# Spatial Thinking and External Representation

Towards a Historical Epistemology of Space

# Edition Open Access



The Edition Open Access (EOA) platform was founded to bring together publication initiatives seeking to disseminate the results of scholarly work in a format that combines traditional publications with the digital medium. It currently hosts the open-access publications of the "Max Planck Research Library for the History and Development of Knowledge" (MPRL) and "Edition Open Sources" (EOS). EOA is open to host other open access initiatives similar in conception and spirit, in accordance with the *Berlin Declaration on Open Access to Knowledge* in the sciences and humanities, which was launched by the Max Planck Society in 2003.

By combining the advantages of traditional publications and the digital medium, the platform offers a new way of publishing research and of studying historical topics or current issues in relation to primary materials that are otherwise not easily available. The volumes are available both as printed books and as online open access publications. They are directed at scholars and students of various disciplines, as well as at a broader public interested in how science shapes our world.

# Spatial Thinking and External Representation

## Towards a Historical Epistemology of Space

Matthias Schemmel (ed.)

**Studies 8**

Max Planck Research Library for the History and Development of Knowledge
Studies 8

Communicated by:
Vincenzo de Risi

Cover Image:
Surveyor's field map, Ur III period (twenty-first century BCE), MVN 10, 214.
Courtesy of the Schøyen Collection. See Chapter 3 for discussion.

ISBN 978-3-945561-08-9
First published 2016 by Edition Open Access,
Max Planck Institute for the History of Science
http://www.edition-open-access.de
Printed and distributed by
PRO BUSINESS digital printing Deutschland GmbH, Berlin


The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

**Max Planck Research Library for the History and Development of Knowledge**

The Max Planck Research Library for the History and Development of Knowledge comprises the subseries, Studies, Proceedings and Textbooks. They present original scientific work submitted under the scholarly responsibility of members of the Scientific Board and their academic peers. The initiative is currently supported by research departments of three Max Planck Institutes: the MPI for the History of Science, the Fritz Haber Institute of the MPG and the MPI for Gravitational Physics (Albert Einstein Institute). The publications of the Studies series are dedicated to key subjects in the history and development of knowledge, bringing together perspectives from different fields and combining source-based empirical research with theoretically guided approaches. The Proceedings series presents the results of scientific meetings on current issues and supports, at the same time, further cooperation on these issues by offering an electronic platform with further resources and the possibility for comments and interactions. The Textbooks volumes are prepared by leading experts in the relevant fields.

**Scientific Board**

To the memory of Peter Damerow (1939–2011)

# Contents


| | | |
|---|---|---|
| | **Preface** | xiii |
| **1** | **Towards a Historical Epistemology of Space: An Introduction** | |
| | *Matthias Schemmel* | 1 |
| 1.1 | The challenge of a historical epistemology of space | 1 |
| 1.2 | Natural conditions of spatial cognition | 5 |
| 1.3 | Culturally shared mental models of space | 10 |
| 1.4 | Social control of space and metrization | 14 |
| 1.5 | Context-independence of mental models resulting from reflection | 17 |
| 1.6 | The expansion of experiential spaces over history | 20 |
| 1.7 | The decline of an autonomous concept of space | 23 |
| 1.8 | Concluding remarks | 26 |
| | Bibliography | 27 |
| **2** | **Spatial Concepts in Non-Literate Societies: Language and Practice in Eipo and Dene Chipewyan** | |
| | *Martin Thiering and Wulf Schiefenhövel* | 35 |
| 2.1 | Introduction | 35 |
| 2.2 | Theoretical frame | 37 |
| 2.3 | Anthropological and linguistic background: Dene Chipewyan | 45 |
| 2.4 | Anthropological and linguistic background: Eipo | 55 |
| 2.5 | Center, periphery and distance in Eipo | 67 |
| 2.6 | Representations of spaces in Eipo and Dene Chipewyan | 74 |
| 2.7 | Conclusion | 84 |
| | List of linguistic abbreviations | 86 |
| | Bibliography | 86 |
| **3** | **The Impact of Notation Systems: From the Practical Knowledge of Surveyors to Babylonian Geometry** | |
| | *Peter Damerow* | 93 |
| 3.1 | Introduction | 93 |
| 3.2 | The origin of notation systems in Mesopotamia in the third millennium BCE | 95 |
| 3.3 | The problem of multiplication and of the calculation of areas of fields | 100 |
| 3.4 | The invention of the *surveyors' formula* | 106 |
| 3.5 | Sophisticated surveying techniques in the Ur III period | 108 |
| 3.6 | From context-dependent to abstract notations of quantities | 109 |
| 3.7 | The heritage of the surveyors in Babylonian mathematics | 113 |
| 3.8 | Conclusion | 117 |
| | Bibliography | 118 |

**4 Theoretical Reflections on Elementary Actions and Instrumental Practices: The Example of the *Mohist Canon***
*William G. Boltz and Matthias Schemmel* ........ 121
4.1 Elementary actions, instrumental practices, and theoretical knowledge ........ 121
4.2 The *Mohist Canon* ........ 124
4.3 Magnitude, filling out, and interstice ........ 126
4.4 Spatial extent and duration ........ 129
4.5 Instruments and arrangements ........ 132
4.6 The epistemic status of Mohist spatial knowledge ........ 137
Bibliography ........ 143

**5 Cosmology and Epistemology: A Comparison between Aristotle's and Ptolemy's Approaches to Geocentrism**
*Pietro Daniel Omodeo and Irina Tupikova* ........ 145
5.1 Introduction ........ 145
5.2 Aristotle ........ 147
5.3 Ptolemy ........ 153
5.4 Conclusions and prospects ........ 169
Bibliography ........ 173

**6 Space and Matter in Early Modern Science: The Impenetrability of Matter**
*Peter Damerow* ........ 175
6.1 The character of early modern science ........ 175
6.2 Ancient atomism ........ 176
6.3 The revival of atomism in the renaissance ........ 178
6.4 Consequences of mechanical models ........ 179
6.5 The rationalist program ........ 181
6.6 Newton's hidden atomism and the problem of force ........ 183
6.7 Euler's solution ........ 185
6.8 Kant's anti-atomistic solution ........ 186
6.9 The impact of Kant's criticism ........ 188
Bibliography ........ 189

**7 Experience and Representation in Modern Physics: The Reshaping of Space**
*Alexander Blum, Jürgen Renn and Matthias Schemmel* ........ 191
7.1 Introduction ........ 191
7.2 The relation between space and time before relativity ........ 194
7.3 The impact of electromagnetism on the classical concepts of space and time ........ 197
7.4 The re-inclusion of gravitation into the concepts of space and time ........ 201
7.5 The role of space and time in quantum theory ........ 204
7.6 Conclusion ........ 206
Bibliography ........ 210

**Index** ........ 213

# Preface

Spatial thinking plays a central role in the life of individuals as well as whole societies. It ranges from everyday orientation in our living environment to the social organization of place and space, and the structuring of a huge corpus of experiential knowledge by means of theoretical concepts in modern science. Spatial knowledge thereby takes on different forms in different contexts, and it does so depending on the spatial experiences accounted for and the available means for its external representation. From this perspective, scientific spatial knowledge is but one form of spatial knowledge and does not represent a stratum independent of non-scientific knowledge. Science is not only based on the specific kind of knowledge expressed in theoretical texts, but essentially involves a broader knowledge base comprising all we have to know in order to master our environment, our technology as well as the specific equipment necessary to gain and validate scientific knowledge. It is part of a hierarchically structured architecture of knowledge which develops over history, just as experiences and means of representation vary over societies and history.

This volume presents and analyses manifestations of spatial thinking in various societal and historical circumstances: in the language and practices of recent non-literate societies, in the administrative institutions of early civilizations, in discursive contexts of ancient Greece and China, in early modern natural philosophy and metaphysics, and in twentieth-century physics. It discusses the historical and structural relations of the different forms of spatial knowledge and thereby attempts to address the question of the epistemic status of this knowledge. The exemplary cases discussed in the different chapters give no exhaustive account of spatial knowledge in human history, but are chosen to highlight important aspects of the cultural development of spatial knowledge. There are various other important topics (e.g. the history of cartography or the history of perspective) that could further contribute to the project of a historical epistemology of space as presented in this book.

This book presents results of the working group *The Historical Epistemology of Space*, conducted jointly by the Max Planck Institute for the History of Science in Berlin and the Humboldt University of Berlin in the framework of the project cluster TOPOI in the period 2008–2012. All contributions are authored by members of the group, its fellows and close collaborators. As the title of the book indicates, a conceptual focus of the group's research was external, or material, knowledge representations, by means of which knowledge is transmitted from one generation to the next, but also between cultures, thus producing continuity in the historical development of thinking. Since the means of knowledge representation at the same time serve as tools for thinking, they propel the interaction of experience and reflection in the historical development of spatial knowledge, a topic recurring in the various contexts presented in the different chapters of this book.

The book starts with a survey of the overall topic, *the historical epistemology of space*, specifying structures of spatial knowledge under different historical and cultural conditions and characterizing their epistemic status (Chapter 1). In the following chapters, different forms of spatial knowledge are presented through exemplary studies, which constituted the core of the group's research.

*Spatial concepts in non-literate societies* are discussed by comparing spatial languages and practices in Eipo and Dene Chipewyan, two independent, recent non-literate societies (Chapter 2). The analysis of these two societies is largely based on fieldwork that had previously been carried out by the authors themselves. Questions of the universality and culture-dependence of spatial thinking in societies that codify spatial knowledge almost exclusively by means of spoken language and joint action are addressed.

*The impact of notation systems* on collectively shared spatial knowledge is discussed with regard to the emergence of the systematic use of signs conveying arithmetical and lexical meaning as a means of knowledge representation (Chapter 3). Such changes took place in different early civilizations and the chapter pursues the development in Mesopotamia from the practical knowledge of surveyors at the beginning of the third millennium BCE to Babylonian geometry in the mid-second millennium BCE.

*Theoretical reflections on elementary actions and instrumental practices* are discussed using the example of spatial, temporal and material concepts documented in the *Mohist Canon*, a theoretical text from Warring States China, ca. 300 BCE (Chapter 4). In particular, the study allows comparative questions to be addressed concerning the independent emergence of theoretical knowledge traditions in ancient Greece and China.

The relation between *cosmology and epistemology* is studied by comparing the different approaches to arguing for the centrality of the earth of the two major classical authorities on cosmology, Aristotle and Ptolemy, and by discussing aspects of their reception up to early modern times (Chapter 5). It is shown how the focus on different parts of experiential knowledge and the use of different means of knowledge representation leads to divergent theoretical constructions in arguing for the same result: Aristotle's approach proceeds from the physical explanation of terrestrial phenomena to the cosmological realm, while Ptolemy follows the opposite direction, starting from mathematical-cosmological considerations and astronomical observations.

Concepts of *space and matter in early modern science* are discussed as a case of theoretical reflection in the context of early modern natural philosophy and mathematics (Chapter 6). Focusing on attempts to distinguish matter from space by assuming that its essential property is impenetrability, the chapter is particularly concerned with an analysis of the empirical foundations of metaphysical concept and system building.

*Experience and representation in disciplinarily structured science* are discussed by delineating the fundamental changes in the concepts of space and time brought about by the advanced formalism of twentieth-century physics, which enabled the integration of a growing corpus of experiential knowledge (Chapter 7). In particular the chapter addresses the question of why certain parts of experiential knowledge had an impact on concepts of space and time, while other parts did not.

This work was supported by the TOPOI project cluster and the Max Planck Institute for the History of Science in Berlin. The book is dedicated to the memory of Peter Damerow (1939–2011), whose support was crucial in shaping the project.

Matthias Schemmel
Berlin, March 2016

# Chapter 1
# Towards a Historical Epistemology of Space: An Introduction

*Matthias Schemmel*

## 1.1 The challenge of a historical epistemology of space

In the history of Western epistemological thought, there is a long tradition of dividing human knowledge into a purely rational part, independent of any experience in the outer world, and an experiential part.[1] Many aspects of spatial knowledge have traditionally been claimed to belong to the rational part. From the Pythagorean-Platonic claims about the ideal existence of geometrical figures, via early modern rationalistic ideas of deriving properties of space from pure reasoning, to the axiomatic deduction of properties of space in the logical positivism of the early twentieth century and later constructivist philosophies, attempts have been made, on very different grounds, to draw a clear-cut line between what is known of space prior to experience and what spatial knowledge is derived from experience.[2] Immanuel Kant's description of space as a pure form of intuition was particularly influential. Theorems from geometry are among Kant's prime paradigms for the existence of synthetic *a priori* judgments.[3] In his *Metaphysical Foundations of Natural Science*, Kant applies his program of isolating the *a priori* part of knowledge to the science of his time.[4]

The historical epistemology of space is similarly concerned with identifying the different sources of spatial knowledge. At the same time it is based on a thoroughly genetic, or developmental, view of cognition. According to this view, experiential knowledge participates in the construction of cognitive structures, which in turn constitute the basis for further experience. From this viewpoint a static separation between preformed structures of cognition and contingent experiences is impossible. Or rather, it is possible only in the snapshot image of a 'cognitive subject'. If the idea of a foundation of human knowledge – and scientific knowledge in particular – is justified, then this foundation must consequently lie in the

[1] This introductory chapter is based on the book: Matthias Schemmel, *Historical Epistemology of Space: From Primate Cognition to Spacetime Physics*, Springer, Cham, 2016.

[2] For Plato, see, for instance, the discussion on geometry in *Republic*, 526c 9 – 527c 11. A prominent rationalistic treatment of space is found in René Descartes' *Principles of Philosophy*, Part 2, in particular §§ 8–21, Descartes 1644, 37–44. For an English translation, see Descartes 1984, 42–49. An example from the early twentieth century of the division of spatial knowledge into an *a priori* and an experiential part is Carnap 1922, 62–67, who distinguishes formal, intuitive, and physical space, of which only the first is completely independent of experience; at the same time Carnap argues that the cognitive structure given by a topological space of infinitely many dimensions is the precondition for any kind of spatial experience. For a constructivist argument about the *a priori* nature of Euclidean space, see, for instance, Lorenzen (1984), who wants to show "how the Euclidean theorems can be proven in Plato's sense solely from definitions and postulates (as fundamental constructions)." ("[…] wie die euklidischen Theoreme im Sinne Platons allein aus Definitionen und Postulaten (als Grundkonstruktionen) zu beweisen sind," Lorenzen 1984, 15, English translation MS)

[3] See Kant's *Transcendental Exposition of the Concept of Space* in his *Critique of Pure Reason*, B 41–42, Kant 1998, 69–70.

[4] Kant 1997.

reconstruction and analysis of the processes that have led to this knowledge over the course of time. Kant's program of exploring which aspects of our knowledge originate in preformed cognitive structures and which aspects involve empirical insights is thus transformed into that of studying the history of the interactive processes between experience and structures of knowledge. It is in this vein that the historical epistemology of space attempts to address the problem of the epistemic status of our spatial knowledge by studying its history.

The developmental view on cognition is strongly suggested by results from different empirical disciplines. First and foremost, evolutionary biology teaches us that cognition is a function of the human organism, in particular the brain, and is therefore to be understood as a product of biological evolution. From studies in developmental psychology it has furthermore become clear that many fundamental cognitive structures are not present at the moment of a child's birth, but are only gradually built up over the years in the long process of growing up. Finally, studies in the history of science and philosophy have revealed the historicity of fundamental concepts such as *space*, *time*, *force*, and *matter*, a historicity that became most obvious through the radical changes associated with the rise of the theories of relativity and the quantum in early twentieth-century physics.

Accordingly, we can distinguish three interwoven strands of development for which we can study the role of experience in the process of building up the perception and conception of space: 1) the phylogenetic strand, that is, the development of the biological species *Homo sapiens*; 2) the ontogenetic strand, that is, the development of individual human beings; and 3) the historiogenetic strand, that is, the development of human society and culture through history.

The phylogenesis of cognition is the subject matter of *evolutionary epistemology*. Continuity of development is produced by heredity. While experience pertains to individuals, against the background of genetic variation it shapes the development of the species stochastically through its impact on an individual's ability to contribute its genes to the next generation's gene pool (i.e., through selection). In this way, since the genes define a species' cognitive potential, the experience of one generation has a bearing on the next generation's basis for experience and thus for further cognitive evolution.[5]

The ontogenesis of cognition is the subject matter of *genetic epistemology*. Continuity of development is produced by the identity of an individual's psyche. Experience may become part of the individual's memory and may shape developing cognitive structures, which are mental reflections of real actions. The cognitive structures in turn constitute the basis for further action and related experience and, as a consequence, for further cognitive development.[6]

The historiogenesis of cognition is the subject matter of *historical epistemology*. Continuity of development is produced by external knowledge representations which serve the social reproduction of cognitive structures within a culture or their transfer between cultures. This reproduction relies on institutions structuring the use of the external representations.[7] Experiential knowledge is encoded in these external representations, which in turn become the precondition for further experience and the construction of new cognitive structures.

[5] See Lorenz 1977 for a classic work on evolutionary epistemology and Vollmer 1994 for a concise overview.

[6] See Piaget 1970 and other works by Jean Piaget cited in this chapter.

[7] 'Institutions' are understood here in the most general sense as social patterns that structure and control collective actions.

These may then become encoded in higher-order representations which are the basis for further experience and further cognitive development.[8]

The historiogenetic strand is the one that will concern us in the following pages.[9] It is closely interwoven with the other strands in two fundamental ways. First of all, in anthropogenesis, the transition from animal to human, phylogenetic and historiogenetic factors are closely intertwined. The emergence of human culture and with it the onset of the historical development of human cognition was a result of biological evolution and, as a consequence, necessarily built upon its biological foundations. But not only did human biology condition the onset of human culture, this culture also conditioned the last steps of anthropogenesis.[10]

The second way historiogenesis is related to the two other strands of cognitive development is based on the fact that the species' development, its phylogenesis as well as its historiogenesis, is realized through the ontogeneses of the individuals. Thus, the phylogenesis of cognitive structures depends on the ontogenetic transformation of the genotypes into phenotypes, and only the latter are subject to natural selection. In a similar way, the historiogenesis of cognitive structures depends on individuals who appropriate collective knowledge available in a given society at a given time in history in their ontogenesis and participate through their cognitive activities in the transmission and transformation of this knowledge.

The intertwining of the ontogenetic and historiogenetic developments of cognition explains the central role of means of external knowledge representation for understanding long-term developments in the history of knowledge. These means of representation – communicative action, spoken language, artifacts, drawings, maps, writing and other symbol systems – mediate between socially shared knowledge, which is the subject of historical development, and the individuals' knowledge which, while being subject to all the contingencies of the individual biographies, is the only actual realization of the shared knowledge. While the external means of knowledge representation define a space of possible transformations of shared knowledge, such transformations actually occur only through the thinking of individuals, which is in turn conditioned by their participation in this knowledge.

The recognition of this dialectic between individual thinking and shared knowledge is crucial for an understanding of the aim of a historical epistemology of space as outlined here. The intention is not to provide a narrative of the world history of individual acts of spatial thinking. Such an aim would not only be unachievable, owing to the sheer magnitude of the task, but also theoretically unsatisfactory, precisely because it neglects the social dimension of thinking. The aim is rather to describe historically identifiable and theoretically interpretable cognitive configurations, or stages, that demarcate the horizon of the forms of spatial thinking that are possible in a given historical situation.[11]

[8] Cf. Damerow 1996b, 371–381. Accounts on historical epistemology as the term is understood here include, among others, Renn 2004, Renn 2005, and Damerow 2007.

