ground of Culture and Self* (Cambridge: Cambridge University Press, 1994), pp. 116–39.

[24] Robert Jütte, *A History of the Senses: From Antiquity to Cyberspace* (Cambridge, MA: Polity Press, 2005).

contexts this essence was thought to flow downward through the body like flood waters, or rotated on the breath.[25] Daily physical and breathing exercises were prescribed to assist this flow, for fear that otherwise it would stagnate around the internal organs and cause disease. The joints were considered natural points on the body where this flow might be interrupted, and so treatment often focused on these areas such as the elbows and the back of the knees. The Zhangjiashan *Yinshu* 引 書 (*Writings on Pulling*) describes techniques to make *qi* move downward and into the limbs. One offers a therapeutic exercise for a hangover which has the following effect:

> When the *qi* of the head flows downwards, the foot will no longer be immobile and numb, the head will not swell, and the nose will not be stuffed up. Whenever there is free time, practice this often."[26]

As the imagery of flowing water guided early technologies of *qi*, so also the empire's waterways were applied to understanding *qi* physiology, and state measures for the prevention of flood and drought were likened to conserving all the essences that were the target of bodily cultivation. Acupuncture theory and the names of acupuncture points are replete with references to water from the "sea of *qi*" to the "crooked spring" and the "celestial well." Treatises on sexual cultivation, in particular, also provide early examples of the propensity to give lyrical, poetic names to points on the body.[27] Elaborate concepts of bodily *qi* and of Yin and Yang are found in the medical canons that date from the centuries around the turn of the millenium. But it is clear from the tomb texts that they were already being aired during the early Former Han, and applied in particular to the fields of breath, exercise, and sexual practice.

Through successive compilations and scholarly syntheses, a wide range of practices were gradually worked into a more unified orthodoxy. At the same time, there was a concerted attempt to bring medical knowledge within an all-encompassing belief system that saw divine order in everything – the stars, the spirits, the seasons, the role of the emperor and his government, the human body. The respect for past traditions – enshrined in rituals honoring the ancestors and the legendary cultural heroes of antiquity – and a desire for continuity gave the impression of an immutable backdrop to an intellectual consensus about the essential unity of the world. Nevertheless, new archaeological evidence demonstrates to us the wide range of ancient healing practices that persisted into the imperial period, challenging histories that would have us believe that the Han intellectual synthesis was universally

[25] Shigehisa Kuriyama, *The Expressiveness of the Body and the Divergence of Greek and Chinese Medicine* (New York: Zone, 1999), pp. 102–4, 223.

[26] *Zhangjiashan* (*Maishu*), nos. 36–7.

[27] *Mawangdui* 4 (*He Yin Yang*), nos. 102–5; Vivienne Lo, "The Influence of Nurturing Life Culture on the Development of Western Acumoxa Therapy," in Hsu (ed.), *Innovation in Chinese Medicine*, pp. 19–51.

effective in practice. Under the veneer of a new and systematic medical ideology, there survived remedy literature and textual miscellanies that contain knowledge and practice of a wide variety of healing arts.[28]

New evidence from the bamboo and silk texts recently excavated from late Warring States and Han dynasty tombs also calls into question the traditional dating of the classical canons of medicine to the Warring States period (475–221 BCE), so that "the burden of proof now falls to those who insist that ... any significant part of [it] was set down before the mid-first century BC."[29] The tomb texts reveal the complexity of the medical landscape from which the medical theory of the received canons emerged, and this evidence helps to support the hypothesis that the years marking the end of the Warring States and the beginning of imperial rule were critical for the writing that was later gathered and compiled in the *Inner Canon of the Yellow Emperor* (*Huangdi neijing* 黄帝內經), forming classical medical theory.

Throughout the Han period, a vast corpus of medical knowledge came to be ascribed to the Yellow Emperor. Compilations known as *jing* 經 (normally translated, not unproblematically, as "canon," and meaning something akin to "standard," or in the context of texts the "main" rather than the apocryphal, or the commentary) were published, setting out many of the cardinal principles of Chinese medical theory. The first imperial bibliography – the bibliographical treatise of *Hanshu* 漢書 (Han History), the official history of the Western Han dynasty (202 BCE–23 CE), compiled by Ban Gu 班固 (ca. 97 CE) – includes twenty-nine medical titles, which give some idea of the kind of medical literature that was officially sanctioned at the beginning of the first millennium. Among them, it lists a *Huangdi nei* and *wai jing* 黄帝內/外經 (*Yellow Emperor's Inner/Outer Canon*).[30] This is the earliest written evidence for the development of a canon of medical treatises attributed to the Yellow Emperor. Unfortunately none of the compilations listed in that bibliography has survived in the form or length indicated there, and the Yellow Emperor canon is now known only through its editors, and from editions based on a printed edition published in the twelfth century. Today the Yellow Emperor corpus exists in three recensions, which have taken on the character of three distinct books, each with a title beginning *Yellow Emperor's Inner Canon*, and subtitled "The Basic Questions" *Suwen* 素問,

[28] Donald Harper, "The Textual Form of Knowledge: Occult Miscellanies in Ancient and Medieval Chinese Manuscripts, Fourth Century BC to Tenth Century AD," in Florence Bretelle-Establet (ed.), *Looking at it from Asia: The Processes that Shaped the Sources of History of Science* (Boston Studies in the Philosophy of Science, 265; Dordrecht/New York: Springer, 2010), pp. 37–80.

[29] Nathan Sivin, "Huang ti nei ching 黄帝內經," in M. Loewe (ed.), *Early Chinese Texts: A Bibliographical Guide* (Early China Special Monograph Series, vol. 2; Berkeley, CA: Society for the Study of Early China and the Institute of East Asian Studies, University of California, 1993), pp. 196–215, p. 200; Lisa Raphals, "Notes on the Baoshan Medical Manuscripts," in Sarah Allan and Wen Xing (eds.), *Xinchu jianbo yanjiu* 新出簡帛研究 (Beijing: Wenwu chubanshe, 2004), pp. 388–408.

[30] David Joseph Keegan, "The Huang-ti nei-ching: The Structure of the Compilation, the Significance of the Structure" (PhD thesis, University of California, 1988).

"The Numinous Pivot" *Ling shu* 靈樞, and "The Grand Basis" *Tai su* 太素.[31] The three texts differ in subject matter; the first is largely devoted to medical theory – the human body as a microcosm, and the origins of disease – though it also describes some therapies (principally acupuncture and moxibustion), as well as a few drug prescriptions. The second deals mainly with acupuncture and moxibustion. The third incorporates aspects of both. There are also vestiges of a "Bright Hall Canon" or *Mingtang jing* 明堂經, which dates to approximately the third century CE.

In the juxtaposition of the tomb texts and the classical works of Chinese medicine we can see how the practice of using fine needles to adjust the essences of the body emerged out of the complex of diverse healing arts during the Han period: a product of *qi* and Yin/Yang practices, divination, numerology, minor surgery, and bloodletting, plus elements of spirit healing. This is affirmed by the "*Jiu zhen*" 九鍼 ("Nine Needles") chapter in *Lingshu: The Numinous Pivot*, in which the Yellow Emperor complains about the crude use of stone lancets in *qi* work. But it should be noted that similar procedures cited in this text refer mainly to small-scale surgery, bloodletting, and the treatment of abscesses, not to *qi* therapy. The discussion about the use of two of the nine needles – *yuan zhen* 員針 (round needle) and *pi zhen* 鈹針 (splitting needle) – in *The Basic Questions* follows similar lines. Those designated for moving *qi* were of very high quality, and likened to "fine hair." The intended outcome of the most subtle form of needling was consistent with the aims of sexual and breath cultivation: adjusting the *qi*, Yin, and Yang would make the senses astute, assure proper flow of the blood, and make the voice and appearance radiant and bright.[32] Technically, Chinese smiths were capable of producing fine metal needles at this time, but none have been found. The earliest extant references to piercing the body at named acupuncture points on the body date to the first century CE. There was nonetheless always a lingering concern about the use of needles: for example, a first-century CE manuscript recovered from an archaeological site in Wuwei, modern Gansu province, warns against harming the seasonal movement of the spirits and souls within the body, and proscribes needling in specific locations at given times of the year.

As Han physicians and thinkers came to grips with the puzzling behavior of sickness, they were guided by a vision of a microcosmic body, united in its essence with the cosmos and the state, and inhabited by the same spirits, which lent it their potency. Just as *qi* connected every phenomenon in nature with the movements of the heavenly bodies and thus with the deities and the spirits of the ancestors, the imperial rulers aspired to extend their sway everywhere under heaven. In an increasingly centralized state, the emperor played the crucial role of mediator between heaven and earth,

[31] Selections in Unschuld, *Medicine in China*, pp. 263–95.
[32] *Lingshu* 4 (9 "Zhongshi").

and this required him to carry out a cycle of complex rituals. Pursuing virtue, venerating one's ancestors, and performing the rituals correctly were ways of securing the gods' approval and ensuring order on earth. Disorder, in the form of civil unrest, natural disasters, famine, or disease, was a sign and consequence of the gods' displeasure.

Early medical theorists saw the interdependence of all the internal organs and bowels as homologous to the relationship between the offices of the empire. Just as good government and the judicious management of highways and waterways mirrored the divine order of heaven, so too did the healthy body. This was more than just an analogy: all formed part of an interconnected continuum. The power and prevalence of this belief goes some way to explaining why Chinese medical theoreticians showed little interest in following up investigations of the internal physical structures of the body known to them through dissections of the cadavers of the execution ground: their vision of the functioning of the human body rested almost entirely on relationships and correspondences perceived in the external world, with which it formed an indivisible bond. This approach to medicine has been called the "Medicine of Systematic Correspondence."[33] So, for example, the *Suwen: The Basic Questions* recension of the *Inner Canon of the Yellow Emperor* relates that:

> The heart is the office of the lord and ruler from whence the brilliance of the spirits emerges; the lung is the office of the minister from whence regulation and economies emerge; the liver is the office of the generals of the army, from whence strategies emerge.[34]

The organs are thus described through their official functions and bureaucratic interdependence. In another example, from the *Discourse on Salt and Iron* (*Yantie lun* 鹽鐵論, composed ca. 60 BCE), the way that the legendary physician Bian Que 扁鵲 used the "needling stone" (*zhenshi* 針石) to redistribute and balance *qi* in the body is taken as resonating functionally with the redistribution of wealth in the state.[35]

The flow of *qi* through the body operated as did the highways, waterways, and canals that provided essential traffic for the wellbeing of the empire; disruption to that flow had parallel consequences and called for homologous remedies. By the time of the Former Han, in the second and first centuries BCE, similar homologies were widely applied to the fourteen channels of the acupuncture body.[36]

By the end of the Han period there were a variety of theories, but *qi* was generally understood to transmit through the *mai*, without being confined

33 Unschuld, *Medicine in China*, pp. 51–92.
34 *Suwen* 3 (8 *Ling lan mi dian*), my translation.
35 *Yantie lun* 鹽鐵論 (Discourse on Salt and Iron, ca. 81 BCE), 3.
36 *Lingshu* 6 (12, *Jing shui*).

to it. In different contexts, charting the flow and influence of *qi* through the *mai* was guided by an understanding of the flow of rivers and canals, and the movement of the constellations as they operated in synchrony across micro- and macrocosmic levels. Han preoccupation with facilitating the movement of goods and people (troops and tax collectors) throughout the empire and with establishing regularities in the calendar for ritual and tax collection purposes found reflection in an unprecedented concern with enumerating and calculating the body and its inner organs and bowels, as well as the physiology of its inner fluids.

Teachers and students in the various medical traditions built up collections of texts and passed them on in manuscripts that had prestige and ritual significance. The content of the texts was often obscure and contradictory, even to readers at the time. Annotations and commentaries, easily differentiated at the time of writing, were designed to elucidate difficulties in interpretation. In successive compilations the scholarly apparatus became embedded into the main text, serving to add confusion and further confound the modern scholar. By accident or design, and particularly at the point of printing over a thousand years later, the compilation of texts and manuscripts would cease to change, and at this point they became the revered canons or classics that we know today. How our received canons of medicine relate to those standard texts, also named *jing*, as they were recorded in the ancient bibliographies and prefaces, is a tricky subject deserving ongoing philological analysis.

It was only towards the end of the Han period that ancient canonical works began to move from secret transmission into the public domain. This period marked an increasing tendency towards the systematization of knowledge. Produced during the early second century CE, the *Canon of Difficult Issues* (*Nanjing* 難經) set out to appraise, analyze, and explain many of the assumptions of the Yellow Emperor corpus. Still using the question-and-answer format, it represents both the apogee and the concluding chapter of that ancient form.

By now, learned physicians were beginning to express and take credit for their individual opinions, outside the framework of canonical literature. Rather than ascribing their work to legendary figures, they attached their own names to treatises, effectively mounting a public challenge to the tradition of secret and ritual transmission of medical texts. Zhang Zhongjing 張仲景 (fl. 196–205), for instance, wrote two treatises on febrile disease in the wake of an epidemic in his town. In these, he charts the progress of the disease in terms of the Yin and Yang system under external attack, and he suggests remedies for each phase. The treatises were later amalgamated to form the *Treatise on Cold Damage* (*Shanghan lun* 傷寒論). From citations in the received and manuscript traditions the text clearly circulated widely in the medieval period, but underwent a renaissance in the eleventh century, when it was heralded in imperial circles as the basis for great theoretical and practical innovations, associated with the treatment of epidemic disorders.

The late Eastern Han period (late second and early third century) witnessed multiple transformations in medical practice, with less prestige attaching to itinerant practitioners and more to elite families of scholar physicians, whose names were associated with books that had an enduring influence.[37] No longer was medical knowledge only passed down from master to disciple and sealed by the ritual conferral of bamboo and silk manuscripts in semiclosed medical lineages. Moreover, late Han soteriological movements entailed ritual transmission from religious leaders (not medical men), their medicine adding redemption for all ills through confession, acts of restitution, and charity. By the late third century large-scale collections of the manuscripts were compiled and would ultimately be transmitted in printed form. Printing culture, in large part developed from the late fifth century within a Buddhist context, was ultimately decisive in the decline of manuscripts as the primary conduit for the transmission of scholarly medical knowledge.

Social disruption marked the close of the Han era, and with it the loss of a centralized government, a hierarchy of nobles, and scholarly officials. New, often millenarian religious organizations developed new understandings of how spiritual matters caused ill health and approaches to treatment. Retrospectively labeled as "Daoist" (*daojiao*) in the Six Dynasties and associated with Laozi 老子 – the foundational deity of Daoism and putative author of its sacred text, *Daode jing* 道德經 (*Scripture of the Way and its Power*) – religious cults had been growing in momentum since the first century CE.[38] By the mid-second century, Daoist cults of healing such as the Celestial Masters (*tianshi* 天師) blossomed, focusing on the confession of sins, spirit possession, and worship of the deity Tai shang Lao Jun Lord Lao (*laojun*), understood to be the divine form of Lao zi, famous for his earlier writings on philosophy and statecraft. By treating illness with incantation and remedies made with infusions of the ashes of talismans, their techniques built on the practices of shamans.

The religious arena provided crucial continuity in the face of dynastic rupture and political transformation. In medieval times, various medical ideas were able to thrive and evolve in the context of religious movements.[39]

[37] Vivienne Lo and Li Jianmin李建民, "Manuscripts, Received Texts and the Healing Arts," in Michael Nylan and Michael Loewe (eds.), *China's Early Empires: A Re-Appraisal* (Cambridge: Cambridge University Press, 2007), pp. 367–97.

[38] Laozi, or Lao Dan 老聃, appears in early accounts as the sixth-century BCE author of the *Daode jing*, although we now understand the text to have a much more complex history.

[39] Unschuld, *Medicine in China*, pp. 117–53; Sakade Yoshinobu 坂出祥伸, *Taoism, Medicine and Qi in China and Japan* (Osaka: Kansai University Press, 2007); Sakade Yoshinobu 坂出祥伸, "Sun Simiao et le Bouddhisme (Sun Simiao and Buddhism)" (Kansai daigaku bunka ronshū, 関西大学文化論集, 1998), pp. 81–98; Ute Engelhardt, "Qi for Life: Longevity in the Tang," in Livia Kohn

At the beginning of the first century CE, millennial cults sprang up across China; some of them posed a threat to the power of the state, like the Yellow Turbans sect of Zhang Jue or Zhang Jiao (張角, d. 184), which led an uprising against the Han ruling house. Though the uprising was crushed, it signified the beginning of the end for the Han empire, which collapsed in 220 amid local wars, famine, epidemics, and waves of refugees. One of the ways in which the Yellow Turbans won converts for their cause was by offering to heal the sick, often by such ancient practices as incantation, and burning talismans and administering the ashes in water. Their main sacred text was the *Taiping Jing* 太平經 (Canon or Scripture of Heavenly Peace), a text grounded, on the one hand, in the theory of the Unity of Heaven and Humanity (天人合一), wherein individual virtue was thought to invite a corresponding response from Heaven. On the other hand, it also contained theoretical descriptions of the body comparable to the *Huangdi neijing*, as well as numerous longevity prescriptions encompassing meditation, breath and *qi* techniques, self-cultivation, diet, plant and animal drugs, and the use of charms and talismans. It was later assimilated into the Daoist canon.[40]

The *Shangqing* 上清 (Highest Clarity) school of Daoism rose to prominence in the fifth century under the guidance of Tao Hongjing 陶弘景 (456–536). A key figure in the history of alchemy and medicine as well as religious Daoism, he not only compiled the Shangqing corpus, but also wrote treatises on alchemy and published the first known critical edition of the Shennong pharmaceutical canon. He enjoyed imperial favor and patronage, especially for his work in the field of alchemy.[41]

The conjunction of medicine, alchemy, and high office is a recurrent theme in the lives of prominent medieval authors.[42] A distinguished example is the scholar-physician Sun Simiao 孫思邈 (581–682 CE), who held government posts at the beginning of the Tang period. Sun Simiao was noted for his eclectic intellectual and religious views, which are exemplified in two massive and wide-ranging medical works in which Buddhist chants and demonic medicine stand on an equal footing with classical scholarly medicine.[43] Classic Chinese alchemy set out to understand and master the

and Yoshinobu Sakade (eds.), *Taoist Meditation and Longevity Techniques* (Ann Arbor, MI: Center for Chinese Studies, University of Michigan, 1989), pp. 263–96.

40 Barbara Hendrischke, *The Scripture on Great Peace: The Taiping jing and the Beginnings of Daoism* (Berkeley, CA: University of California Press, 2006).

41 Michel Strickmann, "The Alchemy of T'ao Hung-ching, " in Holmes Welch and Anna K. Seidel (eds.), *Facets of Taoism: Essays in Chinese Religion* (New Haven, CT: Yale University Press, 1979), pp. 123–92; Mayanagi Makoto, "The Three Juan Edition of Bencao Jizhu and Excavated Sources, " in Lo and Cullen (eds.), *Medieval Chinese Medicine*, pp. 306–21.

42 Strickmann, *Chinese Magical Medicine*; Nathan Sivin, *Chinese Alchemy: Preliminary Studies* (Harvard Monographs in the History of Science; Cambridge, MA: Harvard University Press, 1968).

43 Sabine Wilms, "The Female Body in Medieval China: A Translation and Interpretation of the 'Women's Recipes' in Sun Simiao's Beiji quanjin yaofang" (PhD thesis, University of Arizona, 2002); Elena Valussi, "The Chapter on 'Nourishing inner nature' in Sun Simiao's Qianjin yaofang" (MA thesis, School of Oriental and African Studies, 1997); Fang Ling, "La tradition sacrée de

workings of the cosmos by studying its physical nature. By scrutinizing a substance in all its stages of transformation from its primordial state, an alchemist could learn to apply powerful analogies with cosmic time cycles – from the dawn of time to its end, wherein lies its beginning. Through a carefully calibrated process of successive heating and cooling, the alchemists attempted to speed up the sequences of time so as to transmute imperfect base metal into perfected "gold." These practices were known as *waidan* 外丹 (external alchemy).