[9] Related studies are Damerow 2007 concerning the concept of number, and Elias 1988 and Dux 1992 concerning the concept of time. For histories of concepts of space in science and philosophy over the long term, see Gent 1971, Jammer 1954, and Gosztonyi 1976.

[10] See, for instance, Schurig 1976, in particular, 164–214 for a discussion of the coevolution of anatomy and culture in anthropogenesis. For a more recent account and further references to the literature, see Odling-Smee, Laland, and Feldman 2003, 239–281 who discuss coevolution from the perspective of niche construction.

[11] Cf. Damerow 1994, 312. A common impulse against the idea of historical development of cognition is arguably rooted in the the well-meant attempt to avoid value judgements. But we may speak of development whenever change produces circumstances that serve as a necessary precondition for specific further changes. To deny historical development of cognition would mean to deny the dependency of cognition on its earlier forms and thus,

The identification of stages does not imply that the historical development of the forms of spatial thinking is a linear process. Although governed by entirely different mechanisms, this development actually shares some qualitative features with biological evolution, such as the following:

- *Unpredictability of future developments*: Developmental processes are complex and interconnected, with the result that future developments are, as a rule, unpredictable at any time in history.
- *Dependency of later developments on earlier ones*: Despite this indeterminacy, earlier developments produce the necessary preconditions for later ones.
- *Temporal directedness of overall development*: This dependency of later developments on earlier ones explains aspects of the temporal order of development and makes it possible to define *earlier* and *later* stages of spatial thinking.
- *Asynchrony of development*: The temporal directedness does not imply, however, that all development proceeds uniformly on a global scale: different stages coexist and there may even be local or temporal developments from a 'later' stage to an 'earlier' one.

In the following sections of this chapter we shall discuss six different aspects of the historical development of spatial knowledge. The similar biological constitution of all humans and the fundamental similarities in their physical environments make it plausible to assume that there are structures of spatial cognition that do not vary between different cultures or over the course of history, but constitute the foundation for all cultural manifestations of spatial knowledge (section 1.2). While similar natural conditions hold for some animal species, humans possess unique social abilities to share knowledge, a fact that constitutes the basis for the cultural evolution of human spatial cognition, leading to elaborate cultural systems for orientation (section 1.3). The transformation of human societies from bands and tribes to city states and empires created new forms of the social control of space, involving techniques of surveying, writing, and drawing, which became the precondition for the development of geometry and thereby shaped the further development of spatial thinking (section 1.4). Philosophical and mathematical texts that emerged from cultures of disputation in Greek and Chinese antiquity document theoretical reflections on spatial concepts and their ensuing generalization. Processes of reflection and generalization continued in the subsequent philosophical and mathematical traditions (section 1.5). Processes of concept formation and the generalization of spatial concepts were also promoted by the expansion of experiential spaces, be it the geographical spaces known through political expansion, trade, and exploration, be it the cosmological spaces known through observation, be it meso- and microcosmic spaces known through the integration of technical and experimental knowledge into theories of space. A prominent example is the formation of the Newtonian concept of a homogeneous, isotropic, absolute space independent of its matter content, which can be understood as resulting from reflections on an integrated corpus of mechanical and astronomical knowledge (section 1.6). The expansion of experiential knowledge about the micro

ultimately, to deny its dependency on society and culture. But, as shall be argued below, this dependency is what distinguishes human cognition from animal intelligence. Its denial would mean to assume naively that any thought and insight was possible at any time in history. The outright identification of developmental approaches with value-judgements reveals an (often unconscious) ethnocentrism, since it uncritically presupposes that 'our modern' modes of thinking are more highly valued *per se*.

and the macro cosmos and the reorganization of the knowledge of classical physics at the beginning of the twentieth century led to the demise of the Newtonian concept of space as independent from matter. According to the most advanced theory of space in present-day physics, general relativity, space and matter are inseparably related to one another. At the same time, it is as yet unclear what a theory of space for the whole of physics would look like, since the two fundamental theories of present-day physics, quantum mechanics and general relativity, disagree on basic physical concepts, such as space, time, matter, and force (section 1.7). The chapter concludes with summarizing remarks (section 1.8).

## 1.2 Natural conditions of spatial cognition

In order to understand how human spatial thinking depends on the cultural conditions present at different times in history it is of fundamental importance first to identify spatial abilities and corresponding cognitive structures that are *not* products of human culture, and accordingly not subject to historical change. These may be termed the *natural conditions of spatial cognition*. Starting from such an identification we may then ask how historical and present-day cultural manifestations of spatial thinking relate to this universal basis.

The natural conditions of spatial cognition have a double origin. First, there are biological predispositions of the human species which also involve a cognitive dimension. Second, there are features of the physical environment in which each individual grows up that are so fundamental that they are independent of culture. In the first case, it is the mechanisms of biological evolution by which experience enters the formation of cognitive structures, in the second it is each individual's experience in ontogenesis. The two origins are closely entangled, since the ontogenetic unfolding of biological predispositions always takes place in a physical environment which exhibits certain universal features. While the question of the relation between the two origins will not concern us further here, it is important to note that the idea of universal aspects in human spatial cognition does not in itself imply any kind of nativism.[12]

When trying to identify the natural conditions of spatial cognition we encounter a methodological problem. Cross-cultural studies help to identify aspects of spatial thinking that are human universals, that is, aspects that do not depend on the particularities of any specific culture (for instance on the use of a particular language); yet the universal aspects identified in this manner will include aspects that depend on the very existence of human culture (for instance on the presence of language altogether). From birth (and in certain respects even before that), human beings are immersed in their culture. They are born into a cultural *habitus* that shapes their social and physical experiences and thus potentially exerts an influence on their cognitive development. More importantly, they participate in specifically human modes of cultural learning.[13] As a consequence, when studying the ontogenesis of human cognition, it is practically impossible to abstract from processes of the individual's enculturation. Therefore, to reveal its natural conditions, human spatial cognition has to be compared to animal cognition considered as the cognition of beings without human culture. Of particular interest in this context is the cognition of nonhuman primates, since cognitively

[12] For a critical discussion of 'nativist' approaches, see, e.g., Tomasello 1999, 48–51.

[13] For an explanation of cultural *habitus*, see Tomasello 1999, 78–81; for that of cultural learning, see Tomasello 1999, 61–70, who relates these human modes of learning to the conception of others as intentional beings and argues that its development begins around the ninth month.

they appear closest to humans and are probably similar to our not-yet-human ancestors. We could argue that natural conditions of human spatial cognition comprise their spatial abilities and the corresponding cognitive structures.[14]

To identify the natural conditions of spatial cognition the object of study must therefore be the spatial behavior of animals and humans (children and adults), and in particular of nonhuman primates. Of central relevance in this context are the abilities of *object permanence* and *cognitive mapping*. Let us briefly describe them. *Object permanence* is what developmental psychologists call the mental construction of objects as entities independent of the self, which are understood to exist in a definite location or move along a definite trajectory in space. Studies in developmental psychology suggest that what may be called the *schema* of the permanent object is not present at the time of a child's birth, but only develops during the first two years of childhood.[15] Object permanence skills have been proven for many animal species.[16] There is thus clear evidence that the schemata of object permanence are not unique to humans. On this basis one may argue that they belong to the natural conditions of human spatial cognition.

Besides the smaller-scale skills related to object permanence, humans develop sophisticated abilities of spatial orientation on larger scales. They can quickly accumulate spatial information about previously unknown territories; in known territories they can move flexibly, that is, they can make detours and take short cuts that they have not previously made or taken; and they can optimize their routes by arranging the stations of their travel in a rational manner. They can integrate knowledge about landmarks with knowledge about the motion of their own body to construct route knowledge, and combine their knowledge about intersecting routes to obtain what may be called configurational knowledge: knowledge about the overall configuration of landmarks and their relations.[17] They are also able to make use of cues such as wind directions, the position of the Sun, or distal landmarks. Following a large body of literature, we refer to these abilities here as *cognitive mapping*.[18]

Besides humans, various species of animals exhibit sophisticated performance in spatial orientation.[19] Nonhuman primates in particular have been shown to be able to use spatial information in a flexible manner.[20] Chimpanzees, for instance, who were shown how food was hidden at several locations in a familiar environment were later able to retrieve most of the food, whereby they did not follow the order in which the food was placed, but an order that reflected a minimum-effort strategy. Using this type of strategy they could also be shown first to retrieve the kinds of food they prefer before proceeding to less favored food.[21] Hamadryas baboons, to give another example, were being able to remember the locations of important sites such as sources of water in their local environment, using shortest distance

[14]For a more critical discussion of comparisons between animal and human spatial cognition, see Hazen 1983.

[15]Piaget 1959, 97–101. For a definition of the concept of schema, see, for instance, Piaget 1983, 180–185. A different definition is given in Neisser 1976, 51–57. Below we will introduce the concept of *mental model* to describe relevant cognitive structures.

[16]For a survey of the spatial abilities of nonhuman primates, see Tomasello and Call 1997.

[17]Siegel and White 1975; Kitchin and Blades 2002, 89–90.

[18]See Kitchin and Blades 2002 for a recent account on cognitive maps which surveys a large part of this literature.

[19]See various contributions in Pick and Acredolo 1983.

[20]See Tomasello and Call 1997, 28–39 for a survey of the evidence for different primate species.

[21]See Menzel 1973; Menzel 1987 discusses the interpretation of these findings in terms of cognitive mapping.

strategies as they moved around, and even speeding up when approaching a known site well before they could have perceived it, demonstrating that they knew where they were.[22]

We can summarize these findings as showing that the basic human cognitive mapping skills – similarly to object permanence skills – are not indicative of a peculiarity of human cognition but are part of its natural conditions:[23]

> Overall, primates have the general mammalian spatial skills of cognitive mapping and object permanence [...]. [...] It is also unlikely that humans have any special skills in these domains of spatial cognition. They too possess the general mammalian skills of cognitive mapping and object permanence [...].

Action and perception under control of the schemata of object permanence and the skills of cognitive mapping imply fundamental spatial structures which include the following:

- *Dichotomy of objects and spaces*: Objects are tangible (albeit not always accessible), and between them there are non-tangible (i.e., 'empty') spaces.
- *Definiteness and exclusivity of place*: Every object is in a place and always in one place at one particular time. No other object can be in the same place at the same time.
- *Three-dimensionality of objects and spaces*: Objects are extended in such a way that different sides of an object are perceptible from different perspectives. Each object has a concealed reverse side. The spaces between objects are likewise extended, allowing for objects not only to be located side by side, but also to obstruct the view to another object.
- *Distinction of vertical direction*: There is one direction determined by the tendency of most objects (including one's own body) to fall down or to resist lifting.
- *Continuity of object trajectories*: The mutual spatial relations of objects, including one's own body, may change, which means there is motion. The trajectories of motions are continuous, that is, there are no 'jumps': objects do not vanish in one place and reappear in another, but pass through all the intermediate places during the motion. In their fully developed form, the schemata of object permanence imply continuous trajectories regardless of whether they are perceived or not.[24]
- *Dichotomy of movable and unmovable objects*: Some objects can be moved or move by themselves (e.g., conspecifics); other objects cannot be moved, that is, they have a fixed location (e.g., trees). These latter objects thus define a ground against which one's own motion and the motion of other objects is perceived.
- *Focus on plane of movement*: The space of movement (structured by a network of landmarks, places, and regions) mostly lies within a more or less horizontal plane. (The additional importance of the vertical depends on the mode of life in particular ecologies such as living on different levels of a forest, a mountainous region, or a city with multi-story buildings.)

---

[22] Sigg and Stolba 1981.

[23] Tomasello and Call 1997, 55–56. There are further studies pointing to similarities in animal and human spatial cognition. Thus, Foreman, Arber, and Savage 1984, who carried out experiments with pre-school children in a so-called radial maze, an arrangement previously used in experiments on spatial abilities of animals, have pointed to remarkable similarities between pre-school children and well trained nonhumans in the performance of certain spatial tasks. This was interpreted to suggest a similarity of the role of visuospatial cues in the development and use of cognitive representations of space and the underlying processes across species.

[24] Following Piaget, this is often referred to in the literature as 'stage six abilities'.

- *Path-connectedness of plane of movement*: The topology of the plane of movement is path-connected, that is, between any two locations there is a path connecting them (otherwise it would not be a plane of movement). Generally, there may be different paths for reaching the same location and one may travel along a closed path and come back to one's initial location, even in cases where the path encircles insurmountable obstacles (e.g., trees, mountains, river sections, or buildings).
- *Dependency of effort on path taken*: The effort it takes to get from one location to another generally depends on the path taken.

What is the epistemic status of the natural conditions of spatial cognition and the described cognitive structures? As we have seen, these conditions are rooted in sensorimotor intelligence, which is characterized by a close relation between cognition and concrete action.[25] The development of sensorimotor activity, roughly spanning the first two years of human life, ranges from reflexes via habits to the emergence of practical intelligence. In the course of this development, sensory data are assimilated to cognitive structures called *schemata of action*, which are in turn accommodated to the increasing amount of sensorimotor experience. The result is an increasing coordination, generalization, and differentiation of schemata of action which constitute human sensorimotor intelligence.[26]

It is important to note that the implied spatial structures described above are not in themselves an object of thinking. They allow for successful action, but there is no indication that the related spatial abilities imply any consciousness, that is, any reflection upon the schemata controlling the actions, and thereby go beyond the sensorimotor realm.[27] Thus, without the *dichotomy of objects and spaces*, no object could be perceived or grasped. Without the *dichotomy of movable and unmovable objects* no stable mental representation of the environment would have been possible. Without the *three-dimensionality of objects and spaces* no change of the visual image could be understood as a change of perspective. But while these structures allow for spatial inferences to be drawn, they do so only in the context of action and perception and are otherwise inaccessible to the actor.[28] This becomes clear, for example, when school children who successfully find their way from home to school and back are unable to represent these routes in a map-like fashion.[29] Another example is provided by the well-attested difficulties that children have in rotating a landscape in their minds and describing how it would look from a different point of view.[30]

---

[25] See Piaget 1981, 107–116; Piaget 1959, 86–96; Piaget and Inhelder 1956, 5–13.
[26] See, e.g., Piaget 1981. See also Damerow 1998, 248.
[27] They rely on what Piaget has called *perceptional space* in distinction to *representational space*, which is built up only at the preoperational and operational stages (Piaget and Inhelder 1956, 3–43). See, however, C. Boesch and H. Boesch (1984, 168–169) who interpret certain of their findings as evidence for concrete operational thinking in the spatial reasoning of nonhuman primates and suggest the existence of 'Euclidean' cognitive maps, relating to Piaget's distinction between topological, projective, and Euclidean space; see also Normand and C. Boesch 2009.
[28] It remains an open question to what extent the transfer of spatial abilities to novel and artificial contexts of action presupposes an understanding of the novel situation as involving a representation, e.g., when rhesus macaques using a joystick show that they are able to anticipate the path through a computer-simulated maze; see Tomasello and Call 1997, 51–54.
[29] Piaget, Inhelder, and Szeminska 1960, 3–26.
[30] See the classical experiment by Piaget and Inhelder (1956, 209–246). For a critical discussion integrating recent empirical results, see Newcombe and Huttenlocher 2003, 118–125.

In particular, there is no indication of symbol use or the dependence of spatial cognition on external knowledge representations in general.[31] Accordingly there are also no concepts of space. The cognitive structures forming the natural conditions of spatial cognition common to all humans do not represent general, or abstract, ideas but depend on the specific contexts of action and perception. They are not to be found on the level of concepts but on that of the schemata controlling sensorimotor behavior.[32]

Besides the notion of schema of action we shall employ the concept of *mental model* in referring to these cognitive structures. By this term we understand internal knowledge representation structures which allow current experience to be processed by relating it to former experience. The former experience is coded in the mental model in two distinct but related ways. First, the structure of the model, which consists of *slots* and their mutual relations, is a result of earlier accommodations to experience. The slots are filled by specific instances, that is, by an input from the current situation fulfilling certain conditions required by the slot. But these slots may also have default fillings which are effective whenever appropriate current information is not available. These default fillings result from earlier experience, thus constituting the second way in which experience is coded in the model. In this way, a mental model allows the perception of, understanding of, or even reasoning about a situation whenever the situation can be successfully assimilated to the model – even in cases where the available information is incomplete. A major reason to introduce the concept of mental model here, and not simply to speak of sensorimotor schemata, is that mental models function on different levels of cognition. The sensorimotor and practical mental models inform the models functioning on higher conceptual and theoretical levels (and these may in turn have repercussions on the lower levels).[33]

The sensorimotor mental model of permanent objects is a mental structure to which sensory data are assimilated when objects are perceived and handled. For the assimilation to be successful, the shape, size, location, and position of the object must be identifiable. They do not need to be constant in time, however, although the sensorimotor schemata that underlie the model ensure that certain changes in perception are interpreted as changes of perspective, that is, of the position of the object or one's own body in respect to it, rather than as changes of the object itself. As becomes clear from our discussion above, the sensorimotor model in its fully developed form further implies the mental representation of continuous trajectories.

To describe a range of abilities in large-scale spatial orientation, we have employed the term *cognitive mapping*. This term is widely used, but the precise character of the mental representation underlying the related abilities is a matter of controversy. In particular, it is not at all clear that this representation can be characterized as a bird's eye view of the environment as the term 'map' suggests. Just as the mental model of object does not presuppose a

[31] A possible counterexample of symbol use in spatial communication among bonobos is discussed in Savage-Rumbaugh 1998, 161–165, but does not seem conclusive.

[32] We reserve the notion of concept to describe elements of knowledge structures that are somehow related to linguistic or otherwise symbolic representations, without implying, of course, that there is a one-to-one relation between concepts and words.

[33] On the concept of mental model as understood here, see in particular Renn and Damerow 2007; see also various contributions in Gentner and Stevens 1983. The concept is akin to Marvin Minsky's *frames* (Minsky 1975).

three-dimensional mental image,[34] the mental representation of the large-scale environment need not take the form of a two-dimensional map.[35]

Here the corresponding cognitive structures shall again be described in terms of mental models. The *mental models of large-scale space* may be conceived of as networks of landmarks and their spatial interrelations. It is plausible to assume that the landmarks and their relations are part of a hierarchical structure in which places and regions of different size are defined by reference to landmarks or other places and regions.[36] The landmarks, places, and regions are further endowed with contextual information about what is found there, e.g., kinds of food, water, predators and conspecifics, tools, and places to rest. The spatial relations between landmarks, places, and regions of different size involve topological information (inclusion, order along a route, proximity) as well as information on distances and angles. This latter information is given not in terms of numerical measures, of course, but rather in terms of sensorimotor experiences concerning variations in ease of travel, directions to landmarks, and perspectives. Configurations of landmarks, places, and regions can further be related to reference points outside the realm of motion such as the Sun or distal landmarks like a big mountain, or to overall directions defined, e.g., by a slope of the landscape or by recurring winds. The landmarks that fill the model's slots are permanent objects or configurations of such objects, so that the elementary knowledge about objects in general (their permanence, their change of appearance with perspective and distance, etc.) applies to them. The structural relations between the slots contain the knowledge about the spatial relations among the landmarks. While the individual realizations of the mental models of large-scale space are highly dependent on the concrete features of the respective environment, since they encode the experiential knowledge accumulated as the individual moves through this environment, the basic structure applies universally. This universal structure will be referred to in the following as the *landmark model of space*.