The alchemists' desire to master the physical world led them on a quest for elixirs of longevity and immortality. Highly toxic minerals like cinnabar, mercury, lead, and arsenic were used to preserve the material body in life as well as death. Arsenic, a commonly used "mineral drug," is a nerve poison: when consumed over an extended period, even in small quantities, it results in lapses of consciousness, weakness, cardiac abnormalities, peripheral neuropathy, diarrhea, and delusions. However, it can also induce hallucinations and ecstatic visions; and it seems that this, together with the gradual character of the pathology, allowed users to embrace the symptoms of poisoning as acceptable side effects. Countless Chinese literati and even some emperors of the Tang dynasty perished from the effects of immortality elixirs over the centuries, a tragic irony that brought about the demise of external alchemy.[44]

As commercial and cultural interchange between China and the outside world intensified in the first century CE, Buddhism began to spread into China along the Silk Roads. Early Buddhism was at times misinterpreted (sometimes deliberately) in China as a Daoist sect, and much of Buddhist terminology, thought, and symbols can be found in Daoist sects.[45] Buddhism offered a radical new view of the afterlife centered on the idea of progressive incarnations of an immortal personal soul, and it proposed meditation and prayer as the main path to salvation and healing. The Buddhist devotion to deities struck a particular chord with indigenous popular religion, and the Buddha was readily assimilated into the local pantheon, often in the role of Medicine King, as were the boddhisatvas of healing, undergoing thorough sinicization in the process.

la Médecine Chinoise ancienne. Étude sur le Livre des exorcismes de Sun Simiao (581–682)" (PhD thesis, École Pratique des Hautes Etudes, 2001).

44 Joseph Needham (with Ho Ping Yü), "Elixir Poisoning," in his *Clerks and Craftsmen in China and the West: Lectures and Addresses on the History of Science and Technology* (Cambridge: Cambridge University Press, 1970), pp. 316–39.

45 Erik Zürcher, *The Buddhist Conquest of China: The Spread and Adaptation of Buddhism in Early Medieval China* (Sinica Leidensia, vol. 11; Leiden: E. J. Brill, 1959); Stephen R. Bokenkamp, "Daoism: An Overview," in Lindsay Jones (ed.), *Encyclopedia of Religion* (Detroit, MI: Macmillan Reference USA, 2005), pp. 2176–92; Livia Kohn, *Laughing at the Tao: Debates Among Buddhists and Taoists in Medieval China* (Princeton, NJ: Princeton University Press, 1995).

With increasing prosperity, the Buddhist monasteries became important cultural and social centers, some of them providing cheap hostel accommodation, epidemic relief, or free in-patient care in infirmaries called *Beitian fang* 悲田坊 (Fields of Compassion).[46] As ever, healing proved to be an effective mode of evangelism. However, the increasing material wealth and influence of monastic institutions brought them into collision with the state. Literary depictions of monk and nun healers play upon stereotypes of debauchery and immorality, much as in medieval Europe and India. Monks specializing in the treatment of women's illnesses bore the brunt of these prejudices. In the great suppression of Buddhism under the Tang emperor Wuzong from 842 to 845, thousands of monasteries were closed down or destroyed, their accumulated wealth was seized, and their infirmaries were nationalized.[47] But despite this persecution, monastic centers continued to play a vital role in the preservation and scribal transmission of medical literature. Our current knowledge of Chinese medicine in medieval times is therefore derived in great part from manuscripts copied by Buddhist monks living in far-flung communities along the Silk Roads.[48]

During the four centuries of the Han period classical ideas about the body and its treatment achieved a level of plasticity that allowed them to be seamlessly re-worked through the prisms of religious practice. Several hundred manuscripts recovered from the Buddhist Mogao 莫高 caves outside the town of Dunhuang 敦煌, in modern Gansu province, dating roughly to the Tang period, are particularly valuable for how they demonstrate these continuities and ruptures. Han period medical writers forged a language and set of theories that were integrated with the ritual and philosophy of the imperial process of the time, yet were malleable enough to remain relevant throughout the succeeding political turmoil and radical socio-political transformations. So successful were the Han synthesizers in styling the foundation of a flexible medical tradition that even today many practitioners around the world proudly claim to be practising an authentic Chinese medicine in the style of the early empire.

[46] Charles D. Benn, *Daily Life in Traditional China: The Tang Dynasty* (The Greenwood Press "Daily Life Through History" Series; Westport, CT: Greenwood Press, 2002), p. 227.

[47] Stanley Weinstein, *Buddhism under the T'ang* (Cambridge Studies in Chinese History, Literature, and Institutions; Cambridge: Cambridge University Press, 1987); Joseph Needham, "Medicine and Chinese Culture," in his *Clerks and Craftsmen in China and the West*, pp. 263–93, pp. 277–8.

[48] Lo and Cullen (eds.), *Medieval Chinese Medicine*.

30

CHINESE ASTRONOMY IN THE EARLY IMPERIAL AGE: A BRIEF OUTLINE

Christopher Cullen

INTRODUCTION

A long sequence of people in pre-modern China observed celestial phenomena, recorded and interpreted them, and tried to discover their regularities. Some of them worked as government officials, and others remained in private life. Their activities have generated a rich and complex literature. The object of this brief essay is to outline some of the main features of that activity, concentrating on the period in history when those features first took definite shape – the late third century BCE to the early third century CE.[1]

THE EARLY CHINESE EMPIRE AND THE HEAVENS

At the end of the third century before the Common Era, much of the land mass of East Asia was united for the first time under the rule of a centralized imperial power. This was the empire of Qin 秦, which took its name from the kingdom whose ruler fought his way to become the first emperor of the new realm in 221 BCE. Although empires succeeded one another for the more

[1] More detail will be found in two works by the present author. Christopher Cullen, *The Foundations of Celestial Reckoning: Three Ancient Chinese Astronomical Systems* (London: Routledge, 2017) translates and explains the ancient sources which specify the methods of calculation used in the period discussed, while Christopher Cullen, *Heavenly Numbers: Astronomy and Authority in Early Imperial China* (Oxford: Oxford University Press, 2017) is a narrative history that discusses the development and application of those methods in their political, social, and intellectual context. A short introduction to some of the main aspects of the topic in early imperial China is given in Christopher Cullen, *Astronomy and Mathematics in Ancient China: The Zhou bi suan jing* (Cambridge: Cambridge University Press, 1996); an analysis of developments in a much later period may be found in Nathan Sivin, *Granting the Seasons: The Chinese Astronomical Reform of 1280, with a Study of its Many Dimensions and a Translation of its Records* (New York: Springer, 2009). A discussion of the features of celestial calculation from a somewhat abstract point of view is given in Jean-Claude Martzloff, *Le calendrier chinois: structure et calculs, 104 av. J.-C.–1644. Indétermination céleste et réforme permanente. La construction chinoise officielle du temps quotidien discret à partir d'un temps mathématique caché, linéaire et continu* (Paris: H. Champion, 2009). On these and related works, the review of Christopher Cullen, "Translating Chinese Calendars," *Revue de synthèse* 131.4 (2010), 605–12 may be consulted.

than two millennia until the establishment of the Republic of China in 1912, and although there were periods of division and conflict, the peoples who lived in the broad expanse of territory that Qin had unified came to accept that imperial rule was the natural condition of things. Its supporters argued that it was a political expression of the basic order of the cosmos, and that it offered the only means by which a tolerable and orderly human life could be assured for the mass of people over whom the emperor ruled.

The Qin empire did not long outlast the death of its first ruler in 210 BCE. But the basic pattern it established – centralized administration by a non-hereditary civil service responsible (at least in theory) to the emperor – was continued for another four centuries by the Han 漢 dynasty (206 BCE–220 CE). This essay will discuss the way in which phenomena seen in the heavens were observed, recorded, analyzed, predicted, and interpreted by the people who lived under the empires of Qin and Han. The Qin/Han period, China's early imperial age, set the pattern for thought and practice in many aspects of life for centuries to come, and the study of the heavens was no exception. This is also the period from which we possess the earliest relatively full and systematic records of such matters. It is therefore reasonable to concentrate attention on the Qin and Han in order to outline the basic structures of thought and practice that were followed by those who watched, discussed, and wrote about the heavens in succeeding ages.[2]

The sky-watchers of Qin and Han did not believe that they were engaged in a new task. Although the political organization that they served and sustained was a novel one, they saw themselves as part of a long line of specialists stretching back to the star-clerks who were said to have been commissioned by the sage kings of remote antiquity, legendary rulers whose reigns were conventionally placed in centuries corresponding to the second half of the third millennium BCE. The first self-conscious state unanimously recognized by modern scholarship and archaeology as having existed within what was to be the heartland of the later empire was the kingdom of Shang 商. That state existed from the middle of the second millennium BCE up to its overthrow by its former vassals, the kings of Zhou 周, around 1046 BCE. The diviners employed by the Shang kings have left us copious written records of their work, in the form of inscriptions on bone in an early form of the Chinese script used today. From these texts it appears that they structured time by means of a luni-solar calendar, as did all their successors until the end of empire, and that they recorded and interpreted signs in the

[2] Excellent accounts of many aspects of the early imperial age will be found in Denis Crispin Twitchett, Michael Loewe, and John King Fairbank (eds.), *The Cambridge History of China, vol. 1: The Ch'in and Han Empires, 221* BCE– *220* CE (Cambridge: Cambridge University Press, 1986). Note that the older "Wade-Giles" romanization of Chinese sounds used in that volume differs from the pinyin system used in this essay. Principal differences include the use of Ch'in for Qin, and Ch'ing for Qing, also Chou for Zhou. "Qin/Ch'in" sounds like the English word for part of the face, and 'Zhou/Chou' sounds like the first syllable of "Joseph."

heavens, again as did all their successors. Although the Zhou kings lost all real political power in 771 BCE when they had to flee to the protection of powerful vassals after their capital was sacked, they continued as nominal rulers until their line was extinguished in 256 BCE. From the eighth to third centuries BCE, the supposedly vassal states of Zhou became increasingly independent and competitive, both culturally and militarily. It is from this period that we have the first extensive texts that give evidence of how knowledge of the heavens was gathered and applied to the needs of human society.[3]

OUR SOURCES

What sources are available for writing a history of the knowledge and understanding of celestial phenomena in the early imperial age? Firstly, and most importantly, we have the first three of the great series of standard histories, *zheng shi* 正史, that have come down to us from the centuries of China's imperial past, initially by scribal transmission and later (from around 1000 CE) by printing. These lengthy works, often compiled under official patronage, give systematic accounts of the history of their times, dynasty by dynasty, written from the point of view of the literati who worked in senior posts in central government. From our point of view, the *zheng shi* have two important features: firstly they often transcribe at length from imperial edicts, memorials submitted to the throne, and other official documents; secondly most include specialist monographs giving an account of the activities of particular departments of government – including those responsible for the observation and analysis of celestial phenomena. The *zheng shi* referred to in this essay will include:[4]

(a) *Shi ji* 史記 (*Records of the Historian*). Completed by Sima Qian 司馬遷 ca. 90 BCE.
(b) *Han shu* 漢書 (*Writings on the Han*). Mainly by Ban Gu 班固 (32–92 CE); probably completed in present form ca. 110 CE. Covers the first half of the Han dynasty (Western Han).

[3] The period from the loss of Zhou royal power to the extinction of the line of kings is often called the Eastern Zhou, distinguished from the preceding Western Zhou by the eastwards shift of the royal capital from a site near modern Xi'an 西安 to Luoyang 洛陽. The Eastern Zhou is conventionally divided into (1) the Spring and Autumn *Chun qiu* 春秋 period, named after the annals of Lu 魯 (the home state of Confucius, 551–479 BCE), which covered the years 771–481 BCE; (2) the Warring States *Zhan guo* 戰國 period, which ends with the establishment of the Qin empire in 221 BCE, and is often taken to begin in 403 BCE when the great state of Jin 晉 split into three.

[4] All references to these histories in this essay are made using the modern collated and punctuated editions published by the Zhong Hua press since the 1950s, and available in any sinological library, as well as online via the website of Academia Sinica, Taibei: http://hanchi.ihp.sinica.edu.tw/ihp/hanji.htm.

(c) *Hou Han shu* 後漢書 (*Writings on the Later Han* or *Later Writings on the Han*). Fan Ye 范曄 (398–445 CE). Main text completed ca. 445 CE, monographs by Sima Biao 司馬彪 (ca. 240–c. 306 CE) added later. Covers the second half of the Han dynasty (Eastern Han).

Apart from the *zheng shi*, two further types of sources will be useful to us. Firstly there are a few books handed down from the Han that include at least some material relevant to our discussion. Of these, the most important are the *Huai nan zi* 淮南子, compiled under the auspices of the Han prince Liu An 劉安 and offered to the throne in 139 BCE, which devotes the whole of its third chapter to matters relating to the heavens,[5] and the *Zhou bi* 周髀, which probably reached its final form around 10 CE.[6] Secondly, there are an increasing number of relevant texts excavated by archaeologists from Han tombs, ranging from simple calendrical documents to more complex texts such as the two mentioned in the next section.[7] Such documents give us access to a range of concerns and approaches that sometimes seem radically different to those in the corpus of received ancient literature that was all that was known before recent decades.

THE HERITAGE OF SKY-WATCHING: BEFORE THE IMPERIAL AGE

The Qin and Han empires built on and adapted a complex cultural foundation laid down in preceding centuries, as much in their understanding of the heavens as in other areas. From the earliest Shang inscriptions onwards, it is clear that we are reading material from a society that cared about dates, and that in order to produce those dates someone must have been paying attention to celestial phenomena. The most basic form of dating was the sexagenary day system, in which each day was given a two-character name formed by the systematic permutation of sets of ten and twelve signs so as to designate the number of the day in a sixty-day cycle. Above the day was the month, whose alternation between twenty-nine and thirty days in length indicates that we are dealing with lunar months; these months began in principle on the day that included conjunction of the sun and moon.[8] A year normally consisted of twelve lunar months, but since this would normally have contained no more than $6 \times 29 + 6 \times 30 = 354$ days, it would clearly be

5 See the study and translation of relevant chapters in John S. Major, *Heaven and Earth in Early Han Thought: Chapters Three, Four and Five of the Huainanzi*, with an appendix by Christopher Cullen (Albany, NY: State University of New York Press, 1993).

6 See the study and translation in Cullen, *Astronomy and Mathematics*.

7 Disturbingly, an increasing number of texts looted by tomb-robbers (and hence lacking archaeological provenance and context) are also coming into circulation.

8 In practice, however, calendars mostly followed the mean lunation, so the true conjunction might not always fall during the first day of the month.

too short to follow the annual cycle of the seasons, which repeats at an interval of close to 365¼ days, the period nowadays called the "tropical year." It was therefore necessary from time to time to insert an intercalary month *run yue* 閏月, producing a thirteen-month year, in order to bring the year and the seasons back to a closer correspondence. It is obvious that any ruler who wanted to use such a dating system must have maintained specialists to advise on when lunar months should begin and end, and when a new year should be marked.

In the way of such things, we are not told of such specialists being congratulated when their work was done efficiently, but only hear of them when something goes wrong. Thus, in a story dated to 484 BCE, Confucius[9] is represented as commenting on the anomaly that grasshoppers were still active in the twelfth lunar month of that year, and that moreover Antares (α Scorpii) was still visible in the west after sunset, both signs that the calendar was getting ahead of the seasons because an intercalation was overdue. He remarks "*si li guo ye* 司厤過也 (This is an error by those in charge of the calendar)."[10] The reference to the visibility of Antares makes it clear that running the calendar efficiently involved not only watching the phases of the moon, but also checking on other astronomical phenomena that gave guidance on the cycle of the seasons. Sources from the Warring States period suggest that one important way of tracing the cycle of the seasons involved lists of "centered stars" *zhong xing* 中星 that would culminate due south at dusk and dawn at different seasons of the year.[11] We also begin to see the first signs of another method for checking on the seasonal cycle, which was the use of a simple vertical pole gnomon *biao* 表 to observe the days when its noon shadow was shortest (at summer solstice) and longest (at winter solstice): on this instrument, see the discussion below, in the section "How Did They Make Observations?" By the beginning of the imperial age this had become the standard method of giving a precise reference frame to the seasons based on celestial phenomena, and it was the interval between winter solstices that was seen as the fundamental period underlying the annual cycle.[12]

[9] This is the customary Latinization (introduced by Jesuit missionaries to China) of the name of Kong Qiu 孔丘 (ca. 551–479 BCE), a thinker and teacher whose doctrines were later taken by many to be the basis of all ethical and social practice.

[10] *Zuo zhuan* 左傳, Duke Ai, twelfth year; Ruan Yuan 阮元 (1764–1849), *Shi san jing zhu shu* 十三經註疏 *(The Thirteen Classics with Commentaries and Subcommentaries)* (Taipei: Yiwen Press, 1973 reprint of original of 1815), chapter 59, p. 1027A. The *Zuo zhuan* is a chronicle that roughly parallels the much terser text of the *Spring and Autumn Annals*.

[11] For a listing of these in a text of the early imperial age, and a discussion of how clepsydra timings were related to the stars involved, see Christopher Cullen, "Huo Rong's Observation Programme of AD 102 and the *Han li* Solar Table," *Journal for the History of Astronomy* 38.1 (2007), 75–98, Table 1 and 88–90, also the discussions in Cullen, *The Foundations of Celestial Reckoning*, pp. 224–31, and Cullen, *Heavenly Numbers*, pp. 265–76.

[12] In the period we are discussing, it was assumed that when the sun returned to winter solstice (its furthest southern latitude) it also returned to the same position in relation to the stellar background. In modern terms, no distinction was made between the tropical year and the sidereal year. This

As well as helping to run the calendar, sky-watching also involved the interpretation of what was seen in the heavens, in terms of its significance for human affairs, particularly affairs of state. According to the analysis of Marc Kalinowski, out of 132 instances of divination reported in the *Zuo zhuan*, 19 are connected with the heavens. Of these, the most common type concerns the position of Jupiter, which follows a cycle close to twelve years in period. Next come interpretations of the seasonal risings and settings of the "Fire Star" (Antares). Finally there are also a few discussions of the significance of solar eclipses.[13] Dated records of the latter in both the *Zuo zhuan* and the *Spring and Autumn Annals*, amounting to several dozen, include indications of when an eclipse was total that render them useful to modern astronomers interested in such things as variations in the speed of terrestrial rotation.[14]

The celestial phenomena to be interpreted by diviners of the pre-imperial period included phenomena that were thought at the time to be basically regular and predictable, such as the motion of planets. One of the earliest such records, referring to events in 541 BCE, concerns the answer given to a king of Zhou, who asks his diviner Chang Hong 萇弘 about the likely fate of various feudal lords, and receives a response in terms of the twelve-year cycle of Jupiter.[15] A major treatise on planetary motion, almost certainly dating at least in part from the late Warring States period, was recovered from a tomb closed in the early Han, and given the title *Wu xing zhan* 五星占 (Prognostics on the five planets) by its modern editors.[16] As well as discussing the significance of the motion of the five visible planets through different regions of the heavens (each region being seen as correlated with one of the states in the world below) in terms of statecraft and military conflict, it gives tables for Jupiter, Saturn, and Venus with dates for their risings (and for Venus also settings) that go back into the late Warring States period, beginning in 246 BCE. Significantly, the text also allows for deviations from the predictions made, and gives interpretations of the significance of these in addition to the interpretations relating to regular motion.[17] Thus for Venus we read:

distinction, equivalent to the recognition of the precession of the equinoxes, was first made systematically in East Asia by Yu Xi 虞喜 (281–356 CE). I customarily use the term "solar cycle" for the conflated period in question.

13 Marc Kalinowski, "Diviners and Astrologers," in J. Lagerway and M. Kalinowski (eds.), *Early Chinese Religion, Park One* (Leiden: Brill, 2009), pp. 341–96.

14 See for instance F. R. Stephenson and L. V. Morrison, "Long-Term Fluctuations in the Earth's Rotation: 700 BC to AD 1990," *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences* 351 (1995), 165–202.

15 (*Zuo zhuan*, Duke Shao eleventh year, in Ruan Yuan 阮元 (1764–1849), *Shi san jing zhu shu* 十三經註疏, chapter 45, p. 785A.