## 1.3 Culturally shared mental models of space

If the natural conditions of human spatial cognition are similar to those of some animal species, as has been argued in the previous section, what accounts for the obvious distinction of human spatial abilities and thinking? Rather than attributing this distinction to some specifically human biological disposition for *spatial* cognition, the point shall be made here that the distinction can be explained as resulting from uniquely human abilities of *social* cognition. One argument against the existence of a specifically human *module* for spatial cognition is based on considerations of the necessary timescales for processes in biological

[34]It is the functioning of the model – for instance, the way different perspectives are coordinated to make an object remain constant in size and shape under different views – that implies the three dimensionality. For a suggestion of how a three-dimensional cube and its transformations under different perspectives may be realized mentally without invoking a three-dimensional mental image, see Minsky 1975, 216–221, who uses coordinated *frames*. A more comprehensive discussion of three-dimensional vision is found in Marr 1982.

[35]Objections against the imputed use of cognitive maps, in particular when simpler explanations of the spatial abilities are available, are raised, for instance, by Tuan 1975 and Bennett 1996. Recently, Wang and Spelke 2002 argued against the concept of cognitive map, emphasizing the human use of navigation techniques such as path integration, which are also found in insects and spiders and imply no more than the mental representation of one vector. It seems, however, that the presence of more 'momentary' and 'egocentric' representations in no way precludes the build-up of more enduring and comprehensive mental representations. On the relation of these two types of representations, see, for instance, Cornell and Heth 2004.

[36]See Gärling, Böök, and Lindberg 1985 for a detailed description of possible entities cognitive maps are made of.

evolution.[37] We could adduce another argument if it could be shown that the specifity of human social cognition, together with the historical development of human thought ensuing from it, can satisfactorily explain the characteristics of human spatial cognition such that no further biological factors have to be invoked. Exploring the extent to which this claim can be substantiated is a major task for a historical epistemology of space.

The human ability of social cognition implies that humans are able to communicate, to share knowledge, and to learn from each other. For this kind of cognition to arise it is crucial that humans understand their conspecifics as intentional beings, that is, as beings who act purposefully just like themselves, and are able to imagine themselves in another's place.[38] In order to communicate about space, human children must learn to adopt the perspective of others. To do this they have to construct a mental representation of space that allows conception of all possible perspectives. This means the construction of what Piaget calls *representational space* as distinguished from *perceptional space*.[39] It is the social aspect of human cognition that implies representations that go beyond those closely tied to action and perception occurring at the latest stages of sensorimotor development.[40]

Sharing knowledge crucially depends on what Piaget calls the *symbolic function*, that is, the ability to distinguish events and objects from their meaning. In human ontogeny this ability emerges at the preoperative stage, which succeeds the sensorimotor stage. On the basis of this ability, actions of conspecifics can be understood to mean something, that is, they become potential means of knowledge representation. Purposeful actions with the aim of communicating knowledge, like gestures, and directed joint action become possible. Tools likewise come to represent knowledge in relation to the actions performed with them. Another particularly powerful means of knowledge representation and communication is human language, which phylogenetically is assumed to have developed in the course of the Paleolithic period.[41] Visual representations like drawings are also known from Paleolithic times. They are attested by various kinds of extant artifacts, most prominently the cave paintings of the Upper Paleolithic. In the course of continued cultural evolution, the means of external knowledge representation develop further themselves, for instance, under the particular socio-cultural circumstances of early city-states, when writing and the use of other sign and symbol systems such as numerical notation began to emerge (see section 1.4).

Consequently, the crucial distinction between animal and human cognition is the emergence of a cumulatively evolving human culture, a thoroughly social phenomenon. For every ability of individual humans that may be argued to play a crucial role in the emergence of this culture, such as the ability to use and produce tools, or to understand conspecifics as intentional beings, or to understand symbols and develop language, we find precursors in the animal kingdom.[42] Rather than being attributable to a single distinguishing factor,

---

[37] See Tomasello 1999, 54–55.

[38] On the specifically human ways of learning following from their ability to understand their conspecifics as intentional beings, see Tomasello, Kruger, and Ratner 1993 and Tomasello 1999, 26–55.

[39] Piaget and Inhelder 1956, 3–43.

[40] Piaget 1959, 364–376.

[41] Referring to results from neurology, developmental psychology, and archaeology, it has been speculated that the development of human language was closely related to the communication of cognitive maps (Wallace 1989).

[42] Besides Tomasello and Call 1997, see, for instance, the discussion of cognitive abilities such as categorization as developing independent of language in Langer 2001 and reports on tool-making and tool-using abilities and linguistic capacities of bonobo individuals (Schick et al. 1999, Savage-Rumbaugh and Fields 2000).

the animal-human divide seems to emerge from a process in which social, material, and cognitive developments interact in a complex causal structure.[43]

An immediate consequence of the cultural evolution of human societies for spatial cognition is that the mental models of large-scale space become culturally shared. In addition to those commonalities between two individuals' mental models of space that are due to their similar biological constitutions and their similar experiences within the same environment, human mental models of space display cultural commonalities. In this way the mental models of space themselves become part of an evolving culture, accumulating collective experience over generations and becoming richer and more refined than any mental model a single individual could have produced.

The sharing of mental models of space appears to be common to all human societies, from nomadic tribes to modern urban societies. When considering the impact of the cultural sharing of knowledge on the mental models of large-scale space, the general objects of study are therefore the practices of navigation and spatial orientation and their externalizations in language and other artifacts in all kinds of human societies. In most contemporary societies, however, these practices involve specialized means of spatial representation and advanced technology which have developed over the long course of history. To get an idea of what can be achieved in the absence of maps, compasses, sextants, or GPS receivers, we have to study the spatial practices of nonliterate societies that do not employ such specialized material tools. In the case of prehistoric societies, the archaeological evidence is the only available source for a reconstruction of such practices. In the case of recent nonliterate societies, by contrast, spatial practices, including their spatial language, can be investigated much more directly, which makes them an advantageous object of study.

Recent nonliterate societies show a wide variety of cultural systems for spatial orientation and communication.[44] This cultural diversity is due not only to the self-referential dynamics of cultural evolution, but obviously also to the fact that these systems represent responses to the challenges of widely differing ecologies to which they are adapted. Nevertheless, there are common patterns that may be discerned. It may be observed, for instance, that toponyms play a central role in spatial reference in a wide range of societies. Places and their relations are richly endowed with meanings relating to mythology, the history of places, and the natural knowledge about them. In many societies this practice is additionally complemented by a system of absolute directions, which in some cases plays such a crucial role that members learn always to keep track of these directions.[45]

Two examples of recent nonliterate societies and their spatial language and practices are discussed in this book, the Eipo living in the central highlands of West New Guinea, and the Dene Chipewyan living in the Cold Lake region in Alberta, Canada (Chapter 2). The spatial knowledge described in this context may be characterized as practical knowledge. Among its characteristic features are: its *transmission through external knowledge representations*; its *cultural organization*; its *dependence on the specific contexts of action*; and its *locality*.

*Transmission through external knowledge representations.* In contrast to sensorimotor knowledge, which is built up in the individual's interaction with the physical world, practical knowledge is built up through social interaction and communication. The knowledge representations employed in this context include joint activity and action with the explicit

[43] See, for instance, Damerow 2000 and Jeffares 2010.
[44] See, e.g., Burenhult 2008; Senft 1997; Levinson and Wilkins 2006.
[45] Various examples are given in Levinson and Wilkins 2006.

aim of teaching, the tools and artifacts produced through such actions, and spoken language. While the communication builds upon shared sensorimotor structures, the use of external knowledge representations makes it possible to accumulate knowledge that could never be acquired solely through one individual's experience.

*Cultural organization.* This accumulation is accompanied by the cultural organization of knowledge – which, in fact, makes the mastery of the accumulated knowledge possible in the first place. Culturally shared large-scale space is spanned not only by landmarks, places, regions, and their relations, but by the meanings attached to these entities. These meanings organize the spatial knowledge and are given in form of nomenclatures, narratives (mythical or otherwise), or sets of practices. Place and spatial order play an important role in Eipo myths, for instance, and, conversely, mythical narratives are instrumental in handing down spatial knowledge.[46] In contrast to the sensorimotor mental models of space, large parts of this mental representation may be accessed deliberately by its holder, particularly in order to communicate about space. Besides the cognitive dimension, the cultural organization of knowledge further implies an institutional dimension: The social reproduction of knowledge relies on more or less stable social patterns (institutions) structuring the collective use of the means of knowledge representation.

*Dependence on the specific contexts of action.* The spatial concepts structuring practical knowledge are, as a rule, not abstract or general but depend on the specific contexts of action. They are not applications of more general concepts in concrete situations but are rather conditioned by these situations. Further, the way in which the concepts structuring practical thinking about large-scale space relate to more small-scale spaces remains largely undefined. As a consequence, metrization remains fragmentary. Distances measured in terms of days of travel are not brought into any relation with cubits or other measurements of length which may be employed on a different scale.

*Locality.* The shared mental models of large-scale space are local in character. Practical concepts of space depend on the particular features that make up the space, and are not generally applicable to arbitrary environments. Systems of toponyms, for instance, obviously apply only locally, since they inherit the dependence on the particular environment from the landmarks and relations they refer to. The same holds for most variable cues such as winds or swell-patterns. But more structural elements of the system of orientation may be dependent on local peculiarities as well. Thus, the widespread use of star positions for determining directions by Micronesian expert navigators only works due to the proximity of their islands to the equator, since it is only there that the stars and constellations rise and set nearly perpendicularly to the horizon.[47]

To sum up, culturally shared mental models of large-scale space may be understood as collective elaborations and modifications of sensorimotor models. Just like the latter they are based on the landmark model of space, from which they inherit many structural features. At the same time, they encode a larger body of experiential knowledge than the sensorimotor models: they integrate the experiential knowledge about the environment not

[46] Heeschen 1990. This appears to be a widespread means of organizing spatial knowledge; another example is the practices of the Ngatatjara who live in the Australian desert and use myths and ritualistic sequences of events to memorize and communicate the cultural knowledge about their habitat. A brief description is given in Heth and Cornell 1985, 232–235.

[47] For the use of the *star compass* of the navigators of the Caroline Islands, see Gladwin 1974. See also Schemmel 2016 and the references provided there.

only of one individual but also of whole societies over the course of many generations. This integration is achieved by means of the cultural organization of knowledge, which necessarily reflects features of the local environment and displays cultural characteristics. Elementary knowledge structures thus serve as a foundation for culturally shared practices without determining their cognitive dimension. At the same time, culturally transmitted knowledge has repercussions on the more elementary level of sensorimotor knowledge when action and perception based on a culturally shared mental model of space becomes intuitive.

## 1.4 Social control of space and metrization

An immediate consequence of the cultural evolution of human societies on spatial cognition, which was discussed in the previous section, is the development of elaborate practices of spatial orientation based on shared mental models of large-scale space. Another way in which the cultural evolution of human societies shapes spatial thinking is based on the fact that the organization of society implies the social control of space. How is space divided among different individuals and social groups, what is the social function of different places, what are the places for public, sacred, or private affairs, who is allowed to go where, and who is allowed to use what land or even owns it? Questions of this kind can be observed to arise in the context of the organization of any human society.

The means for the social control of space depend on the respective form of social organization. In the case of small rural communities such as that of the Eipo described in Chapter 2, we may speak of the *mythical control of space*. Under the mythical control of space, knowledge about the social function of different places and about the allocation of space is largely represented by myths, which also ensure its social implementation. Despite the central role the division of land plays in social life, the mythical control of space does not provide standardized tools for measuring lengths and distances or for determining the quantitive measurement of an area. The Eipo's construction of a sacred men's house of defined size and shape, for example, is a complex task which is mastered without recourse to material representations of spatial knowledge such as measuring rods, drawings, or any kind of specialized geometric language. Instead, the spatial knowledge necessary to build the house is embodied in the ritual actions specific to the Eipo culture.[48] The distribution of garden lands among the Eipo is governed by clan-membership, heredity, and the capacity to cultivate the land. There are practices for delimiting fields (the demarcation of land by sacred Cordyline trees), but not for determining or estimating field sizes. Conflicts over the right to use a piece of land may lead to hostilities or be solved by negotiation, but their resolution never involves measurement.[49]

Historically, the earliest evidence for the systematic use of standardized measures for the social control of space stems from the so-called early civilizations. The growth in population of neolithic sedentary communities in some areas of the world went along with the development of increasingly specialized food production, irrigation, and food storage technologies, and resulted in the emergence of stratified societies that controlled progressively

[48] Koch and Schiefenhövel 2009 and Koch 1984, 49–54. See also Chapter 2.

[49] Wulf Schiefenhövel, personal communication. See also Michel 1983. Other instances of the mythical control of space may be identified in the spatial practices and spatial thinking reported for the Bororo of the Brasilian central plateau – see the account of the socio-spatial structure of the village Kejara given by Lévi-Strauss (1955, 244–277) – and the Temne in northern Sierra Leone (Little John 1963).

larger spaces. The formation of city states and larger empires brought about new phenomena in human culture such as centralized administration, property regimes, monumental architecture, centralized religion, and new forms of standardized means of knowledge representation. In particular, it gave rise to new forms of the social control of space which may be referred to as *the administrative control of space*. These forms involved techniques of measuring, surveying, writing, and drawing, which implied a progressive metrization of space and led to a kind of proto-geometry.

A decisive strand in this bundle of developments was the emergence of new forms of the division of labor. Besides gender-specific forms of division of labor (with hunting considered as a predominantly male activity, for instance) or practice-specific forms (as in the case of the experts of Micronesian navigation), a fundamental social division became manifest: the division of physical and intellectual forms of labor. In general we can discern a physical and an intellectual component in the human practices of using and producing tools. Concrete action is preceded by planning, that is, selecting tools, determining the sequence in which they are used, and coordinating work in cases where more than one individual is involved. The growing complexity of the planning and organizational tasks in the stratified societies of the early civilizations led to a division of labor along this intellectual-physical divide. The result was a specialization of intellectual labor which became manifest in the emergence of professions such as the scribe, the administrator, and the surveyor, and in an administrative hierarchy reflecting the emergence of mental activities that coordinated other mental activities.[50]

As these mental activities are themselves dependent on material tools, the development of early civilizations went along with fundamental innovations in the means of external knowledge representation. This holds in particular for activities related to the social control of space such as architecture, urban planning, surveying, and field measurement which involved means of semantic and numerical notation as well as tools for graphical representation such as the compass and the ruler. Among the early civilizations in which such techniques developed are those of Mesopotamia, Egypt, China, and Mesoamerica and South America. The developments are well documented in the case of Mesopotamia, where proto-writing emerged before 3000 BCE on the durable medium of cuneiform tablets so that a large amount of administrative records are preserved. Evidence for similar developments in other early civilizations is more indirect. In the Egyptian case we have depictions of surveyors at work, e.g., the wall painting in the tomb of Menna in Thebes,[51] and mathematical texts on the calculation of areas such as parts of the Rhind Papyrus, but no administrative documents on the determination of field areas have been preserved. Evidence in the Chinese case comes from much later periods and again does not document early administrative practices.[52]

In this book the emergence and early development of the administrative control of space, and the related gradual metrization of space, is discussed using the example of Mesopotamia (Chapter 3). The history, ranging from early Mesopotamian practices of surveying to Babylonian geometry, spans millennia in which fundamental developments

---

[50] Damerow and Lefèvre 1996, 396–397.

[51] See, e.g., Lyons 1927.

[52] Consider, in particular, the *Jiu zhang suan shu* (*Nine Chapters on Arithmetical Techniques*), containing, among other things, problems on the calculation of field areas (Guo 1993, 79–213; for editions in European languages, see Vogel 1968; Kangshen, Crossley, and Lun 1999; and Chemla and Guo 2004).

occurred, such as the invention of the sexagesimal place value number system.[53] Accordingly, the spatial knowledge discussed in Chapter 3 ranges from practical knowledge to mathematical knowledge. It is the expert knowledge of a particular group of administrators and develops over history along with the means of symbolic representation. It is externally represented by measurement devices, drawings, and symbolic notation, which develops into writing on one hand and numerical notation on the other.[54] It thereby reproduces structures found on a more elementary level of cognition, this time, however, endowing spatial entities with arithmetic properties. This arithmetization of spatial entities also leads to an integration of spatial structures which remain separated on a more elementary level. Let us give two examples.

The *conservation of the size and shape of an object* independent of its location and position is implied by the sensorimotor schemata responsible for the coordination of perspectives. It is further implicit in the comparison of the size of objects by means of juxtaposition when no standardized means of measurement are available. The assumption of the conservation of the size of an object when it is moved through space is, in fact, a precondition for the use of measuring rods or ropes. In the context of the use of such tools and in the presence of standard measures of length, area, and volume, the conservation of size becomes manifest on the level of mathematical representation and implies metric homogeneity of space. This arithmetization also serves as a precondition for the integration of spaces of different orders of magnitude through the coordination of the units of measurements on different scales.

The *three-dimensionality of objects and spaces* is another example of the integration of spatial structures through arithmetization. Three-dimensionality is perceptually given on the sensorimotor level. Through the arithmetical dependencies between length, area, and volume it is reproduced on the level of the symbolic means of knowledge representation and enables the reflection on the relations between entities of different dimensionality.

The metric structure of space becomes more generalized through the application of the sexagesimal place value number system with its general procedures for addition, subtraction, multiplication, and division, and in combination with an abstract system of units defined by its internal relations. This illustrates how, in certain historical situations, the emergence of new means of knowledge representation in specialized practical contexts (surveying) may lead to a dynamic of knowledge development that engenders knowledge structures no longer directly related to that context (Babylonian geometry). But this greater generality implicit in the symbolic means of knowledge representation must not necessarily be made explicit, for instance, in the form of a term that represents the concept of a three-dimensional metric space spanning various scales.

Despite its novel degree of abstraction and its thorough metrization of area, the Babylonian surveyors' mental model of space actually differed from Euclidean space. The procedures of Babylonian geometry are of a limited generality which testifies to their origin in administrative practices. In particular, there is the striking absence of the consideration of angles as objects of mensuration, which is rooted in the implicit definition of area by means of what is known as the *surveyors' formula*, that is, the rules of calculation for determining the area of irregular quadrangles of sides $a$, $b$, $c$, and $d$, which corresponds to an application

[53] See also Damerow 2001; Høyrup 2002; and Robson 2008.
[54] See Damerow 2012.

of the formula $(a + c)/2 \cdot (b + d)/2$. Field areas are thus calculated on the basis of lengths without quantitatively accounting for the angles.[55]

## 1.5 Context-independence of mental models resulting from reflection

The cultural developments of spatial thinking discussed in the preceding section show a basic trend towards cognitive structures that are less dependent on the specific practical contexts from which they originated. An example is the emerging practice of area determination by means of a multiplication of lengths within the sexagesimal place value system, which implies a greater degree of generality than any conventional way of relating areas to standard lengths based on specific practices of measurement and notation. The increase in generality is obviously related to the development of the means of knowledge representation such as comprehensive systems of units and a place-value number system. But this development is only the material side of a dialectical process whose other side is mental. Performing operations on external knowledge representations builds up structures which are mental reflections of these operations. Since these operations disregard many aspects of the real-world objects, this mental process may be referred to as a reflective *abstraction*. When the new mental structures are in turn externally represented, e.g., by symbols forming a system, we may speak of a representation of higher order than the one from which the process of reflection started.[56]

Processes of reflective abstraction are a consequence of the exploration of existing means of knowledge representation. Exploration of these means by individuals may happen spontaneously at any time in history. But such individual developments remain without consequences in the history of knowledge unless there are social entities such as organized groups or institutions that ensure that the cognitive products are handed down and – at least for a certain period – become subject to cumulative development. A potential case of this type of institutionalization are the schools of the scribes in Mesopotamia which developed Babylonian geometry as a doctrine of areas independent of the context of surveying – even though the structure of Babylonian geometry still bears witness to its origin in practical surveying (as argued in the previous section; see Chapter 3). The context of teaching and learning the handling of symbolic means of knowledge representation seems to be a natural place for the emergence of exploratory forms of knowledge. Another such context is disputation, traditions of controversial discourse and rational debate. While such traditions are usually oral in origin, they may find expression in text traditions, possibly accompanied by an ongoing oral component. Disputation is a motor for reflection on concepts and, as a consequence, for their generalization. The resolution of apparent paradoxes, for instance, presupposes reflection on language and the delineation of meanings. Spatial knowledge need not be the primary object of these reflections, but if the aim is comprehensiveness it will naturally come into consideration.