16 See the commented translation in Christopher Cullen, "Wu xing zhan 五星占 'Prognostics of the Five Planets'," *SCIAMVS* 12 (2011), 193–249, and the wider discussion in Christopher Cullen, "Understanding the Planets in Ancient China: Prediction and Divination in the Wu xing zhan." *Early Science and Medicine* 16 (2011), 218–51.

17 Relatively frequent deviations from prediction are of course a natural product of the simpler types of predictive theory. In the context of his discussion of early Mesopotamian celestial divination, David

未宜出而出, 未宜入而入, 命曰失舍, 天下興兵, 所當之國亡.

> If it comes out [i.e. becomes visible] when it should not yet come out, or goes in when it should not yet go in, this is called "missing its lodge." Troops will rise up in the empire, and the corresponding country will be lost.[18]

Another document from the same early Han tomb as the *Wu xing zhan* contains depictions and brief discussions of a number of different types of comets. The complexity of this document, given that bright comets are relatively rare events, suggests that it may be based on a long tradition of such records passed on within lineages of diviners.[19]

The sources leave us in no doubt that there were experts who watched, recorded, and interpreted celestial phenomena in the pre-imperial age, but our knowledge of the identity and activities of these experts is at best patchy. Did any of them write books that came down to the imperial age? Writing around 90 BCE, the historian Sima Qian 司馬遷 lists the names of "those who have previously passed on the celestial reckonings" *xi zhi chuan tian shu zhe* 昔之傳天數者 (*Shi ji* 27, 1343). Of the fourteen names he gives, eight are from the Spring and Autumn Period, and the Warring States period that followed it. Two of the experts he names, "the Honourable Mr Gan" Gan Gong 甘公 (whose full name according to later commentators was Gan De 甘德) and Shi Shen 石申, are of particular relevance to this discussion. Sometimes texts refer to these two persons as Gan Shi 甘氏 "Mr Gan" and Shi Shi 石氏 "Mr Shi." Both were to lend their names to collections of omen interpretation frequently cited in later centuries. It is notable, however, that Sima Qian himself gives no more information about these diviners than their names and the feudal states in which they were active: he does not cite material said to have been authored by them. From this and other evidence it seems likely that no books of celestial lore by individual named authors had come down to his time from the pre-imperial period, quite probably because a tradition of individual authorship in technical matters had not yet been established.[20] Given that Sima Qian's responsibilities as a court official included the management of the calendar and the interpretation of celestial omens, as well as keeping the state archives, it is extremely unlikely that he would not have known of any such works, had any been in existence.

Brown has argued that producing omens of this kind for diviners to interpret can be seen as in some sense part of the function of early predictive schemes, rather than a defect: see David Brown, *Mesopotamian Planetary Astronomy-Astrology* (Groningen: Styx, 2000).

18 Cullen, "Wu xing zhan 五星占," p. 231.

19 On comets in early imperial times and before, see M. A. N. Loewe, "The Han View of Comets," in his *Divination, Mythology and Monarchy in Han China* (Cambridge: Cambridge University Press, 1994), pp. 61–84.

20 On the modes of transmission of technical literature in this period, see the discussion in Christopher Cullen, "The Suàn shù shū (筭數書) 'Writings on Reckoning': Rewriting the History of Early Chinese Mathematics in the Light of an Excavated Manuscript," *Historia Mathematica* 34.1 (2007), 10–44, 26–28.

WHY WATCH THE HEAVENS?

We have already seen something of the motivations for those in political power during the pre-imperial age to attend to celestial phenomena. Texts of the Qin and Han period give us explicit statements that clarify and enlarge on what we have already learned. Let us consider first the question of how transient and irregular celestial phenomena were to be interpreted. Each of the three *zheng shi* mentioned above contain a special monograph on this topic, for which the *Han shu* and the *Hou Han shu* both use what was to become the standard title, *Tian wen zhi* 天文志 (*Record of Celestial Patterns/Writing*).[21] In the preface to its monograph, the *Han shu* explains that all phenomena in the heaven are linked to events on earth:

> 其本在地, 而上發于天者也. 政失於此, 則變見於彼, 猶景之象形, 鄉之應聲. 是以明君覩之 . . .
>
> Their origins are on earth, but they manifest themselves on high in the heavens. If government is in error here, then an omen will appear there, just as a shadow follows a shape, or an echo responds to a sound. Therefore the enlightened ruler attends to them . . . *Han shu*, 26, 1273

After a general account of the way that the different regions of the heavens, the sun, moon, and planets may all provide indications of events on earth, the *Han shu* continues with a long list of omens recorded in the past, with notes of the subsequent events related to them. Thus for instance we read of an occultation of Antares by the moon:

> 陽朔元年七月壬子, 月犯心星. 占曰:「其國有憂, 若有大喪. 房, 心為宋, 今楚地.」十一月辛未, 楚王友薨.
>
> In the first year of the "Yang Beginning" reign period [24 BCE], 7th month, cyclical day 49, the moon trespassed on the Heart Star [Antares].[22] Prognostication: "The state will suffer sadness, or there will be great mourning. [The constellations] Chamber and Heart [in which the occultation occurred] stand for [the ancient state of] Song, which is now in the territory of the princedom of Chu."

[21] Compare the title *The Heavenly Writing* chosen by Francesca Rochberg, *The Heavenly Writing: Divination, Horoscopy, and Astronomy in Mesopotamian Culture* (Cambridge: Cambridge University Press, 2004), on the basis of the expression *mul.an* in Babylonian texts.

[22] The date given corresponds to 24 BCE, August 10, on which day there was indeed an occultation of Antares by the moon beginning around 16:50 local time and continuing until about 18:00 local time. Since sunset was not until 18:50, the occultation is likely to have been deduced by observing the positions of the moon and Antares as soon as the star became visible close to the moon after dusk. (Results obtained using Starry Night Pro™ software.)

> In the eleventh month, cyclical day 8, there passed away You, prince of Chu. (*Han shu* 26, 1310)

Immediately after that comes a description of a large meteor seen three years later, for which the prognostication was that there would be a diplomatic mission to the Xiongnu nomads, which did indeed occur the next year, after the Xiongnu leader had died. And so the list continues – rebellions, palace coups, and military emergencies are all clearly signaled in the sky. Clearly, the heavens were a giant display screen with warning lights that required constant watching and interpretation by full-time experts if the ruler wanted to avoid being taken unawares by events.

The importance of *Tian wen* was not therefore in doubt. But there was another way of looking at the sky, dealt with in another set of monographs in the *zheng shi*, and that was the concern of those who constructed and operated the mathematical systems called *li* 曆. *Li* can simply mean a calendar in the ordinary sense of an ordered sequence of dates, but it can also refer to the entire system of constants and algorithms that generates a calendar, and can thus be rendered as "calendrical system," or (following Nathan Sivin) "astronomical system." This last rendering takes account of the fact that a *li* does not simply enable us to calculate when months and years will begin and end, but also provides means for calculating the times when lunar eclipses might be expected, and even the motions of the planets. From 104 BCE we can trace a succession of such astronomical systems, promulgated with the full authority of the state; forty-eight separate systems were created from the beginning of empire up to its end in 1911.[23]

By the early imperial age, it was an accepted tradition that officials had been charged with responsibility for *li* from the time of the earliest known rulers. One text with canonical status, nowadays called the *Shu jing* 書經 "Book of Documents" recounted how in remote antiquity Emperor Yao 堯 had commissioned the fathers and sons of two families to observe the passage of the year, and thus to *jing shou min shi* 敬授民時 "respectfully deliver the seasons to the people."[24] But why was calendrical astronomy so important? We may disregard the suggestion, common in China several decades ago when "historical materialism" was obligatory, that the needs of an agricultural society made an accurate calendar essential. No peasant worth the name has ever needed a robed official with clean hands to tell him when to plant his crops, and the shifting months of a lunar calendar could only be the

[23] See Yabuuti Kiyosi 藪内清, *Chūgoku no tenmon rekihō* 中国の天文暦法 *(Chinese Mathematical Astronomy)* (Tokyo: Heibonsha, 1969), pp. 388–91. For full commented translations of the first three systems in this series, see Cullen, *The Foundations of Celestial Reckoning*, chapters 2, 3, and 4.

[24] See Bernhard Karlgren, *The Book of Documents* (Gothenburg: Elanders Boktryckeri Aktiebolag, 1950), pp. 2–3, and Joseph Needham and Wang Ling, *Science and Civilisation in China, vol. 3: Mathematics and the Sciences of the Heavens and the Earth* (Cambridge: Cambridge University Press, 1959), pp. 186–8. The usual title of this work in the period under discussion was *Shang shu* 尚書 "The honoured writings," or simply *Shu* 書 "The writings."

roughest of guides for such a purpose. A much likelier reason is provided by the pressing need to find the cosmically favored moment at which to conduct not only ritual actions taking place at the highest levels of the state, but also the ordinary affairs of daily life for ordinary people. Thus, for example, it became established practice for the emperor to pray for a good harvest at dawn on the day when winter solstice fell. If in fact winter solstice fell a few minutes outside that day (reckoned as beginning at midnight), the ritual would be ineffective. As for lower levels of society, in recent decades archaeology has recovered many examples of texts known as *ri shu* 日書 "day books," providing detailed guidance about what dates were favorable or unfavorable for activities ranging from getting married to seeing a magistrate on business.[25] If the calendar was out of order, nothing would go right for anybody, from commoners right up to the emperor.

WHO WATCHED THE HEAVENS?

Our sources, which overwhelmingly represent the viewpoints of the mainly official persons who wrote or compiled them, take it for granted that the most important people to occupy themselves with celestial observations, and with analyzing the resultant records, will be officials. And there is certainly some truth in that. All imperial governments did maintain a staff of experts who were paid to watch the heavens, to record and interpret what they saw there, and to make calculations that as far as possible predicted how it would appear at future dates. These are the people who generated the texts on which historical accounts are largely based.

In the *Hou Han shu* we have a listing of the staff of the state observatory (*ling tai* 靈台 "numinous terrace") in the period covered by this work, the first two centuries after the Common Era began.[26] The director of the observatory is rated at an annual salary of two hundred bushels of rice,[27] which places him in a state of modest prosperity, although his superior the Grand Clerk *Tai shi* 太史 (the rank once held by Sima Qian) is rated at three times as much. Below the director there are fourteen observers of the stars, two observers of the sun, three observers of the winds, twelve observers of

[25] See Donald Harper and Marc Kalinowski (eds.), *Books of Fate and Popular Culture in Early China. The Daybook Manuscripts of the Warring States, Qin, and Han* (Leiden: Brill, 2017).