One may distinguish two types of explorative knowledge, which may roughly be designated *mathematical* and *philosophical*. Mathematical explorative knowledge results from systematic reflection specifically on representations related to the use of instruments such as

[55] This method of determining areas was also used by the Roman *agrimensores* (Folkerts 1992, 324) and in demotic Egypt (Neugebauer 1934, 123). There is evidence that it may have also been used by Aztec surveyors (Williams and Carmen Jorge y Jorge 2008). On the origin of angle-geometry, see Gandz 1929.
[56] See Damerow 1996a.

measuring rods and ropes, the straight edge, and the compass.[57] Philosophical explorative knowledge, by contrast, results primarily from systematic reflection upon the linguistic representations of elementary shared knowledge.

Among the most prominent historical settings in which the exploration of the cognitive tools of spatial thinking became productive are the intellectual traditions of ancient Greece. The first-order knowledge that was reflected upon in this context was by no means of purely Greek origin. From the Archaic period on, astronomical, medical, and arithmetical knowledge from Egypt and Mesopotamia entered the Greek world.[58] In contrast to the Babylonian case, which was defined by the needs of central state administrations, the Greek situation was characterized by polycentrism, the encounter of different strata of society, and the negotiation and public justification of political decisions.[59] This was the background for pursuing systematic reflections which aimed at establishing a coherent, encompassing world view, distinct from the received mythology but with the same aspiration to totality. Written texts produced in the context of the Greek philosophers' activities now provide us with the earliest evidence of systematic reflections on the linguistic representation of shared spatial knowledge. A parallel and related development is the formation of a characteristic Greek tradition of mathematics, particularly concerned with questions of geometry.[60]

Among the later historical intellectual places which furthered deliberate and purposeful exploration of the implications of systems of knowledge representation were the Neoplatonic schools of late antiquity, Hellenistic science as pursued at the Museion of Alexandria, court science, philosophy and theology of the Arab Middle Ages as pursued in Bagdad and Córdoba, and the scholasticism of the Latin Middle Ages. In early modern times the theoretical reflection on fundamental concepts such as space and matter gained new impetus in the context of an ideological struggle between different strata of society. In their attempts to formulate encompassing counter world systems against the predominantly Aristotelian world view promoted by the Church, early modern natural philosophers faced the challenge of taking account of an increasing amount of empirical knowledge from practical mathematics and astronomy.[61] In the following centuries, theoretical reflection on space has become increasingly institutionalized in the disciplinary discourses of physics and philosophy.

All of the historical periods and places mentioned so far, in which the exploration of and reflection on representations of spatial knowledge took place, are more or less strongly related by ties of tradition: they all, in one way or another, relate back to the theoretical traditions of Greek antiquity. The example discussed in greater detail in this book, by contrast, presents a rare case of independent emergence of systematic reflections on spatial language; it is documented in the so-called *Mohist Canon*, a text from Warring States China, ca. 300 BCE (Chapter 4). The Mohist reflection clearly represents what we have referred to above as *philosophical* explorative knowledge, the systematic reflection on the linguistic representations of elementary knowledge, although references to mathematical instruments are also found in the text. Compared to the Greek case, the *Mohist Canon* represents a unique

[57] On the role of language as a means of knowledge representation in the emergence of theoretical mathematics, see Lefèvre 1981.
[58] See Schiefsky 2012 for a concise discussion and references to the literature.
[59] Lefèvre 1981; Lefèvre 1984, 306; Hyman and Renn 2012, 86–87.
[60] On the institutional background of the emergence of Greek mathematics, see Høyrup (1994, 9–15), who explicitly contrasts the Greek with the Babylonian case and argues for a close connection between the emergence of Greek mathematics and the contemporary philosophical discourse. See also Asper 2009.
[61] On this point, see Chapter 6.

source for addressing comparative questions in the long-term history of spatial knowledge; questions concerning the conditions for the emergence of traditions of systematic reflection and the necessities and contingencies in their development.

The spatial knowledge considered here can be described as *theoretical knowledge*. This kind of knowledge is largely conditioned by its means, that is, by the external knowledge representations from the exploration of which it emerges. It is handed down in text traditions, mostly in form of written language and symbolic notation, which make it possible to pick up a tradition again even centuries after it has last been actively pursued (although the case of the *Mohist Canon* shows that it may also be handed down without ever having been actively taken up again). It is aimed at consistency and comprehensiveness and thereby gives rise to more general and abstract concepts such as those of Euclidean distance and the atomistic absolute void, sometimes including a general concept of space.

The explorative reflection upon elementary structures of spatial thinking creates theoretical structures which preserve many of the spatial properties implied by sensorimotor intelligence. At the same time, the theoretical context of generalization and aspiration toward consistency leads to questions about these properties which could never have occurred in elementary or practical contexts. At the level of fully developed sensorimotor activity, the mental models have their clear-cut realm of applicability. At the level of theoretical thinking, by contrast, there is an inherent uncertainty about what aspects of the mental models to build upon. This ambiguity derives from the absence of the concrete contexts of action that limit the meaning of the linguistic representations of knowledge in their everyday use. The operations on external representations in reflective thinking are dissociated from these original contexts and produce structures inherent in the system of representations. The result of such processes of reflective abstraction are not predetermined in general, because the space of possible structures spanned by the means of representation is much richer than any particular realization in it. The analysis of the Mohist passages and their comparison to Western sources in Chapter 4 shows, among other things, that the occurrence of elementary mental models in theoretical thinking on space is indeed a cross-cultural phenomenon. The connection of such reflections with encompassing worldviews, by contrast, is a peculiarity of the Greek case and depends on the timing of specific theoretical traditions such as the construction of cosmologies on the one hand and the reflection on the meaning of words on the other.

There is a striking difference between philosophical and mathematical explorative knowledge. While the former depends on individual decisions motivated within more encompassing knowledge systems and remained controversial throughout the history of philosophical thinking, the latter was, from early on, considered to present inevitable truths. The well-defined object of reflection of mathematical explorative knowledge, the first-order representations related to the use of instruments (figures drawn by means of a straight edge and compass in the case of Euclidean geometry), allowed for a consistent representation within a deductive structure. The reflection on first-order representations thus led to a generalization of spatial concepts which implied a de-contextualization: what had been a theory of constructed figures became interpreted as a theory of space, decoupled from what fills space.[62]

[62]For an outline of the long-term transformation of the object of geometry from figures to second-order properties of figures, and eventually to space, which was a precondition for the formulation of non-Euclidean geometries, see

The reflection on the higher-order representations of Euclidean geometry (deductively organized sets of statements) further generalized the spatial concepts when the possibility of non-Euclidean geometries was discovered. It thereby led to theoretical alternatives in the case of mathematical knowledge as well, theoretical alternatives which could not be evaluated on purely rational grounds. As a consequence, it led to a re-contextualization of geometry, because there was a new appreciation of the role of rigid bodies (and light rays) for establishing the geometry of physical space. The emergence of non-Euclidean geometries thus functioned as a historical reminder of the empirical origins of Euclidean geometry in instrumental action. Accordingly, and in spite of deviating epistemological claims, the question of the applicability of non-Euclidean geometries was revealed as an empirical question. In this context, first-order representations of spatial knowledge (measuring rods), became higher-order representations that relate abstract structures to physical space by connecting theoretical knowledge with other layers of knowledge.[63]

## 1.6 The expansion of experiential spaces over history

In the previous section we argued that reflection on the external representations of elementary and practical knowledge may lead to new and more general spatial concepts. In such cases of theoretical thinking, novelty arises from the structures inherent in the means of knowledge representation and tools for intellectual labor becoming explicit through being explored and through reflective abstraction. But the history of theoretical reflection does not unfold before a background of unchanging spatial experience. If we are concerned with the relation of experience and theoretical reflection in the historical development of spatial concepts, we have to take into account a complementary long-term trend: the expansion of experiential spaces. This expansion of experience not only implies an accumulation of spatial knowledge but also plays an important role in creating new spatial concepts and stabilizing them within more comprehensive knowledge systems.

Starting with the first steps of ontogenesis, experience plays an instrumental role in shaping human spatial cognition (section 1.2). Beyond the immediate experiential environment of the individual, different socially shared spaces can be experienced in different societies. This experiential basis of spatial knowledge expanded in the course of history, not monotonically and not universally, but within a long-term, global perspective. One may distinguish three realms of experiential space to which this expansion pertains. First of all it pertains quite literally to the *geographic spaces* known to human societies, which have grown through travel, trade, exploration, and military campaigns. Such activities led to the expansion of the space for movement of various societies or even of their organized space, as in the case of expanding empires which take political and economic control of more and more territory. These spaces have grown in many local historical contexts and in a long-term perspective, spanning the time from prehistoric nomadic and sedentary tribes to modern global societies that enable intercontinental travel and communication.

Another experiential space that has expanded over history is *cosmological space*. Cosmological space is the entire universe known, or assumed to exist, by a given society. Society transfers spatial concepts and knowledge acquired in terrestrial contexts to this space. It is,

De Risi 2015, 1–13. For a general discussion of first and higher order representations in the history of mathematics, see Damerow 1994.

[63] On this point, see Chapter 7.

in particular, also the space of mythical realms of experience. Cosmological space is experiential through the observation of the sky, especially systematic astronomical observation. This space has grown enormously, from observations of the Sun, the Moon, the planets, and the stars in early societies, to the modern observation of astronomical objects billions of light years away. It has also grown with respect to its wealth of physical contents. With the increasing refinement of celestial mechanics from antiquity to modern times, and with the rise of astrophysics in the course of the nineteenth and twentieth centuries – developments clearly related to the progress of observational instruments and techniques – the import of knowledge from terrestrial science into cosmology has vastly increased. With the observation of the flight of the galaxies, cosmological space itself has been turned into an object to which elements of physical description, such as the field equations of general relativity or the model of a black body, may be applied. Visible light has become just one of a wide range of sources for knowledge about the universe, and present-day astronomy is reaching the brink of the observable universe: looking far away means looking back in time, and with the most recent breakthrough in the detection of gravitational waves[64] there is the justified expectation that we will soon be able to 'look through' the early universe which is opaque with respect to electromagnetic radiation.

*Microcosmic space*, just like macrocosmic space, has been a target for projection of experiential knowledge from the mesocosmic realm, as the example of atomism discussed in the previous section illustrates. On the background of such theoretical world views, knowledge about physical objects acquired through practical experiences in dealing with technological artifacts or even through systematic experimentation has potential implications for spatial concepts. The expansion of experiential knowledge about the micro-world was not only due to new instruments of magnification – from the optical microscope to the particle accelerator – but also to the systematic exploration of chemical, electric, and magnetic phenomena. In particular the increase, in modern times, of empirical knowledge in the fields of mechanics and electrodynamics led to fundamental changes in the concept of space, the first being related to Newtonian absolute space, the second to the spacetime of special relativity.

When considering the impact of the expansion of experiential spaces on spatial thinking, the objects of study are processes of concept formation fostered by the increase of experiential knowledge in the three realms described above: geographical, cosmological, and microcosmic space. Two examples are discussed in the present book: the geographical and cosmological knowledge on which the insight into the spherical shape of the Earth and the idea of its central position in a spherical universe are based, and the different ways to argue for this idea that are found in Aristotle and Ptolemy (Chapter 5); and the transformation of natural philosophical considerations on the relation between matter and space through the growth of the corpus of empirical knowledge on mechanics and astronomy (Chapter 6).

The knowledge discussed in these chapters is once again theoretical knowledge. Unlike the knowledge discussed in the previous section, it is theoretical knowledge resulting from systematic extensions of its experiential base. The accumulation of experiential knowledge takes place within institutions specifically designed for the purpose of knowledge acquisition[65] and often occurs using instruments specifically designed for the purpose of knowledge acquisition such as astronomical instruments and laboratory equipment. The empirical

[64] Abbott et al. 2016.

[65] 'Knowledge acquisition' or 'knowledge production', depending on whether one wishes to stress the objective or the constructive aspect of knowledge growth.

knowledge is organized in integrative structures based on symbolic and formalistic tools such as numerical coordinates, analytic geometry, calculus, and differential equations. The way the symbolic tools are used is shaped by the experiential knowledge to be integrated. At the same time, the symbolic tools are related to concepts and have a repercussion on conceptual structures. It is via the interaction of experience, symbolic representation, and concepts that experiential knowledge shapes conceptual structures. In this process of reflection upon the institutionally accumulated empirical knowledge, the mental models, which were based on elementary and practical experience, are transformed. The accumulating knowledge and its symbolical-formal integration thereby produce and stabilize models and concepts that are highly counter-intuitive. Examples of such counter-intuitive knowledge structures are:

- *The Earth has a spherical shape* (cf. Chapter 5). The idea of a spherical Earth violates the *distinction of the vertical direction* in elementary spatial cognition.
- *Matter is nothing but empty space permeated by forces* (cf. Chapter 6). This idea (formulated by Kant in his *Metaphysical Foundations of Natural Science*) violates the *dichotomy of objects and space* in elementary spatial cognition. In a certain sense it anticipates the later field concept that emerged in nineteenth-century research on electromagnetism, a concept that represents a hybrid of bodily and spatial properties.

The theoretical knowledge resulting from the expansion of experiential spaces has repercussions on different layers of knowledge. Global, geographical coordinates, for instance, attained practical importance in deep-sea navigation. Coastal shipping primarily relies on landmarks. Mediterranean seafaring from the late Middle Ages on could use the magnetic compass complemented by portolan maps displaying compass directions and distances. But for deep-sea navigation knowing one's absolute position is crucial, since in vast regions there are no landmarks and the distances are too large for dead reckoning. After the discovery of electromagnetic radiation, radio navigation became an important tool for spatial orientation at sea.

Theoretical knowledge resulting from the expansion of experiential spaces also has repercussions on theoretical knowledge in general. The insight into the sphericity of the Earth, for instance, which was stabilized by the expanding geographical knowledge, had far-reaching consequences for theories of space, as its central role in Aristotelian physics and cosmology illustrates. The success of electrodynamics, to give another example, inspired the electromagnetic worldview which held that all matter should be reducible to fields. Further, the application of the field model to gravitation lay at the foundation of the development of general relativity, as will be discussed in the following section.

But theoretical knowledge resulting from the expansion of experiential spaces may also have an impact on meta-theoretical knowledge. This is strikingly demonstrated by the influence of Newton's concept of space on Kant's epistemology. Long before writing the *Critique of Pure Reason*, Kant had read the Leibniz-Clarke correspondence and occupied himself with the concept of space, considering aspects of Leibniz's as well as of Newton's conceptions. In the *Critique*, Kant presents space as the pure form of outer intuition and states that[66]

> [w]e can never have a presentation of there being no space, even though we are quite able to think of there being no objects encountered in it.

[66]Kant 1996, 78.

While space is thus a precondition of experience, rather than being derivable from experience, matter is not so, as Kant explains in his post-critical *Metaphysical Foundations of Natural Science*, in which he endeavors to provide a sound metaphysical foundation for Newtonian mechanics. In contrast to space, matter is an 'empirical concept', that is, it requires perceptually given instances in order to attain objective reality.[67] This epistemic divide between space and matter was not part of Newton's philosophy of space. But it was only the autonomy of Newton's concept of space with respect to the concepts of things in space (matter, force) that made Kant's epistemic separation possible. Kant clearly argues on the basis of a container model of space,[68] even though he does not argue for the reality of this container but only for its necessity in cognition.[69] Kant's epistemic separation of space and matter would not have been possible against the background of Aristotelian physics or general relativity, both representing frameworks in which space is (in very different ways) inseparably intertwined with matter.

## 1.7 The decline of an autonomous concept of space

In the previous sections we have argued that more and more general concepts of space emerged under more and more specific cultural conditions. In societies where centralized state administrations took over the social control of space, spatial measures became more standardized and integrated and eventually assumed general arithmetic properties (section 1.4). In societies where oral and written disputation became a social practice, spatial terms formerly used in the context of specific contexts of action attained abstract meanings defined by their position in more encompassing conceptual systems (section 1.5). Under specific historical circumstances in early modern Europe, the integration of different historical strands of knowledge culminated in Newtonian mechanics and gave rise to a concept of space that was not only general but, at the same time, implied the autonomy of space from other physical entities represented by fundamental concepts such as *matter*, *force*, and *time* (section 1.6). With regard to its autonomy the space of this conception was similar to the void of ancient atomism, yet it was clearly not conceived of as *nothing*, but rather as a physical entity in its own right, sometimes even as a *substance*, and often as conceptually prior to the things filling space.

The trend for increasingly general spatial concepts under ever more specific cultural conditions did not continue, however, within institutionalized physics and its neighboring disciplines over the course of the twentieth century. It is true that the concepts of space employed in modern physics are more general than the Newtonian concept in that they pertain to theories that are able to integrate a larger corpus of empirical knowledge. We can give an obvious illustration of this fact by referring to general relativity, which contains Newtonian gravitation theory as a limiting case and, in addition, is able not only to pre-

---

[67] On Kant's empirical concept of matter, see Friedman 2001.

[68] Einstein, on p. xiv in his foreword to Max Jammer's *Concepts of Space* (Jammer 1954, xi–xvi), introduces and discusses the fundamental distinction between the concepts of space as the container for all things and space as the positional quality of all things.

[69] Compare Kant's statement above to the following statement contradicting it, made by David Hume in his *Treatise concerning human nature*: "the ideas of space and time are […] no separate or distinct ideas, but merely those of the manner or order, in which objects exist": "[…] 'tis impossible to conceive either a vacuum and extension without matter, or a time, when there was no succession or change in any real existence" (Hume 2007, 31). Hume is clearly advocating a position-quality concept of space (see the previous footnote).

dict the advancement of the planets' perihelia as well as the bending of light by gravitation with high precision, but also to describe the spacetime dynamics of massive objects such as galaxy nuclei and, in fact, of the universe in its entirety. Yet, in two important respects the Newtonian concept constitutes the historical acme of the generality of concepts of space: it was thought of as fundamental not only for the theory of mechanics from which it arose, but for the physical world in general, regardless of what was considered to be in that space and what discipline described things in space. It was further considered to be universal in the sense that space was the same everywhere: it was homogeneous and isotropic. This property was closely related to its autonomy from other fundamental concepts; since the distribution of things in space (matter and forces, say) is obviously not homogeneous, space has to be decoupled from these things in order to be so.