[26] *Hou Han shu, zhi* 25, 3572.

[27] The "bushel" *shi* 石, was equivalent to about 20 litres. At UK prices in 2013, good quality rice sells for about £1.50/kg, so given that a litre of rice weighs close to a kg, a *shi* would be worth about £30, and the Director's salary rating would be worth about £6,000. Even allowing for the great difference between ancient and modern values of commodities, such calculations make Han officials look badly off, but as pointed out in Hans Bielenstein, *The Bureaucracy of Han Times* (Cambridge: Cambridge University Press, 1980), pp. 125–31, by the period we are discussing the nominal salary in grain was little more than a grade ranking. Part of the salary was paid in cash equivalent, and there were additional allowances and regular gifts in kind. As a result, the Director and his family were able to enjoy a standard of living far higher than that of most imperial subjects.

vapours, three observers of the sundial, and seven charged with observing the notes given by bells and musical tubes,[28] together with one janitor. The responsibilities laid on the Grand Clerk, assisted by his subordinates such as the observatory director and his staff, included producing the calendar for each new year, the choosing of fortunate days for state ceremonies, and the interpretation of omens.

No doubt the Grand Clerk and his team really did carry out many important activities in relation to celestial phenomena. But if we inspect the historical records closely, we can see clear signs that there were important actors in this area who were not part of the bureaucracy with relevant defined responsibilities related to the heavens, or who were even outside the world of officialdom altogether. As an example of the first group, we may consider the so-called "Expectant Officials," *dai zhao* 待詔, literally "[those] awaiting an edict." Persons so designated occur frequently in the record of astronomical controversy in Eastern (Later) Han. In the words of an Eastern Han commentator on the *Han shu* 諸以材技徵召, 未有正官, 故曰待詔 "all those summoned to serve by reason of their talents, but who do not yet have a regular official post are accordingly called '[those] awaiting an edict'" (*Han shu* 11, 340). Overall, the surprising thing about the records in the *Hou Han shu* is not just that Expectant Officials frequently comment on astronomical matters – more than that, they are heard from rather more often than those who hold substantive posts concerned with astronomy. One particularly striking example is Jia Kui 賈逵, who held the title of *Zuo zhong lang jiang* 左中郎將 "left (i.e. senior) leader of the palace gentlemen" which placed him at the head of those in "general service" positions (see *Hou Han shu, zhi* 25b, 3575). It was detailed evidence on lunar motion submitted by him that forced the Grand Clerk's observers to accept a new armillary instrument with a ring representing the ecliptic in 103 CE (*Hou Han shu, zhi* 2, 3030).[29] Examples could easily be multiplied, if space permitted.

But there was skill of an advanced kind well outside official circles, and there were times when the officials had no choice but to call on it. One striking example dates from the time of the creation of the Grand Inception system in 104 BCE. The account given in the *Shi ji* would seem to imply that responsibility for creation of the new system was largely in official hands. But the *Han shu* gives us extra information that throws a different light on matters:

> 姓等奏不能為算, 願募治曆者, 更造密度, 各自增減, 以造漢太初曆. 乃選治曆鄧平及長樂司馬可, 酒泉候宜君, 侍郎尊及與民間治曆者, 凡二十餘人, 方士唐都, 巴郡落下閎與焉. 都分天部, 而閎運算轉曆.

[28] If the cosmos was running smoothly, then these would sound the right notes at appropriate seasons.
[29] Cullen, *The Foundations of Celestial Reckoning*, pp. 385–90; Cullen, *Heavenly Numbers*, pp. 245–50.

> [The Senior Star Watcher She] Xing and the rest memorialized that they were unable to perform the calculations, and wished to summon experts on astronomical systems so as to achieve greater precision, with each [expert] making adjustments, so as to create a Han Grand Inception astronomical system. So they chose the experts on astronomical systems Deng Ping, Sima Ke from Changle, Yi Jun the marquis of Jiuquan, the Attendant Gentleman Zun, and experts on astronomical systems from amongst the people – more than twenty persons. Tang Du, a gentleman knowledgeable in techniques, and Luoxia Hong from Ba prefecture were among them. [Tang] Du, distinguished the divisions of the heavens, while [Luoxia] Hong carried out calculations to revise the astronomical system. (*Han shu* 21a, 975)[30]

As a final example of non-official expertise, we may consider Lang Yi 郎顗 (fl. ca. 133 CE) who was summoned to court under Shundi (r. 127–44 CE), and whose lengthy written submissions have been preserved in the *Hou Han shu* 後漢書, chapter 30b. His skills seem to have included hemerology, the interpretation of such transient celestial phenomena as parhelia, and unusual vapors in the night sky. His discussions of planetary omens and their significance show that he is capable of carrying out the quite complex calculations to predict where a planet should be seen on a given date, in accordance with the methods current in his day. But from where was this skill-set derived? It turns out that he was running a private college (inherited from his father) by the seaside in the Beihai 北海 region that taught the lore of the heavens as well as the classics:

隱居海畔, 延致學徒常數百人. 晝研精義, 夜占象度 . . ..

> He lived in retirement by the seashore, and often had several hundred visiting students. By day they researched the essential significance [of the classics], and at night they divined by the degrees of the observed phenomena. (*Hou Han shu* 30b, 1053)

This school of classics, celestial calculation, and divinatory skills was clearly a major center of learning quite outside official structures. For the Grand Inception reforms, the central government had called on the services of "over twenty" experts "from amongst the people." But by the time of Lang Yi at least, it appears that they could have called on hundreds had they wished.[31]

30 Cullen, *The Foundations of Celestial Reckoning*, pp. 369–70; Cullen, *Heavenly Numbers*, chapter 3.
31 For a discussion of Lang Yi, and of "outsiders" in connection with astronomy, see Cullen, *Heavenly Numbers*, pp. 293–301.

HOW DID THEY MAKE OBSERVATIONS?

For most of the imperial period, Chinese specialists in celestial phenomena used three basic types of instrument to make quantitative observations: the gnomon *biao* 表; the clepsydra *lou ke* 漏刻; and the armillary sphere *hun tian yi* 渾天儀. It was the early imperial age of Qin and Han that saw this repertoire assembled, and in this section I shall sketch what we know of that process.

Firstly, we may note how much can be done without instruments of any kind. The basic data of the calendar can be established to a quite high degree of precision by simple naked-eye observation at the beginning and end of a long enough period. If one fixes the date of a new moon correct to within a day (which is not really very difficult, if one looks for the first and last crescents on either side of the event), and records a similar event to the same precision ten years later after having counted the number of lunations intervening, then since there will be about 124 lunations in that period, simple division will fix the mean length of a lunation within less than 2 percent of a day, even if the errors in the two determinations are in the opposite sense. Two estimates of the date of the same event in the solar cycle (such as the rising of the sun in line with a local geographic feature) within five days either way can yield an estimate of the cycle length correct to within one day over a ten-year span; if the two dates are a century apart, then we are down to a fraction of a day. The first explicit mention of the use of a gnomon of standard length to measure the noon shadow of the sun near a solstice is not found until the *Huai nan zi* book, ca. 139 BCE.[32] But vaguer references are found in texts of the late Warring States period, and the occurrence in the *Spring and Autumn Annals* of two winter solstice dates very close to the real solstices strongly suggests the use of gnomons as far back as the seventh century BCE. Using those two dates, which are 133 years and 48578 days apart, produces a very good value of 365.25 days for the solar cycle.[33] This is in fact the earliest value to which we find explicit reference – again, in the *Huai nan zi* book.

As for the night sky, a naked-eye observer who watches the moon for several nights in a row will soon be able to point to where it will be found on future evenings, since the moon's daily displacement is roughly constant at

[32] The classic gnomon of early imperial times was a vertical pole of height 8 *chi* 尺. The *chi* (sometimes rendered as "foot") of Han times was about 23cm, 0.8 imperial foot. It was divided into ten *cun* 寸 "inches" (lit. "thumbs"). In later times, much larger gnomons were sometimes constructed. *Huai nan zi* chapter 3 tells us that the noon shadow of such a gnomon on the day of summer solstice is 1 *chi* 5 *cun*, 1.5 *chi*. This suggests a latitude of observation of something over 34 degrees, which is close to that of major centers of culture of the Qin and Han. See Cullen, *Astronomy and Mathematics*, pp. 101–6.

[33] The two dates given are those for the solstices near the start of the fifth year of Duke Xi (Dec 25, 656 BCE) and the twenty-first year of Duke Zhao (December 26, 523 BCE). The actual solstices fell on Dec 27 in both cases.

a "moon-step" of thirteen degrees, a quantity that anyone can measure against the sky by extending an arm to the full, and spreading thumb and index finger to their maximum extent. The moon itself is a useful half-degree template.

With a little practice, one can easily make a quantitative estimate of the position of the moon or a planet within the reference frame commonly used in early imperial times. This was the system of the *er shi ba xiu* 二十八宿 "twenty-eight lodges," a sequence of constellations of varying width forming a belt round the sky.[34] The belt of lodges follows the general region of both the celestial equator and the ecliptic, without following either very closely. The extent of the lodges was quantified using the unit known as the *du* 度 "measure," a quantity defined by the daily displacement of the sun against the background of the stars. Throughout the period discussed in this essay it was assumed that this displacement was constant, and as a result of this the number of *du* in a complete circuit of the sky was equal to the number of days in a complete solar cycle, around 365 ¼.[35] Each lodge defines a portion of the circuit of the sky running from a reference star at its western end, up to the reference star of the next lodge to the eastwards. In modern terms, a lodge corresponds to a slice of the celestial sphere in right ascension, so that the sphere is divided up like an orange with irregular segments.[36] Since, however, the system of the lodges appears to predate the notion of the celestial sphere in Qin and Han, its origins are better understood in terms of a less precise concept of "how far round the sky." Our earliest explicit listing of the widths of the lodges in a text is found in chapter 3 of the *Huai nan zi* book, and thus dates from a little before 139 BCE. As to how these widths were ascertained, the first clear statement appears in connection with events only a few decades later, in a passage of the *Han shu* referring to preparations for the great reform of 104 BCE, where Sima Qian and his colleagues are said to have

定東西, 立晷儀, 下漏刻, 以追二十八宿相距於四方 . . .