In twentieth-century physics these two aspects of generality became inapplicable to the developing concepts of space. While the aspiration of formulating fundamental concepts underlying all of physics has always remained a part of the agenda of theoretical physics, and unification is one of the major challenges of present-day theoretical work, there is no concept of space in twentieth-century physics that could consistently be applied to all fields of physics. The same applies to the concepts of time, matter and force. The most advanced concept of space in a well-established theory of modern physics is clearly that contained in the dynamic spacetime of general relativity, which also plays a central role in modern cosmology. At the same time, this concept of spacetime is not compatible with quantum theory, which has so far provided us with the most advanced theory of matter and radiation. Thus, quantum field theory usually presupposes a special-relativistic spacetime, and quantum mechanics is mostly done in non-relativistic space. It is unproblematic, of course, to understand fundamental concepts such as *matter* and *space* differently in the different fields of physics. The point is that, if these different usages are understood as resulting from the consideration of limiting cases to a unifying theory,[70] such a unifying theory has not yet been established and we do not know what its concept of space will look like. There is not even agreement on the way the two fundamental theories of twentieth-century physics, quantum theory and general relativity, are to be combined for an advanced understanding of their relation. Is quantizing general relativity the solution? Or, on the other hand, can gravitation theory explain quantum mechanical measurement?[71]

[70] A limiting case to a theory is understood as the theory that results from the original, more general theory when some dimensional constant of it is taken to be zero, which is just how special-relativistic spacetime results from general relativity in the limiting case of weak gravitational fields. For a detailed account of limiting relations between physical theories, see Ehlers 1986.

[71] This latter view has, for instance, been expressed by Roger Penrose (1989, 348–373). A similar view was expressed by Richard Feynman in a letter to Victor Weisskopf dated January 4 to February 11, 1961: "[…] how can we experimentally verify that [gravitational] waves are quantized? Maybe they are not. Maybe gravity is a way that quantum mechanics fails at large distances" (Feynman papers, Box 66, Folder 7, p. 15, Caltech Archives). In current approaches to an integration of gravity with quantum theory, one can still discern the different viewpoints on the nature of spacetime of the different physics communities. Thus, most varieties of string theory (which grew out of quantum field theory) start with a special-relativistic container-model spacetime (albeit of ten or more dimensions), within which the attempt is made to unify all fundamental interactions, including gravity, in a quantum theoretical framework. A different approach (closer to the spirit of general relativity) is to 'quantize general relativity', thereby attempting to preserve its position-quality view of spacetime (usually referred to as *background independence*). Thus, in Loop Quantum Gravity, a currently successful candidate of this approach, the fundamental objects, the quanta of the gravitational field, are not *in* space. They are nodes in a network of relations (a spin network, technically speaking) and it is quantum superpositions of their aggregates that *constitute* space (Rovelli 2008, 368–369).

The autonomy of space, its independence from time, matter, force, and motion, which was a precondition for its universal homogeneity and isotropy, is lost in twentieth-century physics, too. With special relativity, space becomes entangled with time in such a way that their separation depends on the relative state of motion of the observer and the system under consideration. With general relativity, this spacetime becomes further entangled with matter and force; where the geometry of spacetime is determined by matter (and other forms of energy), and determines the motion of matter and radiation under what was classically considered the gravitational force. Spacetime and matter are entangled so closely that a consideration of the two separately (what is the geometry of spacetime? – How is matter distributed in that spacetime?) can only be done in special cases and only approximatively, while the full theory always demands consideration of both at the same time. Quantum theory provides further intriguing instances of an intertwining of spatial and material concepts, as may be illustrated by reference to non-local phenomena such as quantum entanglement. But however radical the changes quantum theory has effected with respect to the concepts of matter and radiation, it has not (yet) led to a new concept of space. In this book the discussion relating to the decline of an autonomous concept of space is focused on the question of which parts of the experiential knowledge of modern physics had an impact on the concept of space and which parts did not, and how this disparity can be explained (Chapter 7).

The spatial knowledge under discussion in this context is a particular kind of theoretical knowledge, knowledge that develops only in a science that is highly structured in terms of disciplines and sub-disciplines.[72] This knowledge is characterized by a hierarchy of divisions into areas that display specific knowledge structures comprising area-specific concepts, models, and methods. At the same time, different areas are connected by the overlap of certain concepts, models, and methods. Particularly, fundamental concepts such as space, time, energy, matter, and force relate different areas, without necessarily being understood in the same way in every area. Areas may further be connected by objects of study whose treatment requires specific knowledge from more than one area. The knowledge structures within these areas are comparatively stable over periods of time, but knowledge integration across area-boundaries leads to fundamental changes of structure.

The boundaries between the areas shift in various ways over the course of time, resulting in knowledge integration and disintegration, but overarching theories remain a challenge. Thus, the theory of special relativity resulted from the integration of mechanics and electrodynamics into a unified spacetime framework. This led to the temporary disintegration of gravitation, which had formerly been a part of mechanics. The re-integration of gravitation, mechanics and electrodynamics in a unified spacetime framework resulted in the development of general relativity. Quantum mechanics, which had emerged from the consideration of problems on the boundary between thermodynamics and electrodynamics, and further integrated knowledge from mechanics, was faced with the challenge of integrating relativistic field theory. The integration of special-relativistic electrodynamics into a quantum framework – quantum electrodynamics – left the gravitational force – general relativity – standing alone again.[73] In this sub-disciplinary landscape, the two theories of relativity play quite different roles. Special relativity provides the spacetime framework for a large number of

[72] On the differentiation of scientific disciplines from the late eighteenth to the early twentieth centuries, for the case of the physical sciences in Germany, see Stichweh 1984 and Jungnickel and MacCormmach 1986.

[73] The observation of this latter shift of frontier – from a divide between quantum mechanics and field theory to one between quantum field theory and general relativity – is a result of research done by Alexander Blum; see Blum and

sub-disciplinary fields, while general relativity, albeit the more fundamental theory, is comparatively isolated.

The knowledge organized according to disciplines is represented by means of highly specialized technical languages, often employing symbol systems, in particular mathematical formalisms. Empirical knowledge is systematically produced in various subfields. The knowledge resources in their disciplinary configuration define a space of possible transformations and thereby condition the outcome. This means that even in cases in which, historically, the development relies on the particular contribution of a single individual, the configuration of knowledge conditions the outcome of the transformation. In particular the invention of general relativity may appear as contingent on Einstein's peculiar insistence on the incorporation of the equivalence principle in a relativistic theory of gravitation, and his isolated work ensuing from it. Nevertheless, granting the necessity of consolidating gravitation with relativity, and given the knowledge resources of classical mechanics, one is almost inevitably led to spacetime curvature. Thus, despite Gunnar Nordström's more conservative approach to a relativistic theory of gravitation, his final theory exhibits a curved spacetime, as could later be shown.[74] One can advance very basic arguments, requiring energy conservation, deriving the equivalence principle from it, and then showing that in special relativity this inevitably leads to curved spacetime geometry.[75] Theories with a tensor potential starting off in a flat Minkowski spacetime also turn out to exhibit a curved spacetime, once their inconsistencies are eliminated.[76] One may thus conceive of very different historical pathways, probably distributing innovative contributions across more individuals, and combining the classical resources in different temporal order, all eventually leading to a theory very similar to general relativity – or maybe, much less probably, directly to a theory of quantum gravity![77]

## 1.8 Concluding remarks

This introduction started by raising questions about the epistemic status of spatial cognition. What is the relation between predetermined cognitive structures and experience? To what extent are the structures of spatial cognition universal or how far do they depend on cultural conditions? The argument underlying our reasoning was that it is only by studying the history of spatial thinking that the epistemic status of spatial knowledge can be assessed. We then attempted to substantiate this claim by discussing different aspects of the historical development of spatial knowledge and analyzing the epistemic status of the related structures of spatial thinking. In particular, we encountered the following forms of space:

---

Rickles forthcoming. The very synoptic outline given in this paragraph neglects, among other things, the nuclear forces that also played an important role in the history of twentieth-century physics.

[74] Einstein and Fokker 1914. Historically, this result was again a consequence of Einstein's intervention; see Norton 1992. Max Abraham's introduction of a variable line element is another case in point; see Renn 2007, 311–312.

[75] Misner, Thorne, and Wheeler 1973, 177–191.

[76] Misner, Thorne, and Wheeler 1973, 424–425.

[77] One such counter-factual scenario assumes the implementation of the equivalence principle in Newtonian science, leading to a form of classical mechanics that involves an inertio-gravitational field curved in spacetime, so that the step to general relativity becomes almost trivial, once special relativity appears (Stachel 2007). See also Renn and Stachel (2007), who discuss the convergence of David Hilbert's work on the *foundations of physics* with Einstein's theory.

- *Naturally conditioned space* is structured by elementary mental models controlling action and perception, such as the permanent object model and the landmark model.
- *Culturally shared space* is represented in language, culturally conditioned actions and cultural artifacts, and builds upon the mental structures of naturally conditioned space, endowing them with cultural meaning.
- *Administratively controlled space* is represented by measuring tools, arithmetic and linguistic symbols, and schematic drawings, and adds metric significance to structures of the previous forms of space.
- *Mathematically reflected space* generalizes metric structures by abstraction, using diagrams, formalized language, and other symbol systems for its representation.
- *Philosophically reflected space* generalizes linguistically represented elementary structures by elevating them to the rank of principle and exploring the consequences.
- *Empirically and disciplinarily imposed space* results from the integration of knowledge acquired by systematic observation and experimentation employing conceptual-mathematical formalisms.

A central concern of this chapter was to indicate in which ways these forms of space are genetically related. As we have seen, the occurrence of each new form of space depends on specific socio-cultural conditions. Its concrete realization, by contrast, does not solely depend on these conditions, but also on the cognitive structures it builds upon and on further experience. There is thus always both an aspect of construction and an aspect of experience in these spaces. Both aspects are closely intertwined, of course, because experience is always informed by cognitive structures already present in the mind, and, at the same time, it is experience that shapes the development of cognitive structures. One can thus say that there is no experience that is not structured by the mind, but there is also no mental structure that has not been shaped by experience. Our cognitive structures are the sediments of experience. But sedimentation is a historical process. This is why the understanding of the architecture of cognition requires the historical analysis of its genesis. The different forms of space represent not only successive historical stages, however. They also represent forms of thinking that are simultaneously present within single societies. Different forms of spatial knowledge are shared either by the entire society, or by specialized groups, and may affect each other. Within different societies, they coexist in varied manifestations, each society displaying its unique spectrum of expressions of spatial thinking. The following chapters will highlight this diversity of cultural manifestations of spatial thinking through history and provide ample material for the discussion of their genetic relatedness.

## Bibliography

Abbott, B. P. et al. (2016). Observation of Gravitational Waves from a Binary Black Hole Merger. *Physical Review Letters* 116:061102.

Asper, Markus (2009). The two cultures of mathematics in ancient Greece. In: *The Oxford Handbook of the History of Mathematics*. Ed. by Eleanor Robson and Jacqueline Stedall. Oxford: Oxford University Press, 107–132.

Bennett, Andrew T. D. (1996). Do Animals have Cognitive Maps? *The Jounal of Experimental Biology* 199:214–224.

Blum, Alexander and Dean Rickles, eds. (forthcoming). *Quantum Gravity in the First Half of the Twentieth Century: A source book*. Berlin: Edition Open Access.

Boesch, Christophe and Hedwig Boesch (1984). Mental Map in Wild Chimpanzees: An Analysis of Hammer Transports for Nut Cracking. *Primates* 25(2):160–170.

Burenhult, N., ed. (2008). *Language and Landscape: Geographical Ontology in Cross-Linguistic Perspective*. 30. Special Issue of Language Sciences 2–3. Elsevier.

Carnap, Rudolf (1922). *Der Raum. Ein Beitrag zur Wissenschaftslehre*. Berlin: Reuther & Reichard.

Chemla, Karine and Shuchun Guo, eds. (2004). *Les neuf chapitres: Le classique mathématique de la Chine anciennne et ses commentaires*. Paris: Dunod.

Cornell, Edward H. and C. Donald Heth (2004). Memories of travel: Dead reckoning within the cognitive map. In: *Human Spatial Memory: Remembering where*. Ed. by Gary L. Allen. Mahwah, NJ: Erlbaum, 191–215.

Damerow, Peter (1994). Vorüberlegungen zu einer historischen Epistemologie der Zahlbegriffsentwicklung. In: *Der Prozeß der Geistesgeschichte. Studien zur ontogenetischen und historischen Entwicklung des Geistes*. Ed. by Günter Dux and Ulrich Wenzel. Frankfurt a.M.: Suhrkamp, 248–322.

— (1996a). Abstraction and Representation. In: *Abstraction and Representation: Essays on the Cultural Evolution of Thinking*. Boston studies in the philosophy of science. Dordrecht/Boston/London: Kluwer, 371–381.

— (1996b). *Abstraction and Representation. Essays on the Cultural Evolution of Thinking*. Boston studies in the philosophy of science 175. Dordrecht/Boston/London: Kluwer.

— (1998). Prehistory and Cognitive Development. In: *Piaget, Evolution, and Development*. Ed. by Jonas Langer and Melanie Killen. Mahwa, NJ: Erlbaum, 247–269.

— (2000). How Can Discontinuities in Evolution Be Conceptualized? *Culture and Psychology* 6(2):155–160.

— (2001). Kannten die Babylonier den Satz des Pythagoras? Epistemologische Anmerkungen zur Natur der Babylonischen Mathematik. In: *Changing Views on Ancient Near Eastern Mathematics*. Ed. by Jens Høyrup and Peter Damerow. Berlin: Reimer, 219–310.

— (2007). The Material Culture of Calculation. A Theoretical Framework for a Historical Epistemology of the Concept of Number. In: *Mathematisation and Demathematisation. Social, Philosophical and Educational Ramifications*. Ed. by Uwe Gellert and Eva Jablonka. Rotterdam: Sense Publ., 19–56.

— (2012). The Origins of Writing and Arithmetic. In: *The Globalization of Knowledge in History*. Ed. by Jürgen Renn. Berlin: Edition Open Access, 153–173.

Damerow, Peter and Wolfgang Lefèvre (1996). Tools of Science. In: *Abstraction and Representation. Essays on the Cultural Evolution of Thinking*. Ed. by Peter Damerow. Dordrecht/Boston/London: Kluwer, 395–404.

De Risi, Vincenzo, ed. (2015). *Mathematizing Space: The Objects of Geometry from Antiquity to the Early Modern Age*. Cham: Springer.

Descartes, René (1644). *Principia philosophiae*. Amstelodami (Amsterdam): Elzevirium.

— (1984). *Principles of Philosophy*. Synthese Historical Library 24. Dordrecht: Reidel.

Dux, Günter (1992). *Die Zeit in der Geschichte. Ihre Entwicklungslogik vom Mythos zur Weltzeit*. Suhrkamp Taschenbuch Wissenschaft. Mit kulturvergleichenden Untersuchungen in Brasilien (J. Mensing), Indien (G. Dux / K. Kälble / J. Meßmer) und Deutschland (B. Kiesel). Frankfurt a.M.: Suhrkamp.

Ehlers, Jürgen (1986). On Limit Relations between, and Approximative Explanations of, Physical Theories. In: *Logic, Methodology and Philosophy of Science VII*. Ed. by Ruth Barcan Marcus, Georg J.W. Dorn, and Paul Weingartner. Amsterdam: North-Holland, 387–403.

Einstein, Albert and Adriaan D. Fokker (1914). Die Nordströmsche Gravitationstheorie vom Standpunkt des absoluten Differentialkalküls. *Annalen der Physik* 44:321–328.

Elias, Norbert (1988). *Über die Zeit*. Frankfurt a.M.: Suhrkamp.

Folkerts, Menso (1992). Mathematische Probleme im Corpus agrimensorum. In: *Die römische Feldmeßkunst: Interdisziplinäre Beiträge zu ihrer Bedeutung für die Zivilisationsgeschichte Roms*. Ed. by Okko Behrends and Luigi Capogrossi Colognesi. Göttingen: Vandenhoeck & Ruprecht, 311–336.

Foreman, Nigel, Margaret Arber, and Joe Savage (1984). Spatial Memory in Preschool Infants. *Developmental Psychobiology* 17(2):129–137.

Friedman, Michael (2001). Matter and Motion in the 'Metaphysical Foundations' and the First 'Critique': The Empirical Concept of Matter and the Categories. In: *Kant and the sciences*. Ed. by Eric Watkins. Oxford: Oxford University Press.

Gandz, Solomon (1929). The Origin of Angle-Geometry. *ISIS* 12(1929):452–481.

Gärling, Tommy, Anders Böök, and Erik Lindberg (1985). Adults' Memory Representations of the Spatial Properties of Their Everyday Physical Environment. In: *The Development of Spatial Cognition*. Ed. by Robert Cohen. Hillsdale: Erlbaum, 141–184.

Gent, Werner (1971). *Die Philosophie des Raumes und der Zeit. Historische, kritische und analytische Untersuchungen, Bände I und II*. Hildesheim: Georg Olms.

Gentner, Dedre and Albert Stevens, eds. (1983). *Mental Models*. Hillsdale: Erlbaum.

Gladwin, Thomas (1974). *East is a Big Bird. Navigation and Logic on Puluwat Atoll*. Cambridge, MA: Harvard University Press.

Gosztonyi, Alexander (1976). *Der Raum. Geschichte seiner Probleme in Philosophie und Wissenschaften*. Freiburg: Alber.

Guo, Shuchun, ed. (1993). *Zhongguo ke xue ji shu dian ji tong hui: Shu xue juan yi* 中國科學技術典籍通彙*:* 數學卷一. Zhengzhou: Henan jiaoyo chubanshe.

Hazen, Nancy L. (1983). Spatial Orientation. A Comparitive Aproach. In: *Spatial Orientation: Theory, Research, and Application*. Ed. by Herbert Pick and Linda Acredolo. New York: Plenum Press, 3–37.

Heeschen, Volker (1990). *Ninye bún. Mythen, Erzählungen, Lieder und Märchen der Eipo (im zentralen Bergland von Irian Jaya, West-Neuguinea, Indonesien)*. Mensch, Kultur und Umwelt im Zentralen Bergland von West-Neuguinea 20. Berlin: Reimer.

Heth, C. Donald and Edward H. Cornell (1985). A Comparative Description of Representation and Processing During Search. In: *Children's Searching. The Development of Search Skill and Spatial Representation*. Ed. by Henry M. Wellman. Hillsdale: Erlbaum, 215–249.

Høyrup, Jens (1994). *In Measure, Number, and Weight: Studies in Mathematics and Culture*. Albany: State University of New York Press.

— (2002). *Length, Width, Surfaces: A Portrait of Old Babylonian Mathematics and Its Kins*. New York: Springer.

Hume, David (2007). *A Treatise of Human Nature: A Critical edition*. Ed. by David Fate Norton and Mary J. Norton. Oxford: Clarendon Press.

Hyman, Malcolm and Jürgen Renn (2012). Survey: From Technology Transfer to the Origins of Science. In: *The Globalization of Knowledge in History*. Ed. by Jürgen Renn. Berlin: Edition Open Access, 75–104.