> . . . fixed east and west, erected instruments for observing shadows, and set water-clocks working, in order to find the extents of the 28 lodges in the four quarters [of the heavens]. (*Han shu* 21a, 975)[37]

In the general context of our knowledge of Qin and Han observational practice, this statement clearly implies that the widths of the lodges were

[34] I have explained my reasons for translating *xiu* 宿 as simply "lodge" without qualification in Christopher Cullen, "Translating 宿 *sukh/xiu and 舍 *lhah/she: 'lunar lodges', or just plain 'lodges'?," *East Asian Science, Technology and Medicine* 33 (2011), 80–92.

[35] As explained above (footnote 9), in our period of interest the modern concepts of the tropical and sidereal years were not distinguished.

[36] The lodges differ considerably in width.

[37] Cullen, *The Foundations of Celestial Reckoning*, p. 369.

found by timing how long it took for each lodge to pass by a north-south sight line of gnomons. Observations over a large part of a year would enable the total 365 ¼ *du* circuit of the sky to be apportioned amongst the lodges in proportion to their transit times. In later times, we know of multi-vessel clepsydras that could provide a steady rate of flow for a whole day and night.[38] The only such devices surviving from the Han are relatively small single-vessel outflow devices.[39] This would not have rendered them useless for astronomical measurements – quite the reverse, in fact, since a clepsydra that ran dry in (say) fifteen minutes (during which time the heavens rotate by under four degrees) and was immediately refilled would provide an indefinitely repeating time interval that was easy to reproduce accurately as well as to subdivide.[40]

But what about devices that would have enabled measurements of angles on the heavens in any plane, not just in right ascension? The earliest object with a graduated circumference connected with the heavens is a lacquer disk from a tomb of 165 BCE, marked round its rim with the names of the twenty-eight lodges, and bearing graduations corresponding to their width in *du*. It has been speculated that this "lodge dial" might have been used to measure angles in non-equatorial planes, but it does not seem well adapted to such a purpose and seems more likely to have been a calculating device of some kind.[41] A later citation of a memorial submitted in 52 BCE refers to the use of a "diagram instrument" *tu yi* 圖儀, for making observations of the motions of the sun and moon round the system of twenty-eight lodges. This may have been a reference to the use of an observational instrument related to the lodge dial.[42]

It is only when we move away from measurements of right ascension, which can be done by transit timing, to more general measurements of angular distances in the heavens, that we are forced to assume the presence of some kind of graduated ring capable of being placed in a number of different

[38] Needham and Wang Ling, *Science and Civilisation in China*, vol. 3, 320, translate one fragmentary text by Zhang Heng 張衡 (78–139 CE) as referring to a multi-vessel clepsydra; Henri Maspero, "Les instruments astronomiques des Chinois au temps des Han," *Mélanges chinois et bouddhiques* 6 (1939), 183–356, 202 is rather more cautious. Explicit descriptions of such devices are not found before the eleventh century, although it is tempting to read them back into earlier texts.

[39] The three devices described and illustrated in Chinese Academy of Social Sciences Institute of Archaeology, *Zhong guo gu dai tian wen wen wu tu ji* 中国古代天文文物图集 *(Collected Illustrations of Objects Connected wth Astronomy in Ancient China)* (Beijing: Wenwu Press, 1980), pp. 5, 39–41, and 116–17 are between 22 and 24cm in height, with internal diameters ranging from about 9 to 19cm.

[40] I have verified experimentally that observations of the transits of stars using a simple gnomon and clepsydra make it possible to obtain measurements of the differences of right ascension of stars to an accuracy similar to that found in ancient Chinese sources: see Christopher Cullen, "Early Chinese Measurements of Right Ascension before the Armillary Sphere," paper presented at the First International Conference on the History of Chinese Science (Louvain, 1982), and Cullen, *Heavenly Numbers*, pp. 194–8.

[41] Christopher Cullen, "Some Further Points on the shih," *Early China* 6 (1981), 31–46, 35.

[42] See Cullen, *The Foundations of Celestial Reckoning*, p. 388 and Cullen, *Heavenly Numbers*, pp. 202–7.

planes. Careful analysis of a set of data found in later sources, giving the distances from the north celestial pole of the standard stars of the twenty-eight lodges suggests that such observations may have been made around 70 BCE.[43] That would definitely have required the use of some kind of simple graduated ring with sights, capable of being set in a plane perpendicular to the celestial equator. By the second century CE, there is no doubt that such rings were in use, and that they had been brought together in instruments largely identical to the "armillary spheres" of the west – the term in China was *hun tian yi* 渾天儀 "spherical heaven instrument." As we have already seen, a memorial of Jia Kui 賈逵 in 92 CE insisted that the devices then in use by the officials of the Grand Clerk should have an ecliptic ring added to them, to facilitate observations of the motions of the sun and moon.

At this point we may add a note on what is a relatively minor topic in the context of this essay, and that is the question of cosmography. In the ancient Hellenic world (or at least in commonly accepted accounts of its history) the question of what might be the shapes, sizes, dispositions, and motions of the heavens and the earth played an important role in debates about the heavens. In the world of Qin and Han similar discussions took place, but as something of a sideshow in comparison with other topics. Briefly we may say that two main cosmographical pictures were to some extent in conflict. The first and earlier of these was the *gai tian* 蓋天 "heaven as a chariot-umbrella" view. According to this, the human race (or at least that part of it known to Qin and Han) lived on a basically flat earth, approximately 103,000 *li* 里 (about 55,000km) from a position corresponding to the north geographical pole. The heavens were a vast disk 80,000 *li* above the earth, and rotated once a day around an axis through the pole, carrying with it the heavenly bodies. The apparent rising and setting of those bodies was due to their passing into and out of a critical range of 167,000 *li*, beyond which human sight failed. Winter came when the sun moved far out from the pole on the heaven-disk, and summer came when it moved in closer. At the pole itself, there was said to be day for six months and night for six months – just as happens in reality. An account of this cosmography, with details of the calculations behind the dimensions, was given in the book *Zhou bi* 周髀 "The gnomon of Zhou," probably assembled in final form around 10 CE, but there are signs of similar views in the pre-Qin period.[44]

Such a concept of heavens as a rotating disk is not incompatible with the transit timing methods using gnomons and clepsydas that gave dimensions to the lodge system, since this only concerns itself with the meridian passing from the pole through the observer. But outside that limited zone, it is not

[43] Yabuuti Kiyosi 藪内清, *Chūgoku no tenmon rekihō* 中国の天文暦法 *(Chinese Mathematical Astronomy)*, pp. 54–64; also Xiaochun Sun and Jacob Kistemaker, *The Chinese Sky during the Han: Constellating Stars and Society* (Leiden: Brill, 1997), pp. 56–67.

[44] For a detailed study and translation of this work, see Cullen, *Astronomy and Mathematics*.

difficult to show that the *gai tian* produces strange results. Thus, for instance, at some time during the period between 9 and 23 CE, Huan Tan 桓譚 pointed out to his friend Yang Xiong 楊雄 that on the *gai tian* view, an east–west line through an observer far from the pole would divide the sun's daily track into two unequal portions. Thus at the equinoxes, when the sun rises (or as the *gai tian* has it, comes within range of sight) due east and sets due west, the lengths of day and night could not be equal.[45] It was not difficult to find other similar criticisms. By the time of Cai Yong 蔡邕 around 180 CE, he could write:

周髀數術具存, 考驗天狀, 多所違失, 故史官不用.

> The calculation methods of the *Zhou bi* are all extant, but if they are compared with the celestial phenomena they are mostly in error, so the [Grand] Clerk's officials do not use them. (*Hou Han shu, zhi* 10, 3215, commentary)

He goes on, however, to say:

唯渾天者近得其情, 今史官所用候臺銅儀, 則其法也.

> It is only the *Hun tian* that gets near the truth. Now the bronze instruments used by the [Grand] Clerk's officials on the observation platform follow its pattern. (*Hou Han shu, zhi* 10, 3215, commentary)

The *Hun tian* "spherical heaven" here is clearly not the instrument itself, but a cosmography such as the *gai tian* view set out in the *Zhou bi*. Like the *gai tian* it assumed a flat earth – but in accordance with the model given by the rings of an armillary sphere, it saw the heavens as a giant sphere rotating daily about an inclined axis (the inclination to the horizon being what in modern terms would be the latitude of the Han dynasty observer), carrying with it the heavenly bodies. Risings and settings now meant simply rising up over the edge of the earth or sinking down out of sight again. The Han observer was assumed to be at or close to the center of the sphere, and no attempt was made to depict polar conditions. Essentially, the universe was a planetarium giving a display to a group of privileged observers near the center of the earth.[46] The Grand Clerk's officials did not need to think explicitly about cosmography, since it was in fact inscribed into the instruments they used

[45] See the translation in ibid., pp. 59–60.

[46] Since the radius of the celestial sphere was typically thought to be much larger than known geographical distances, most human observers were relatively close to the center in practice. Zhang Heng suggested that the diameter of the celestial sphere was 232,300 *li*, about 100,000 km even if we understand the number *yi* 億 used here in the smaller of its two senses, 100,000 rather than 1,000,000. That was ten times the maximum north–south extent of the Han empire. See *Hou*

each night. Nor, as we shall see, did explicit questions of cosmography enter into the actual calculations those observers might make to predict the future state of the heavens, or to retrodict how they had appeared in the past.

HOW DID THEY MAKE CALCULATIONS?

To turn to calculation: we are fortunate to have full specifications for the *li* used for three out of the four centuries of the Han empire. The system used from 104 BCE is set out in full (perhaps with additions from around 10 CE to specify planetary motions) in *Han shu*, chapter 21b. The system that replaced it in 85 CE is set out in *Hou Han shu, zhi* 3b. A further system created under the late Han, but not used until after its fall, is recorded in *Jin shu*, chapter 17. The interested reader may consult my complete translations of these texts.[47] Here I shall give only a basic sketch of how such systems functioned.