Jammer, Max (1954). *Concepts of Space: The History of Theories of Space in Physics*. Cambridge, MA: Harvard University Press.

Jeffares, Ben (2010). The Co-Evolution of Tools and Minds: Cognition and Material Culture in the Hominin Lineage. *Phenomenology and the Cognitive Sciences* 9:503–520.

Jungnickel, Christa and Russell MacCormmach (1986). *Intellectual Mastery of Nature: Theoretical Physics from Ohm to Einstein*. 2 Vols. Chicago: University of Chicago Press.

Kangshen, Shen, John N. Crossley, and Anthony W.-C. Lun, eds. (1999). *The Nine Chapters on the Mathematical Art: Companion and Commentary*. Oxford: Oxford University Press.

Kant, Immanuel (1996). *Critique of Pure Reason: Unified edition; with all variants from the 1781 and 1787 editions*. Indianapolis: Hackett.

— (1997). *Metaphysische Anfangsgründe der Naturwissenschaft*. Hamburg: Meiner.

— (1998). *Kritik der reinen Vernunft (1. & 2. Aufl.)* 505. Philosophische Bibliothek. Hamburg: Meiner.

Kitchin, Rob and Mark Blades (2002). *The Cognition of Geographical Space*. London: Tauris.

Koch, Gerd (1984). *Maligdam. Ethnographische Notizen über einen Siedlungsbereich im oberen Eipomek-Tal, zentrales Bergland von Irian Jaya, West-Neuguinea, Indonesien.* Mensch, Kultur und Umwelt im Zentralen Bergland von West-Neuguinea 15. Berlin: Reimer.

Koch, Gerd and Wulf Schiefenhövel (2009). *Eipo (West-Neuguinea, Zentrales Hochland) - Neubau des sakralen Männerhauses in Munggona*. DVD. Produktionsjahr: 1974, first published in 1979; IWF Bestellnummer/Bandzählung E 2475.

Langer, Jonas (2001). The mosaic evolution of cognitive and linguistic ontogeny. In: *Language Acquisition and Conceptual Development*. Ed. by Melissa Bowerman and Stephen C. Levinson. Cambridge: Cambridge University Press, 19–44.

Lefèvre, Wolfgang (1981). Rechensteine und Sprache. In: *Rechenstein, Experiment, Sprache. Historische Fallstudien zur Entstehung der exakten Wissenschaften*. Ed. by Peter Damerow and Wolfgang Lefèvre. Stuttgart: Klett-Cotta, 115–169.

— (1984). Die Wissenschaft in der geschichtlichen Entwicklung des Menschen. In: *Kindlers Enzyklopädie: Der Mensch*. Ed. by Norbert Loacker. Band 7. Zurich: Kindler, 295–328.

Levinson, Stephen C. and David Wilkins, eds. (2006). *Grammars of Space*. Cambridge: Cambridge University Press.

Lévi-Strauss, Claude (1955). *Tristes tropiques*. Paris: Plon.

Little John, J. (1963). Temne Space. *Anthropological Quarterly* 36(1):1–17.

Lorenz, Konrad (1977). *Behind the Mirror. A Search for a Natural History of Human Knowledge*. London: Methuen.

Lorenzen, Paul (1984). *Elementargeometrie. Das Fundament der Analytischen Geometrie*. Mannheim: Bibliographisches Institut.

Lyons, Henry (1927). Ancient Surveying Instruments. *The Geographical Journal* 69(2): 132–139.
Marr, David (1982). *Vision: A Computational Investigation in the Human Representation of Visual Information*. San Francisco: Freeman.
Menzel, Emil W. (1973). Chimpanzee Spatial Memory Organization. *Science* 182(4115): 943–945.
— (1987). Behavior as a Locationist Views it. In: *Cognitive Processes and Spatial Orientation in Animal and Man. Volume I. Experimental Animal Psychology and Ethology*. Ed. by Paul Ellen and Catherine Thinus-Blanc. Dordrecht: Martinus Nijhoff Publisher, 55–72.
Michel, Thomas (1983). *Interdependenz von Wirtschaft und Umwelt in der Eipo-Kultur von Moknerkon. Bedingungen für Produktion und Reproduktion bei einer Dorfschaft im zentralen Bergland von Iran Jaya, West-Neugiunea, Indonesien*. Mensch, Kultur und Umwelt im zentralen Bergland von West-Neuguinea 11. Berlin: Reimer.
Minsky, Marvin (1975). A Framework for Representing Knowledge. In: *The Psychology of Computer Vision*. Ed. by Patrick Henry Winston. New York: McGraw-Hill, 211–277.
Misner, Charles W., Kip S. Thorne, and John A. Wheeler (1973). *Gravitation*. New York: Freeman.
Neisser, Ulric (1976). *Cognition and Reality: Principles and Implications of Cognitive Psychology*. San Francisco: Freeman.
Neugebauer, Otto (1934). *Vorlesungen ueber Geschichte der antiken mathematischen Wissenschaften: Band 1, Vorgriechische Mathematik*. Berlin: Springer.
Newcombe, Nora S. and Janellen Huttenlocher (2003). *Making Space. The Development of Spatial Representation and Reasoning*. Cambridge, MA: MIT Press.
Normand, Emmanuelle and Christophe Boesch (2009). Sophisticated Euclidean Maps in Forest Chimpanzees. *Animal Behaviour* 30:1–7.
Norton, John D. (1992). Einstein, Nordström and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation. *Archive for History of Exact Sciences* 45:17–94.
Odling-Smee, F. John, Kevin N. Laland, and Marcus W. Feldman (2003). *Niche Construction: The neglected process in evolution*. Princeton, NJ: Princeton University Press.
Penrose, Roger (1989). *The Emperor's New Mind: Concerning computers, minds, and the laws of physics*. New York, NY: Oxford University Press.
Piaget, Jean (1959). *The Construction of Reality in the Child*. 5th print. The Basic Classics in Psychology. New York: Basic Books.
— (1970). *Genetic Epistemology*. New York: Columbia University Press.
— (1981). *The Psychology of Intelligence*. Totowa: Littelfield, Adams & Co.
— (1983). *Biologie und Erkenntnis*. Frankfurt am Main: Fischer.
Piaget, Jean and Bärbel Inhelder (1956). *The Child's Conception of Space*. London: Routledge & Kegan Paul.
Piaget, Jean, Bärbel Inhelder, and Alina Szeminska (1960). *The Child's Conception of Geometry*. Digital reprint 2007. Abingdon: Routledge.
Pick, Herbert and Linda Acredolo, eds. (1983). *Spatial Orientation: Theory, Research, and Application*. New York: Plenum Press.
Renn, Jürgen (2004). The Paradox of Scientific Progress. Notes on the Foundation of a Historical Theory of Knowledge. In: *Research Report 2002-2003*. Max Planck Institute for the History of Science, 21–49.

Renn, Jürgen (2005). The Relativity Revolution from the Perspective of Historical Epistemology. *Isis* 95(4):640–648.

— (2007). The Summit Almost Scaled: Max Abraham as a Pioneer of a Relativistic Theory of Gravitation. In: *Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy*. Ed. by Jürgen Renn and Matthias Schemmel. Dordrecht: Springer, 305–330.

Renn, Jürgen and Peter Damerow (2007). Mentale Modelle als kognitive Instrumente der Transformation von technischem Wissen. In: *Übersetzung und Transformation*. Ed. by Hartmut Böhme, Christof Rapp, and Wolfgang Rösler. Berlin: de Gruyter, 311–331.

Renn, Jürgen and John Stachel (2007). Hilbert's Foundation of Physics: From a Theory of Everything to a Constituent of General Relativity. In: *Gravitation in the Twilight of Classical Physics: The Promise of Mathematics*. Ed. by Jürgen Renn and Matthias Schemmel. Dordrecht: Springer, 857–973.

Robson, Eleanor (2008). *Mathematics in Ancient Iraq. A Social History*. Princeton: Princeton University Press.

Rovelli, Carlo (2008). *Quantum Gravity*. Cambridge: Cambridge University Press.

Savage-Rumbaugh, Sue (1998). Scientific Schizophrenia With Regard to the Language Act. In: *Piaget, Evolution, and Development*. Ed. by Jonas Langer and Melanie Killen. Mahwah, NJ: Erlbaum, 145–169.

Savage-Rumbaugh, Sue and William Fields (2000). Linguistic, Cultural and Cognitive Capacities of Bonobos (Pan Paniscus). *Culture and Psychology* 6:131–153.

Schemmel, Matthias (2016). *Historical Epsitemology of Space: From Primate Cognition to Spacetime Physics*. Cham: Springer.

Schick, Kathy D., Nicholas Toth, Gary Garufi, E. Sue Savage-Rumbaugh, Duane M. Rumbaugh, and Rose A. Sevcik (1999). Continuing Investigations into the Stone Tool-Making and Tool-Using Capabilities of Bonobo (Pan paniscus). *Journal of Archaeological Science* 26:821–832.

Schiefsky, Mark (2012). The Creation of Second-Order Knowledge in Ancient Greek Science as a Process in the Globalization of Knowledge. In: *The Globalization of Knowledge in History*. Ed. by Jürgen Renn. Berlin: Edition Open Access, 191–202.

Schurig, Volker (1976). *Die Entstehung des Bewußtseins*. Frankfurt a.M./New York: Campus.

Senft, Gunter, ed. (1997). *Referring to Space: Studies in Austronesian and Papuan Languages*. Oxford: Clarendon Press.

Siegel, Alexander W. and Sheldon H. White (1975). The Development of Spatial Representation of Large-Scale Environments. In: *Advances in Child Development and Behavior*. Ed. by Hayne W. Reese. New York: Academic Press.

Sigg, Hans and Alexander Stolba (1981). Home Range and Daily March in a Hamadryas Baboon Troop. *Folio primatologica* 36:40–75.

Stachel, John (2007). The Story of Newstein or: Is Gravity Just Another Pretty Force? In: *Gravitation in the Twilight of Classical Physics: The Promise of Mathematics*. Ed. by Jürgen Renn and Matthias Schemmel. Dordrecht: Springer, 1041–1078.

Stichweh, Rudolf (1984). *Zur Entstehung des modernen Systems wissenschaftlicher Disziplinen: Physik in Deutschland 1740 - 1890*. Frankfurt: Suhrkamp.

Tomasello, Michael (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Tomasello, Michael and Josep Call (1997). *Primate Cognition*. Oxford: Oxford University Press.

Tomasello, Michael, Ann Cale Kruger, and Hilary Horn Ratner (1993). Cultural learning. *Behavioral and Brain Sciences* 16(3):495–511.

Tuan, Yi-Fu (1975). Images and Mental Maps. *Annals of the Association of American Geographers* 65(2):205–214.

Vogel, Kurt, ed. (1968). *Neun Bücher arithmetischer Technik: Ein chinesisches Rechenbuch für den praktischen Gebrauch aus der frühen Hanzeit (202 v. Chr. bis 9 n. Chr.)* Braunschweig: Vieweg.

Vollmer, Gerhard (1994). *Evolutionäre Erkenntnistheorie*. Stuttgart: Hirzel.

Wallace, Ron (1989). Cognitive Mapping and the Origin of Language and Mind. *Current Anthropology* 30(4):518–526.

Wang, Ranxiao Frances and Elizabeth S. Spelke (2002). Human Spatial Representation: Insights from Animals. *Trends in Cognitive Sciences* 6:376–382.

Williams, Barbara J. and María del Carmen Jorge y Jorge (2008). Aztec Arithmetic Revisited: Land-Area Algorithms and Acolhua Congruence Arithmetic. *Science* 320(72):13–27.

# Chapter 2
# Spatial Concepts in Non-Literate Societies: Language and Practice in Eipo and Dene Chipewyan

*Martin Thiering and Wulf Schiefenhövel*

## 2.1 Introduction

This chapter focuses on the linguistic representation of spatial concepts in two little known and unrelated languages with a non-written tradition. It explores the degree to which environmental experience and spatial orientation is reflected in language, i.e., it is in line with anthropological linguistic approaches placing language in its social and cultural context, and its cultural practices.[1] As such, spatial knowledge is not only encoded in concepts or categories, but is embodied in the lived histories of human beings, and their cultural and linguistic practices.[2]

The unrelated cultures under survey present interesting environmental terrains: one is an alpine region (Eipo), the other comprises vast prairies (Dene). The mental and perceptual course-maintaining processes in these cultures rely on cognitive maps.[3] We assume very fundamentally that *Homo sapiens*, like all other animals, is equipped with biological, especially neurobiological dispositions enabling orientation in space and thereby ensuring survival and, ultimately, reproduction. As has been argued in Chapter 1 of this book, the ability of cognitive mapping is part of this biological disposition. Cognitive maps are structures of spatial reasoning; they are processes of unconscious inference.[4] We understand cognitive maps as establishing a relation between the 'real world' cues (such as objects and places) and their mental equivalents. This will give us the opportunity to relate environmental conditions to structures of spatial cognition as they are reflected in linguistic and enactive presentations.

This chapter deviates from the descriptions of landscape features in the sense that it adopts cognitive maps that are referred to in navigation techniques of orientation, i.e., navigating without instruments. We argue that this kind of navigation is based on dynamic cognitive maps and mental triangulation. This enables the navigators to have a spatial conception of their position at any time. It is argued here that this is of special importance not only for piloting but also for orienting oneself on land. We show this for the alpine regions of the Eipo and the vast prairies extensions of the Dene in Alberta.

We adopt the premise that[5]

---

[1] Foley 1997, Mark et al. 2011.
[2] Foley 1997, 177.
[3] Portugali 1996.
[4] Knauff 2013.
[5] Siegel and White 1975, 11.

> descriptions of space, or allusions to space in language, must rest on two kinds of knowledge. The first appears to be based on models (maps, representations) which people construct to guide *spatial behavior*. The second appears to consist of a linguistic symbol-system that allows the models to be shared within a community of discourse.

The question is whether there are commonalities between the two unrelated languages, and if differences appear, what form do they take linguistically and conceptually? The following quote summarizes our point of departure.[6]

> Man, in confronting reality, faces a kaleidoscope of phenomena ranging from the natural to the man-made, to the imaginary, to the totally abstract. Comprehension of such a broad inventory of reality and non-reality requires language, the tool that permits man to take verbal stock of objective and subjective experiences alike. In man's ongoing endeavor to conceptualize and verbalize a world that can never be fully known, language is the vital intermediary.

Our question here concerns the relationship between non-linguistic information and spatial language. One language, Eipo, is spoken in the central mountains of the Indonesian Province of Papua, formerly the province of Irian Jaya, West New Guinea. The other language, Dene Chipewyan, is spoken in Cold Lake, Alberta. The point of departure in our argumentation is that non-linguistic information has its impact upon spatial language and categorization, i.e., with reference to space and its relation to semiotic systems. We present language data indicating the influence of environmental landmarks and cultural heritage in shaping spatial categorization in the two languages. In this chapter landmarks are defined as any kind of environmental reference points. This can be a mountain, a river, a house, or even a tree (see section 2.2).

In accordance with the exposition given in Chapter 1, it is assumed that spatial concepts develop in the course of ontogeny on the basis of cognitive structures resulting from phylogeny. This development depends on the experiences of a speaker and the common concepts in the speaker's community in a particular culture at a particular time. In the course of our argumentation we present some fundamental spatial concepts and representations based on anthropomorphic spatial knowledge in Eipo and Dene Chipewyan. Knowledge members of both cultures developed on the basis of human phylogenetic adaptations throughout their ontogenesis in a remote area in West New Guinea and Western Canada. The term *culture* has several meanings and theoretical backgrounds. We adopt the specific idea of *culture* following Clifford Geertz's *Interpretation of Culture*:[7]

> The concept of culture is essentially a semiotic one. Believing that man is an animal suspended in webs of significance he himself has spun, I take culture to be those webs, and the analysis of it to be therefore not an experimental science in search of law but an interpretative one in search of meaning.

We show such webs of basic spatial categorization in the two cultures, i.e., we present a snapshot of spatial semantics represented by the two languages. Moreover, this chapter posits its

[6] Malotki 1983, 13.
[7] Geertz 1999, 5.

arguments on the basis of species-specific cognitive organization that matures and shapes in the course of ontogenesis during sensorimotor action and sociocultural learning.[8] Spatial cognition is externally represented in language as well as in cultural-specific practices.[9] Note that language is understood here as an external representation of mental concepts, or, as Boas puts it, human language is one of the most important manifestations of mental life.[10]

The chapter is structured as follows: we first present some theoretical fundamentals of cognitive linguistics (section 2.2), followed by anthropological outlines of Dene Chipewyan (section 2.3) and Eipo (section 2.4). We then present some selected examples of spatial concepts in Eipo (center and periphery and natural limitations, distance, and orientation in Eipo; section 2.5). Finally, we compare representations of spaces in Dene and Eipo based on a variety of data sets (section 2.6). For the case of the Eipo, data are used from the dictionary of the Eipo language containing actual usages of the recorded utterances as well as published material from Schiefenhövel and Heeschen.[11] Additionally, we rely on a collection of myths, songs, and stories from Eipo speakers.[12] For the case of Dene, first hand data were elicited by Thiering with Dene Chipewyan speakers, based on various elicitation tools and interviews.[13] We conclude the chapter with some general comments (section 2.7).

## 2.2 Theoretical frame

### 2.2.1 Cognitive maps

Descriptions of space are based on internal models of knowledge representation of the environment. Such models are defined in cognitive psychology as *mental models* (or, depending on the authors, using concepts such as *scripts*, *slots*, *frame-systems*, *fillers*, *schemas*, *idealized cognitive models*, *mental spaces*). More specifically, *cognitive maps* represent the geometric layout of the differentiated topography of a space (via toponyms). By definition, a cognitive map or survey representation of a spatial layout encodes relations (distances and directions) among behaviorally relevant *landmarks* within a coordinate reference system centered on the environment. We use the term coordinate system rather loosely, or as an analogy, since, in the context of practical orientation, we do not believe in a mathematical coordinate system represented in the brain. Still, in the case of spatial conceptualization the analogy helps to model and describe the cognitive function of representing environmental frames of reference as a cognitive device.

Cognitive maps function to support navigation, and, in turn, are created by navigation and exploration of large-scale space. During navigation and exploratory spatial behavior, landmarks are experienced sequentially in space and time. The process of constructing a cognitive map can be thought of as a process that places a mental 'copy' of each sequentially experienced landmark into a simultaneous system that preserves metric information about the linear distance between landmarks, and their direction relative to one another.

---

[8] Piaget and Inhelder 1956.
[9] Foley 1997, 169–178. See also Chapter 1 of this book.
[10] Boas 1977, 68.
[11] Heeschen and Schiefenhövel 1983.
[12] Heeschen 1990; there is also a rich collection of film material on the Eipo's daily activities and cultural practices; see Blum et al. 1979–1996.
[13] Thiering 2006, Thiering 2009a, Thiering 2010; field notes by Thiering.

An important, emergent property of a simultaneous system is that the spatial relations between landmarks entered in the system, even those relations not directly experienced, are also available.

Cognitive maps express the essential structure of spatial information encoded in our memories through learning processes. Like cartographic maps, cognitive maps can be constructed using many different sources of information and encoding processes. Some cognitive maps may be stored as permanent structures in long-term memory, e.g., a cognitive map of a familiar city, while others may be temporary structures for the current state of a dynamic environment, e.g., parents keeping track of the locations of children as they play in a park. In either case the characteristics of objects are thought to be stored along with their spatial locations. Hence, a cognitive map is, in the simplest terms, the encoding of a structure in our memory of what is where, i.e., such maps are essentially individualized internal representations or models of the worlds in which we live.