In the Han period and for over a millennium thereafter, all *li* shared one basic feature, in that they began calculations from a stated *li yuan* 曆元 "system origin." This was a point in the past, perhaps in the very remote past, when it was posited that all significant elements in the system had been at initial conditions: the time of day was midnight, the day then beginning was the first of the sexagenary cycle, a conjunction of sun and moon fell at that midnight, which began the first month of the count for that year,[48] and all five visible planets were in conjunction with the sun. Particularly if the *li yuan* was remote from the time when the system was in use, it was not necessary to show that such conditions had in fact obtained at the specified instant. What mattered, rather, was whether using that system origin produced accurate results in the period to which the system was applied.

To predict the state of the heavens at later dates, it was only necessary to find the number of years elapsed since system origin, and count off the number of complete cycles of lunations, solar cycles, and planetary periods that had elapsed, to find out the position of the current instant in the various cycles currently in progress. The specifications for cycle lengths had two important features:

(a) For convenience of calculation, fractions were avoided by the use of a procedure we may call "scaling": the principal number given for a quantity was not its actual magnitude, but instead that magnitude multiplied by a scaling

Han shu, zhi 10, 3215, commentary; for further discussion, see Cullen, *Astronomy and Mathematics*, pp. 61–6.

47 All systems created in the Han dynasty are fully translated and explained in Cullen, *The Foundations of Celestial Reckoning*.

48 For the purposes of calculating *li*, the first month of the sequence used, the "astronomical" first month, was two months before the first month of the civil year, which after 104 BCE began with a new moon in late January or February. The astronomical first month thus contained the winter solstice, the start of the solar cycle.

factor large enough to eliminate fractions. So, for instance, in the system promulgated in 104 BCE, the length of (in modern terms) the mean synodic month was specified by the two quantities 2392 for the *yue fa* 月法 "lunation factor," representing the basic magnitude, and 81 for the *ri fa* 日法 "day factor," representing the scale. In modern terms, the system is working with a mean synodic month whose length in days is: 2392/81 = 29 + 43/81 = 29.5309 to six significant figures.[49]
(b) To simplify the process of finding the current position in the cycle of interest, cycle lengths were specified so that the basic cycles had common multiples – what Nathan Sivin has called "resonance periods" – that were conveniently small. An important example of such a common multiple, used in a number of *li*, was the *zhang* 章 "Rule [cycle]," made up of precisely 19 solar cycles and 235 mean lunations.[50] In the system of 104 BCE the longest resonance period involving the sun and moon (but not the planets) was the *yuan* 元 "Origin" of 4,617 years, after which initial luni-solar conditions and the sexagenary day cycle recurred. The existence of that period meant that if one was many years from the system origin, calculations could immediately be simplified by casting out whole Origins from the years elapsed.

All calculations were specified algorithmically, as step-by-step instructions to calculate named quantities, which in their turn might be used for further calculations. It is not difficult to translate such instructions into modern programming languages, or to reproduce them in spreadsheet form.[51] As for the physical means of calculation used, there are no explicit references to this in the text of any early *li*. We do know, however that the abacus was not in use in China until centuries later than the Qin/Han period, and although the data for calculations and the results obtained might be represented using writing, the actual process of calculation was not carried out in writing until the seventeenth century. Instead, as other texts of the period explain, numbers were represented by coordinated arrays of counting rods arranged on a flat surface, with each digit in a decimal place-value system represented by a small group of rods. As with the abacus, the process of calculation effaced the starting data.[52]

As already stated, one feature not found in any ancient *li* was the use of a geometrical model of the cosmos as a basis for calculation. Of course, there

[49] A modern value for the mean synodic month is 29.5306 days (to the same precision); the two values agree to within 0.0003 day, just under half a minute.

[50] These cycles, equivalent to the so-called "Metonic cycle" known in the ancient Mediterranean world, implied that during nineteen years there would have to be seven intercalary months, since 12*19 + 7 = 235. See Cullen, *Heavenly Numbers*, chapter 1, and on Meton, G. J. Toomer, "Meton," in F. L. Holmes and C. C. Gillespie (eds.) *Dictionary of Scientific Biography* (New York: Scribner's, 1974), pp. 337–9.

[51] Christopher Cullen, "Translating Ancient Chinese Astronomical Systems with Excel: How Not to Stew the Strawberries?," *Journal for the History of Astronomy* 36 (2005), 336–8.

[52] Elementary accounts of counting rod procedures can be found in (for instance) Ho Peng Yoke, *Li, Qi and Shu: An Introduction to Science and Civilization in China* (Hong Kong: Hong Kong University Press, 1985), pp. 55–70.

are some problems posed by the use of the concept of the celestial sphere that seem to demand such procedures, such as the need to predict (using modern terms) the changing right ascension of a body (measured with reference to the celestial equator) from the fact that it is moving uniformly along the ecliptic, a great circle inclined at about 23 ½ *du* to the equator. While early imperial specialists in *li* were clearly aware of this problem, the fact that the tools available to them did not include spherical trigonometry led them to resort to more makeshift methods for this purpose, apparently based on measurements made on model celestial spheres.[53]

DISAGREEMENTS AND HOW TO SETTLE THEM

The fact that a *li* was adopted at a given date and later replaced by a different system clearly points to the fact that someone, somewhere must have made some decisions. And given the importance of such changes for the imperial government, such decisions are likely to have been made (or at least approved) at a high level in state structures. But who actually made such decisions, and on what basis, and by means of what process of deliberation?

Some scholars have put forward a general view under which decision-making in the early imperial period consisted essentially of proposals being sent up from a relatively low level through vertical lines of communication, with the decisions being ultimately made at the level of the emperor himself, and issued to his submissive vassals for implementation.[54] Now there is no doubt that under some circumstances personal imperial influence could be strong, even decisive. I have told the story of how the decision to carry out a reform of the official *li* in 104 BCE depended crucially on the way that Emperor Wu was persuaded that doing so would be helpful in realizing his personal project to communicate with the spirits and become an immortal.[55] But the more general pattern, exemplified by the summoning of experts to settle details of the implementation of the reform, as set out earlier, was one of consultation and discussion through basically horizontal structures set to work by higher authority. In some cases this might involve long-term programmes of observation designed to test the effectiveness of proposed changes to the system in use. Our earliest example of this comes only twenty-

[53] See Christopher Cullen, "Seeing the Appearances: Ecliptic and Equator in the Eastern Han," *Zi ran ke xue shi yan jiu* 自然科學史研究 *(Studies in the History of Natural Sciences)* 19.4 (2000), 352–82.

[54] See the views of Nathan Sivin in G. E. R. Lloyd and Nathan Sivin, *The Way and the Word: Science and Medicine in Early China and Greece* (New Haven, CT: Yale University Press, 2002), chapter 6, 239–51.

[55] See Cullen, *Heavenly Numbers*, chapter 3, and Christopher Cullen, "Motivations for Scientific Change in Ancient China: Emperor Wu and the Grand Inception Astronomical Reforms of 104 BC," *Journal for the History of Astronomy* 24.3 (1993), 185–203.

six years after the 104 BCE reform, in 78 BCE, when the person who held the office of Grand Clerk suggested that the new system had been improperly constructed. The decision against him was arrived at only after a two-stage programme extending over three years, in which his predictions and those of the official system were carefully compared with observation (*Han shu* 21a, 978).[56]

By the time we reach the Eastern Han, the records of such evidence-based decision-making are quite extensive.[57] When in 85 CE the system adopted in 104 BCE was replaced by a new system, there had already been a long sequence of discussions going back to 32 CE, when the first memorials were submitted pointing out that prediction no longer matched observation. A faulty date prediction for the lunar eclipse observed on September 8, 62 CE resulted in further discussions, and some elements of a new system were submitted to experimental trial. It was not until thirteen years later that a completely new system was promulgated, and even then the work of those commissioned to carry out this project was corrected at the initiative of Jia Kui, in company with ten other named persons (*Hou Han shu, zhi* 2, 3025–7).[58]

As the Eastern Han proceeded, there are several fairly detailed records of occasions on which considerable numbers of officials, by no means all of them experts in *li*, gathered to hear discussions and make decisions on problems with the system in use and proposals for change. I have analyzed some of these debates in historical context, leading up to the great debate on *li* that took place in 175 CE, in which Cai Yong led for the defence of the official system against two men who had demanded that those responsible for running the system should be subjected to criminal sanctions for incompetence.[59] It appears that openness in discussion and the testing of theory against observational data were favored modes of dispute settlement amongst specialists in *li* in this period.

POSTSCRIPT

I have no space here to attempt a comparison of the story set out in this essay with similar accounts dealing with the same topic in the ancient Mesopotamian, Indian or Greek worlds. But in such comparisons

[56] Cullen, *The Foundations of Celestial Reckoning*, pp. 371–3; Cullen, *Heavenly Numbers*, pp. 109–10.

[57] For a translation of a collection of documents detailing the debates that led up to these decisions, see Cullen, *The Foundations of Celestial Reckoning*, chapter 5; the content of that collection is analyzed in Cullen, *Heavenly Numbers*, chapters 6 and 7. Christopher Cullen, “Actors, Networks and ‘Disturbing Spectacles’ in Institutional Science: 2nd Century Chinese Debates on Astronomy,” *Antiquorum Philosophia* 1 (2007), 237–68, may also be consulted.

[58] Cullen, *The Foundations of Celestial Reckoning*, pp. 376–81; Cullen, *Heavenly Numbers*, p. 236.

[59] See Cullen, *The Foundations of Celestial Reckoning*, pp. 403–10, and Cullen, *Heavenly Numbers*, pp. 310–23.

I suggest two points are likely to emerge. Firstly the richness and thoroughness of the documentation available from China is unparalleled in the period I have discussed. Secondly – and paradoxically – there is the great advantage in China that the story is documented from a stage where some aspects of the topic are at a relatively simple state of development. When the documented story of celestial learning in Qin and Han begins, there is no knowledge of the celestial sphere. We can even see the discovery of the importance of the ecliptic as a frame of reference being argued out before our eyes for the first time by a man like Jia Kui. While in later times intercultural contacts and transmissions weave the Chinese story inextricably into the world story, the very fact of early China's separateness and difference serves to render it all the more interesting and illuminating.

INDEX