The processes used to acquire spatial knowledge appear to have a fundamental impact on the character of a cognitive map. The nature of cognitive maps produced by different encoding processes and the focus on understanding the circumstances that produce cognitive maps with fixed orientations and those that produce orientation-free cognitive maps is at issue here. Cognitive mapping is[14]

> the process composed of a series of psychological transformations by which an individual acquires, stores, recalls, and decodes information about relative locations and attributes of the phenomena in his everyday spatial environment.

The end product of a cognitive mapping process is a cognitive map.[15] Cognitive mapping is a recording process in memory of the existence of an object and its known location in space. Within a given visual image, a large number of landmarks are simultaneously visible, so relative distances and directions are easy to judge.[16]

The next subsection examines the usage of cognitive maps with respect to landmarks serving as anchorage points to navigate and orient oneself in a known and unknown environment.

### 2.2.2 Landmarks

At focus in the very different environments under review, i.e., alpine vs. prairies, are landmarks as external points of reference. Moreover, in this chapter landmarks are defined as any kind of cultural-specific environmental reference points. This can be the above-mentioned mountains, rivers, houses, rocks, or even a tree. Landmarks are points of reference external to the person. In a city, landmarks may be distant buildings or geographical features that can be seen from many angles and distances, or they may be primarily local such as buildings, signs, trees, storefronts, doorknobs, or other urban details.[17] Siegel and White argue that landmarks are unique configurations of perceptual events (patterns). They identify a specific geographical location. A person's account of his spatial representations generally

[14]Downs and Stea 1973, 7.
[15]Tolman 1948.
[16]Kuipers 1982, 203.
[17]Miller and Johnson-Laird 1976, 378.

begins with landmarks, and these landmarks are the strategic foci to and from which the person moves or travels. Landmarks are used as proximate course-maintaining devices. They not only identify beginnings and endings, but also serve to maintain course.[18]

Arguably, landmarks shape and determine a detailed topographical map of the environment as represented via language. The following quote by Fowler and Turner summarizes the function of landmarks or geographic features in particular. This quote also summarizes our point of departure with respect to the function of environmental knowledge and its reflection in language.[19]

> The naming of geographic features as part of territorial marking and orientation is a common occurrence in all cultures […]. Usually, topographical names reflect specific cultural interests and historical developments within the possibilities given by the morphology of the language.

Fowler and Turner clearly point out that the process of naming geographic and territorial landmarks is crucial in all cultures. More specifically they conclude that topographical names indicate particular cultural interests as represented by the language repertoire, toponyms, or the language-specific affordances. Indeed, data presented here show a rather dense linguistic system of topographical maps represented, e.g., in place names serving as mental maps for orientation. It is furthermore argued that human beings instantiate relations between objects relying on various frames of reference that, as the name implies, serve as reference points to locate participants (see below). These reference points anchor a specific orientation between objects and the viewer.[20] These linguistic coordinates are important for the description of topographical spatial relations in Dene and Eipo, as they are for the description of projective relations in general.[21]

It is believed that travelers locate their current position on the Earth's surface symbolically within a cognitive map. For orientation in the environment relying on toponyms the traveler must compare the necessary direction of travel toward the destination using the respective cognitive map for orientation. To conduct a survey without instruments, distance and heading are conceptualized as movement, or change of position, within a cognitive map. At any time the traveler can estimate distance from and direction of known points such as the starting point. Hence, the difficult aspect is to retain a sense of direction especially without any visible landmarks, as in dead reckoning navigation.[22] As we argue, orientation processes on sea as well as on land are based on some fundamentals in mental triangulation and gestalt theoretic conceptions of spatial relations (figure-ground asymmetries; see below).

A prominent example from orientation on water comes from navigation without instruments.[23] More specifically, one method in navigation is dead reckoning. It depends on determining one's position at any time based on the distance and direction traveled since leaving the last known location.[24] The navigator monitors the motion of the boat to deter-

[18] Siegel and White 1975, 23.
[19] Fowler and Turner 1999, 424. For a detailed categorization of spatial relations, see Miller and Johnson-Laird 1976, 377.
[20] L. Carlson-Radvansky 1993, L. Carlson-Radvansky and G. Carlson-Radvansky 1996, Carlson and Logan 2001, Carlson 2003, Levinson 2003, Levinson and Wilkins 2006.
[21] Malotki 1983, 16.
[22] Gladwin 1974, Hutchins 1996, Sarfert 1911.
[23] Hutchins 1996, 65–93; see also Hutchins 1983 and Chapter 1 in the present book.
[24] Gladwin 1974, 144.

mine the displacement from a previous position.[25] This mental computing or mental triangulation, i.e., the transformation and propagation of representational states, is arguably also used on land.[26] In addition to this method, travelers' reports, stories, symbols, icons or any other kind of representation are are also examined to reconstruct cognitive maps of spatial orientation based on implicit knowledge systems.

Cognitive maps underly cognitive information-processing systems of spatial perception.[27] As is argued here, the specific encoding patterns vary in the orientation reference systems. Moreover, we consider spatial reference frames that construe a complex mental model or gestalt-like representation of knowledge. As such, course-maintaining systems on land and at sea based on different sorts of texts are of specific interest. The rationale behind this is to argue for describing cognitive maps as gestalt-like representations of environmental cues forming a dynamic mental model or cognitive map. What might be common to all cultures and hence be universal is the gestalt-like constructive process of cognitive maps. These cognitive maps function as implicit knowledge systems that enable people to navigate in a specific environment at a given time and space.

With respect to spatial orientation, Fowler and Turner point out:[28]

> If peoples choose to orient themselves to coasts or seas, rivers or mountains, the Sun's path, or some other feature, some aspect of this will usually show up in their place names.

Adopting Fowler and Turner's point it will be shown that people in both of the cultures discussed here use place names in their specific environments to construct a linguistically dense topographical reference system for orientation. Hence, environmental experience is also represented via language, and language in turn shapes spatial concepts or mental models.[29] We will also present the rich fabric of terms of spatial deixis in both cultures under study. This highlights the importance of the notion of frames of reference here since they profile spatial relationships between the speaker-hearer and the environment.

### 2.2.3 Frames of reference

It is argued that human beings instantiate relations between objects relying on various frames of reference. Reference points are fundamental in ascribing specific orientations between objects.[30] These linguistic coordinates are important for the description of spatial topographical relations such as *an*, *on*, and *in*, in Dene and Eipo, as they are for the description of projective left-right relations in general.[31] Following Malotki, the term linguistic coordinate here means the division of a spatial configuration into a speaker, a hearer and a third part (a person or a thing the speaker-hearer refers to). Hence, a linguistic coordinate system is not a geographically or mathematically abstract concept, but a means of spatial categorization in the linguistic encoding.

---

[25] Hutchins 1996, 56.
[26] Hutchins 1996, 49.
[27] Marr 1982.
[28] Fowler and Turner 1999, 424.
[29] Thiering 2012.
[30] L. Carlson-Radvansky 1993, L. Carlson-Radvansky and G. Carlson-Radvansky 1996, Carlson 1999, Carlson and Logan 2001, Carlson 2003, Levinson 2003, Levinson and Wilkins 2006.
[31] Malotki 1983, 16.

The encoding of spatial relations depends on certain spatial (and temporal) parameters that set the linguistic coordinate reference system for the speaker-hearer. In general, spatial marking is based on three different reference frames to be selected from. These are assigned to the objects profiled in the situation.[32] The three frames of reference can be divided into

1. a viewer/ego-centered or relative frame, as in the English example *he's to the left of the house* (assuming that from the perspective of the viewer, a person is situated to the left side of the house),
2. an object-centered or intrinsic frame, as in *he's in front of the house* (assuming that the front is where the main door is located; the object has an inherent front and back side), and
3. an environment-centered or absolute frame, as in *he's north of the house*.

In (1), the viewpoint depends on the location of the viewer's vantage point and his/ her relation to the figure and ground. The intrinsic frame in (2) is an object-centered reference system determined by natural or culture-specific inherent features of the object. Finally, the absolute frame (3) is a fixed direction provided by, e.g., cardinal direction.[33]

With respect to the figure-ground asymmetry we follow Talmy's adaptation of the *Gestalt* psychologist approach arguing that certain cognitive categories play an important role in attributing the primary and secondary objects of a scene.[34] These functions are encoded by the figure and ground of a scene, the variable element or positive space versus the reference element or negative space.[35] The former is usually the smaller and moveable object whereas the latter is usually the permanently located, larger object.[36] For more details, see the subsection following the next.

The Language and Cognition group at the Max Planck Institute in Nijmegen provides an exception to standard procedures in armchair linguistics. Elicitation tools developed by the researchers of this group facilitate the gathering of data from actual speakers and their usage of a particular language.[37] We argue that these ascriptions are determined by cultural, environmental and language-specific affordances.[38] These, in turn, depend on speaker-imposed figure-ground asymmetries that are attributed to the respective objects.[39] Another important concept for the discussion of spatial concepts is that of *ideas of space*.

### 2.2.4 Ideas of space

We argue that ideas of space (*Raumbilder*),[40] i.e., the speaker's basic delimitation of his/her world of experience, are important in Eipo and Dene, as in any other language and culture. A selection of such ideas of space are, for example, the deictic parsing of space into 'here', 'there', and 'over there' or simply 'celestial space' versus the 'Earth' as encoded via 'above'

[32] L. Carlson-Radvansky 1993, Carlson 1999, Carlson 2000, Carlson and Logan 2001, Carlson 2003, Levinson 2003, Coventry and Garrod 2004.
[33] For an extensive overview, see Levinson 2003.
[34] Talmy 1983, 230.
[35] Talmy 1978, 627, Hofstadter 1980, Talmy 1983, 232, Talmy 2000.
[36] See Talmy's 20 parameters for the domain of spatial configurations of figure-ground asymmetries; Talmy 1983, 277.
[37] Levinson 2003, Levinson and Wilkins 2006.
[38] Whorf 1956, Wygotski 1964, Watzlawick 1981, Hunt and Agnoli 1991, Lucy 1992b, Lucy 1992a.
[39] Talmy 1978, Talmy 1983, Talmy 2000.
[40] Malotki 1979.

and 'down'. We have also *ideas of space* such as the 'left' and 'right' asymmetries, 'in front of' and 'behind', 'up' and 'down', 'near' and 'far away', 'inside' and 'outside', 'in' and 'on', the cardinal directions 'North', 'South', 'West', and 'East', 'back' and 'forward', man-made places such as a 'house' and 'geographic places' or 'surfaces'.[41] Note that in contrast to Hopi, the Eipo language does not have true terms for cardinal directions, yet we believe that expressions such as 'downstream' and 'upstream' have a similar semantic function. Hence, it may be stated that Eipo also evokes a tripartite system of deictic reference. Note that this three way separation is similar to, e.g., German *hier* 'here', *da* 'there', *dort* 'over there' differentiating between proximal and distal distances, taking the speaker as the anchor of her/his perspective.

Malotki's survey presents various facets of Hopi encoding of spatial relations and demonstrates a 'degree of specificity'.[42] This linguistic phenomenon of the figure's location with respect to the ground is related to the amount of detailed expressive content with which spatial relations are described in various languages.[43] It is claimed that, for example, the English prepositional phrase 'X is on the table' has a lower degree of specificity than the corresponding expression in other languages such as 'X is located at the table's upper surface'[44] as is the case in Ewe, a language spoken in the south-east of Ghana. The latter specification encodes further partitions of the table into smaller regions.[45]

In Chapter 5 of his analysis, Malotki[46] gives a detailed account including various illustrations of the different representations of space and spatial semantics as linguistically summarized in a total of 43 locational morpheme markers specifying space in Hopi;[47] the alphabetically ordered spatial morphemes are described in terms of content or semantic fields in Malotki's concluding remarks.[48] He states that Hopi uses a fine-grained linguistic system to encode spatial relations and, we would add, spatial concepts that also differ, to a certain degree, from most other languages.[49] This should be of no surprise since every language presents language-specific affordances, i.e., the semantic content hard-wired into specific morphosyntactic devices or morphosyntactic patterns. As such, spatial concepts are linguistically represented in different forms which are based in the respective language system. Malotki concludes that[50]

> owing to its differentiated construction of the locative with its punctive and diffuse subsystems as well as the locative and the destinative with their extreme and non-extreme partitions, respectively, the Hopi language forces its speakers to a sharper observation of certain areas of spatial reality than most other SAE languages.

---

[41] Malotki 1979, 294,297.
[42] Svorou 1993, Thiering 2013.
[43] Svorou 1993.
[44] Svorou 1993, 6–8, Langacker 2008, 19, 43, 55–57.
[45] Ameka 2006, 371.
[46] Malotki 1979, 144–261.
[47] Malotki 1979, 145–146.
[48] Malotki 1979, 295, 298.
[49] Malotki 1979, 293.
[50] SAE stands for *Standard Average European*. The German original reads: "die Hopi-Sprache auf Grund ihrer differenzierten Gestaltung des Lokativs mit seinen punktiven und diffusen Subsystemen sowie des Lokativs und Destinativs mit ihren extremen bzw. nicht-extremen Untergliederungen ihre Sprecher zu einer schärferen Beachtung gewisser Bereiche der räumlichen Realität zwingt, als dies die meisten SAE-Sprachen tun." (Malotki 1979, 299).

Thus, Malotki claims that Hopi-speakers are forced by their language, and, as we assume, by the environment, to pay more attention to spatial reality. He does not claim that this necessarily implies radical differences between the Hopi's 'Weltbild' and that of speakers of other languages.[51] He points out that the Hopi's idea of space might contain culture- and language-specific elements. Malotki believes that in particular aspects of spatial relations a difference in focus might lead to differences in thinking about space. This belief may be interpreted as an adherence to a modest form of linguistic relativism.[52]

Summing up, Malotki concludes that the Hopi language uses a fine-grained linguistic system to encode spatial relations. We would add that this language additionally uses spatial concepts that also differ from most other languages.[53] Similarly fine-grained spatial distinctions can be found in languages of other cultures. As an example we point to the spatial deixis terms used by peoples in the Alpine regions of Europe, which reflect a very precise relationship between the environment and language similar to that of Hopi.[54] As we demonstrate, Eipo and Dene Chipewyan also present crucial environment-dependent encoding patterns mirrored in the languages. The mountains and rivers as important limitations in Eipo, or lakes, in particular Cold Lake, rocks, trees and rivers in the Dene culture, show their repercussions in the language patterns and the carving-up of spatial concepts on the language level.

### 2.2.5 Figure-ground asymmetries

As we have seen, one of the major hypotheses in cognitive psychology (which was the precursor to cognitive linguistics) is the idea of mental representations as abstract schemas or mental models.[55] We know from gestalt psychological approaches that such schemas are supposedly universal and not language-specific. Moreover, they are non-linguistic mental representations of experience. They are extracted from more specific structures and categorize such structures through relations of full or partial schematicity.

The idea of mental representations leads more specifically to the general claim in cognitive linguistics that all grammatical structures are symbolic. Additionally, the lexicon, morphology, and syntax form a continuum of symbolic units, each residing in the association of a semantic and a phonological structure or pole.[56] Moreover, the meanings of linguistic expressions are conceptualizations shaped in accordance with the linguistic system. In addition, all facets of our general knowledge of a conceived entity contribute to the meaning of an expression which designates this entity and, given that, any sharp distinction between semantics and pragmatics is gratuitous.[57] Semantics is, in this view, not an autonomous cognitive module, nor is the linguistic system overall.

With respect to semantic structures it is claimed that they are predications that are characterized relative to cognitive domains such as time, space, and color. Most domains of

[51] Malotki 1979, 301.
[52] Malotki 1979, 301.
[53] Malotki 1979, 293.
[54] Berthele 2006.
[55] Gentner and Stevens 1983, Johnson-Laird 1983, Penrose 1991, Ritter, Martinetz, and Schulten 1991, Schade 1992, Schreuder and Flores d'Arcais 1989, Strube 1996.
[56] Langacker 1987.
[57] Nunberg 1978, Sweetser 1990.

linguistic relevance are non-primitive. That means they are interrelated networks.[58] As such, they involve cognitive structures of indefinite complexity, i.e., we have layers of interrelated networks that can be modeled in a connective fashion.[59] Any cognitive structure can function as the domain for a predication.[60] Moreover, meaning is conceived as cognitive processing, and even expressions used to describe a presumably objective situation may differ in meaning, depending on how the situation is construed. This is known from figure-ground reversals.[61] An expression imposes a particular image on its domain. Imagery is used as a technical term for the cognitive capacity to construe a cognitive domain in alternate ways.

The cognitive linguist Leonard Talmy introduced the figure-ground asymmetry stating that a physical object is located or moves with respect to another object which serves as a reference point.[62] This asymmetry is embedded in schematization. Schematization is the process involving the profiling of specific aspects of a reference point of a scene representing the whole gestalt.[63] Talmy defines the basic asymmetry in a schematization process as follows:[64]

> The Figure object is a moving or conceptually movable point whose paths or site is conceived as a variable [...]. The Ground object is a reference-point, having a stationary setting within a reference-frame, with respect to which the figure's path or site receives characterization.

Talmy presents a list of various characteristics of the figure-ground asymmetry specifying the relationship, such as the figure being of greater concern or relevance (more salient) as opposed to the ground being of lesser concern or relevance (more backgrounded).[65] This semantic distribution is clearly different from the gestalt notion, which is perceptually based on geometric coordinates instead.[66]

Three basic factors determine the contrast between figure and ground: size, movement, and position of the figure in relation to the ground in the shared knowledge of the discourse participants. Talmy states that, e.g., adpositional phrases profile relationships such as the location of the figure in relation to the ground, the time of the unfolding event, the manner in which the event unfolds, and the transition, motion and path of the figure.[67]

An alternative dichotomy is introduced by Langacker who defines the asymmetry as a trajector (corresponding to the figure) in a relational profile to a landmark (corresponding to the ground).[68] He argues furthermore that[69]

> [w]ith a few if any exceptions, relational predications display an inherent asymmetry in the presentation of their participants. This asymmetry is not reducible

[58] Wender 1980, Zell 1994.
[59] Bechtel and Abrahamsen 1991, Birbaumer and Schmidt 1993, Edelman 2002, Hillert 1987, Hillert 1992, Kandel and Hawkins 1994, Murre and Goebel 1996.
[60] Langacker 1987, 56.
[61] Thiering 2011.
[62] Talmy 1978, 627.
[63] Talmy 2000, Sinha and Kuteva 1995.
[64] Talmy 1978, 627, see also Talmy 2000, 315.
[65] Talmy 2000, 316.
[66] Lewin 1936.
[67] Talmy 2000.
[68] Langacker 1987, 231.
[69] Langacker 1987, 231.

> to semantic roles, i.e. the nature of participants involvement in the profiled relationship. [...] it is observable even for predications that designate symmetrical relationships: X equals Y is not precisely equivalent semantically to Y equals X, nor is X resembles Y equivalent to Y resembles X. [...] In the expression X equals Y [...], X is referred to as a trajector, and Y as a landmark. This terminology reflects the intuitive judgment that Y provides a reference point with respect to which X is evaluated or situated [...].

Clearly, the semantic distinction between the two conceptually based categories reflects the fundamental notion in gestalt psychology of figure and ground.[70] It is believed here, however, that the gestalt psychologist's definition is much more complex and broader than the notions adopted in cognitive semantics. Nevertheless the basic idea of a reference object and an object that needs an anchor is similar. Conceptually, the cognitive semantic notion is very specific in the distribution of meaning components in a sentence. Talmy shows that arguably similar sentences such as (a) 'The bike is near the house' and (b) 'The house is near the bike' are not the same semantically. They present two different (inverse) forms of a symmetric relation.[71] In (a) the house is the reference object, and in (b) it is the bike. This latter profiling seems to be at odds with speakers' expectations. Depending on the real world situation, however, a speaker might refer to the bike as the reference object for various reasons. Zlatev presents a similar example in support of construed situations. In the expressions (a) 'The tree is by the car' and (b) 'The car is by the tree' different situations are encoded. These differences indicate different worlds of human experience, i.e., a non-objectivist approach is favored here.[72] Hence, the semantic function chosen by the speaker does not necessarily correspond to the world of part-whole partitioning, but constitutes language-specific information. This might be due to pragmatics or culture-specific decisions or biases. This example already reveals that language, or rather, speakers choose to reverse natural figure-ground asymmetries. The selected empirical evidence presented in this chapter supports this observation as well.

With this description of some basic theoretical features at hand we shall now consider the two cultures at focus here. The theoretical notions just outlined are important for the analysis of the following language examples.

## 2.3 Anthropological and linguistic background: Dene Chipewyan

This section presents anthropological background information of the Dene culture and linguistic knowledge that speakers of Dene relied on in their daily interaction with the environment.[73] We provide information on the cultural backgrounds as well as language examples of spatial orientation. The Eipo language and culture is then presented in section 2.4.

[70] Koffka 1935, 177–210, Rubin 1921.
[71] Talmy 2000, 314.
[72] Zlatev 2003, 332.
[73] The past tense indicates the drastic change the Dene culture has undergone in the past decades.

### 2.3.1 Contact history and recent acculturation

Dene Chipewyan presents a rather interesting status quo in terms of the actual cultural heritage and the influence of Western culture.[74] Dene Chipewyan belongs to the Northern branch of the Athapaskan language family (spoken primarily in northwestern Canada). The Dene territory extends (or rather, extended) from the southern shore of the Great Slave Lake (Northwest Territories) east to Churchill, Manitoba and south to central Alberta/ Saskatchewan.[75] Perhaps partly due to this geographic isolation similar to Hopi, the Dene dialect of the Cold Lake region is rather conservative with a particularly rich morphology.[76] Only about 2,000 speakers are left in Cold Lake, and only 10% at most speak Dene fluently and on a daily basis.

The Cold Lake First Nations Dene Chipewyan people live near Cold Lake, Alberta, approximately 300 kilometers north-east of Edmonton on the Alberta and Saskatchewan border. Genetically, the Dene language is related to Bearlake, Beaver, Carrier, Chilcotin, Dogrib, Eyak,[77] Hare, Kutchin, Sarsi, Sekani, Slavey, Tahltan, Tsetsaut, Tutchone, and presumably all the languages found to the north-east of these also belong to the Northern Athapaskan phylum.[78] Sapir hypothesized that the Athapaskan language family is part of a larger language phylum which he called Na-Dene.[79] The history of First Nation people in North America was highly influenced by the arrival of the white people. It is fair to state that the initial clash between the native people and white people had a devastating, often lethal effect for most of the aboriginal cultures. European colonialists killed about 50 million indigenous people between 1795 to 1945 worldwide.[80] Bodley also claims rightfully that the colonial encounter was not only a human but also a cultural disaster:[81]

> Colonialism was the first phase of a dramatic world-wide cultural transformation that produced a single global-scale culture based on the commercial market economy.

Nevertheless, the arrival of Europeans in the subarctic region also brought new technology, schools and economic opportunities. The native First Nation of Canada's subarctic region were traditionally caribou hunters. The caribou was the most important source for food, clothing etc. The Dene people followed the caribou migration routes. This is exemplified by the term *edagha* 'a narrow place or area in the lake where the caribous are accustomed to cross and where people sit a little way above (referring to the current) to wait for them'. Moreover, and importantly, following the caribou determined and structured the seasonal cycle and socioterritorial organization.[82] The Dene Chipewyan culture was strongly in-

[74]Eipo, by contrast, has been very isolated until the 1970s, as we shall explain in section 2.4.
[75]Sarsi, Beaver, Slavey, Dogrib and all the languages occurring north-east of these also belong to the Northern Athapaskan phylum.
[76]Malotki 1979, Malotki 1983.
[77]Assumed to be located between Athapaskan and Tlingit; Hoijer 1946, 11.
[78]Hoijer 1946; K. Rice 1989, 11.
[79]Including Tlingit and Haida; Sapir 1915, 12, Hoijer 1946.
[80]Bodley 1999, 465.
[81]Bodley 1999, 465.
[82]Smith 1981, 273.

fluenced by the Canadian Hudson Bay company[83] and the widespread settlement of white people during the Gold Rush years.

Historically, the Dene people lived in family groups on lands encompassing roughly 150.000 square kilometers. They were apparently a mobile people of hunter-gatherers who maintained both summer (*-sine, ziné*) and winter (*háye*) camps, traveling between them on foot or with dog teams. This aspect is important since building a tent (*bét'asi* 'outside of the house, tent') or trap while traveling or following big game (see below) depended on the actual material resources of the particular place.

After the signing of Treaty (or Contract) Six in 1876, many families worked on their reserve farms in summer raising cattle and horses. In winter, they continued to travel north to hunt, trap, and fish. In the early 1950s, the Federal Government turned the traditional Dene Chipewyan territory into an aerial weapons range.[84] It is important to note that the people lost access to their lands and hunting and fishing grounds. Moreover, they were relocated to three small reserves near Cold Lake totaling approximately 18.720 hectares in size (as opposed to 150,000 hectares previously).

Although the Dene people live partly in their original habitat (around Cold Lake), the historical hunting grounds are off limits. The Canadian government bases its largest air military base on the former hunting territory of the Dene. This simply means that Dene people can no longer use their old hunting and spiritual grounds, or family locations of the ancestors. A map measuring 3 × 4 meters at the Cold Lake reserve (band house) actually shows the degree and dimension of the former grounds.

This map indicates every band member, band family etc. and their origin, i.e., it shows that every place or location in Cold Lake once had a human place holder. This topology of names is similar to the topology of names that the Eipo have in their mountainous environment (see below).

Additionally to the military base, the world's second largest oil sands is situated around Cold Lake, meaning that the territory is off limits for the Dene people. Not much is visually left in terms of native traditions in Cold Lake and the village is similar to most other West Canadian villages or small cities, i.e., it is dominated by the fast food stores such as 'Subway', 'McDonald's', grocery stores, and shopping malls etc. typical of North American villages, towns and cities. Hence, Cold Lake is merely a Western Canadian town located in Alberta far away from the next large city (Edmonton) and dominated by Western culture. Dene people speak primarily English and the younger people in particular strive to simply assimilate to the white Canadians in terms of job opportunities or education. The idea of language, and hence cultural preservation, is of lesser importance for the daily life of the Dene.

A general problem with elder speakers of Dene is that some of them simply refuse to speak Dene even though their language is not officially discriminated against today. This is due to the painful past with respect to their treatment in the boarding schools where speaking Dene was prohibited. This led also to complete reluctance to speak Dene at home. The result is that the next generation (aged 45 to 55) were already crucially affected by language attri-

[83] One of the oldest companies in the world, established in 1670; mainly trading fur in British colonies of North-America; see the Hudson Bay Company Archive for further information; http:// www.gov.mb.ca/ chc/ archives/ hbca/.

[84] Named the 'Canadian Forces Cold Lake'.

tion, not to mention the young generation today.[85] As such, Dene presents an interesting, but difficult language and culture where one has to dig deep to obtain an idea of the culture and the practices of the speakers in terms of traditional habits and their history. Some of those traditional habits have survived through oral history. In particular older people remember various hunting techniques or the different functions of traps. On a daily basis this knowledge is not important anymore since their traditional way of life has changed so drastically. It should be pointed out that the future of Dene, or rather, the Cold Lake dialect, seems very bleak. In fact, this chapter is an attempt to glimpse into the intricacies of the interplay of culture, rituals, habits, and language in Dene. It is also an attempt to capture some of the spatial knowledge as long as it is available.

### 2.3.2 Material culture and subsistence techniques

> The aboriginal inhabitants of what is now northeast British Columbia are the inheritors of one of the purest forms of hunting economy; purest in the sense that they are peoples who are flexible in the face of every changing circumstance, to whom material possessions are more of a hindrance than a help, and whose skills and mobility secured a life of relative affluence and good health as long as they could hunt successfully.[86]

The introductory quote indicates the importance of flexibility in the Dene culture in which hunting was the main source of survival. Dene Chipewyan people were mainly Caribou hunters and the most important food animals were the caribou *etthén* of the northern transitional forest and the tundra. Moose and woodland caribou were also important for survival. Generally, caribou were concentrated during their migrations between winter and summer, and in other times scattered at small groups. These behavioral characteristics often determined the manner in which the animals were hunted. The extent to which the migration of the caribou structured the Dene's life is indicated by specific expressions in their language. An example is the classificatory verb stem[87] for the caribou arriving, i.e., *etthén níltah* 'arrive' as opposed to *-tl'ah* which is the verb stem used for caribou only, as in *The caribou arrived*. The semantic difference is in the momentaneous resultative act of arriving as opposed to the telic end result of the arrival indicated by the perfective form. Another specification is the process of the caribou's return as in *etthén nahéltah* 'return' (only used for caribou) *The caribou returned*. It is apparent that knowledge of the caribou's location has been vital for the Dene since the caribou migration structured the Dene people's seasonal distribution, socioterritorial organization, and technology.[88] The caribou are also a key element of religious beliefs and oral literature.

The Dene people used the chute and pound method during the migration phase. A number of people and dogs circularly enclosed an area with a circumference of up to a mile or more containing the caribou herd, using a variety of snares (traps) fastened to poles or tree stumps. The construction of a snare or a deadfall is a highly sophisticated technology. However, it does not require a sound understanding of fundamental principles of physics, but

[85] Thiering 2009b.
[86] Brody 1982, 85.
[87] For the specific linguistic terminology, see below.
[88] Smith 1981, 273.

rather the behavioral characteristics of the particular species. Indeed, it is practical knowledge transmitted from one generation to the next that enables such techniques. Their material components are largely comprised of materials which can be found scattered across the boreal forest landscape. Dene deadfalls were used mainly for *tha* 'marten', *thachogh* 'fisher', *thelchuzi* 'mink', *nágídhi* 'fox', *sas* 'bear' (*dlézí* 'grizzly bear', *sas delgai* 'polar bear', *sas delzeni* 'black bear'), *nábie* 'otter', *dzen* 'muskrat', *tsá* 'beaver', and *nághai* 'wolverine'. Snares were set chiefly for grouse, hare, fox, bear, caribou, and moose. Hence, different techniques were required for different animals. Since caribou were the most important animal, the methods of hunting them will be specified as an example.

Once a caribou herd was detected the caribou were manoeuvered into the mouth of a prepared chute and driven to the pound. Once inside the pound the caribou were entangled by snares or traps. In addition, single caribou were hunted with spears or shot with arrows. Knowing the caribou tracks, another option was simply spearing them while they crossed the rivers and lakes. Hence, it was important to know the specific water conditions or the respective river as linguistically represented in expressions such as *des dánét?á* 'the river is full' or *des héli nálthah* 'the river is flowing fast'. Both expressions were important for fishing and for locating caribou. Hunting techniques were adapted with respect to the behavioral characteristics of the animals. Big game use rivers or lakes for their water supply. Of course, since the arrival of white men, rifles were used more frequently. Unlike caribou, moose do not gather in larger herds, but tend to live in isolation. After eating the moose turns back on its trail to the windward to rest. Hunters adapted to this habit. They followed the trail to one side and windward, checking every once in a while whether the animal had returned. When this was the case, the hunter knew the moose's exact location. Beside caribou and moose, bears were also hunted, but only occasionally. Beaver, on the other hand, were an important food source. Usually they were caught during winter when their homes could easily be located. The ice conditions limited the beaver's movements. The idea of catching beaver was simple: it was sufficient to block the entrance and then break into their lodge. A variety of traps were used such as tossing-pole, springpole, stationary snare, deadfalls of various sizes and trigger mechanisms, bows and arrows.[89] Snares were used to catch hare. Only after European contact began were small mammals hunted or trapped solely for their fur.

The dog was the only domesticated animal used for hunting moose, bear, beaver, and geese. Fishing was an important food source only for some clans. In general, big game like caribou was sufficient. Seasonal climatic conditions in conjunction with the behavioral characteristics of the fish indicated the appropriate seasons of exploitation and the techniques to be employed to hunt them. Trout were taken by hook in open water or through ice holes in late winter. Fish spears were also used. Fishnets were usually made of willow or babiche in prehistoric times, while industrially produced twines and nets were introduced after European contact.

With the approach of fall, people left the summer gathering centers to seek food in preparation for the long and rather cold winter. People carried little with them, because many things could be made relatively quickly with local materials at hand. Although the land required unique skills to survive, these skills did not require a highly specialized manufacturing technology in order to act within the environment (the exception was making

[89] Bows were made of birch; strings were of twisted babiche, rawhide, or sinew. Arrows were made of straight-grained spruce or birch.

traps). This is not to say that indigenous technology was not sophisticated – quite the opposite, it was extremely complex, but its production did not require specialized labour. Most people could make most things used in the society. Indigenous people of the North accommodated to the sense of balanced needs with respect to what was available to them locally within their environment. They did not need many things in order to make a living. Their inventory of plants used for food and other material purposes was extensive.

### 2.3.3 Social structures

Regional bands ranged in size from about 200 to 300 people. Local bands varied from 30 to 100 people and their movements were again based on the migration of the herds. Shift of families was common and hence the bands became amalgamated and heterogenous. It can be assumed also that dialects changed or intermingled.[90] Most families were related to each other. Band membership was known to be fluid, i.e., bilateral kinship and marriage provided avenues for new affiliations.[91] Due to European-introduced diseases, substantial social realignments occurred. Smallpox, tuberculosis and influenza affected the Dene people in the 1920s.[92]

After 1945, most children were sent to Catholic residential schools off the reserve to receive a Euro-Canadian education. The entire community was adversely affected by the almost total separation of the family unit, which persisted except for the few weeks each year when children returned to their families. Elders and children lost the ability to communicate with one another. These schools had an especially devastating effect on the Dene language[93] and way of life, not only because children were discouraged from or actively punished for speaking their native tongue in these schools, but also normal linguistic and cultural transmission between the generations was vastly disrupted.

This is quite different from the Eipo situation, as will be outlined below. In Eipo, strong family and community bonds have been maintained and hence a detailed topography of their environment is still known. Parallel to the linguistic loss in Dene went the loss of songs, games, rituals, stories, techniques, e.g., practical knowledge of how to build the highly complicated traps, and ceremonies. All in all this implies an almost complete loss of community life and culture. The last 50 years have seen a steady decline in the numbers of Dene Chipewyan at Cold Lake able to fully communicate in their heritage language.[94]

A 1998 survey carried out in accordance with the Department of Indian and Northern Affairs Registration System identified 285 persons.[95] At present the number is down to about 200 speakers; fluent or conversant speakers of Dene Chipewyan out of an official band membership of 1,908. Thus, only about 10% of all band members speak an Aboriginal language to some degree of competency. The 1960s must have been traumatic for the Caribou Eater Chipewyan people since their contact-traditional way of life changed drastically and suddenly. The five bands, which were named after geographic areas, were relocated,

[90] As in Eipo; see below.
[91] Smith 1981, 276.
[92] Smith 1981, 274.
[93] Thiering 2009b; Thiering 2010
[94] Thiering 2009a.
[95] http://jan.ucc.nau.edu/~jar/ILAC/ILAC_10.pdf; accessed 25 February 2014. For an earlier census, see Smith's table of Chipewyan population in 1970: Smith 1981, 75.

e.g., to a subarctic town notorious as one of Canada's worst slums.[96] The result of this relocation had a devastating effect on the people and left them disoriented and demoralized.[97] The imposed village life profoundly changed the traditional living habits of the hunter-and-gatherer culture. Men were supposed to leave families behind while hunting, i.e., the former division of labor was disrupted. The distance from the village to the hunting grounds made it difficult to kill a large number of animals simply because only a limited amount of meat could be transported by a dog team.[98]

### 2.3.4 Traditional religion

Myths about places, rituals and used objects, powers, spiritual and medical knowledge, stories, dances and music were religious. Hunting and gathering were the most important activities for survival, and spirituality was linked to finding food and was important for survival in the harsh climatic conditions. Hence, spirits were thanked for when finding food. If no food was found the Dene people tried to appease the spirits with offerings. One important spiritual figure was the *Kakhani*, a supernatural being, half-man and half-monster. It was believed to steal children. Unlike the Eipo, who did not decorate most of their tools, not even objects like the holy digging-stick (see below), Dene people decorated their snowshoes with paint, strings of shells, and amulets woven into the snowshoe to keep the wearer safe from unfriendly spirits.

### 2.3.5 Physical environment

The environment of the Dene Chipewyan people is made up of tundra, forest (black spruce, white spruce, birch, aspen, also known as the 'land of the little sticks'), and boreal forest. The seasons are basically bicyclic: long and severe winters, short and moderately warm summers. The severe winters limited activities and required maximal effort for survival. Variation in snow conditions affected the behavior of the fauna (providing food and clothing) and hence affected native techniques for its exploitation. During summer, traveling was on foot, following water courses or by canoe on open water. Around late autumn (September/ October) water began to freeze, which limited traveling. In winter, dog sleds and snowshoes were used. Game animals provided most of the raw materials, e.g., bones, antlers, hide (skin) to produce beamers, needles, spear, arrowheads, fishhooks, bowstrings, fishing lines, bags, lodge coverings. The forest (forest-tundra) provided most of the remaining raw materials for bows, arrows and spear shafts, containers, dishes, net gauges, snowshoe, and canoe frames, snow shovels, toboggans, bark for making dishes, boxes, and coverings for lodges and canoes.

Generally, the climate was a dominant and active element in the subarctic environment. This region belongs to the cold snow forest category, a circumstance which profoundly affects the life circle of the Dene people. Rivers and lakes played an important role in transportation and communication. The drainage grids and water surfaces were important movement and communication routes and therefore attracted settlement and other activities

[96] The five regional bands are: Duck Lake/Churchill band ('east people'), Barren Lands band ('flat-area-dwelling people'), Hatchet Lake band ('hatchet-lake people'), Black Lake band ('upland or western people'), and Fond du Lac band ('pine-house people').
[97] Smith 1981, 282.
[98] Smith 1981, 282.

during both winter and summer. In addition, knowing the game routes, e.g., along rivers, helped in finding enough food for the band. The richness of fish, lumber, and wood pulp attracted white enterprises, particularly the Hudson Bay Company. This, of course, changed the life habits of the Dene people as well.

### 2.3.6 Relationships to neighboring groups

The only known enemies were the Cree to the south and the Inuit to the north. The landscape features forming the borders were not crossed by the Dene except for warfare. Regarding contact to the Europeans, at the beginning the marginal location to the transportation and trade routes, the dependence on caribou, and the low interest in European trade goods led to a rather slow and limited sociocultural change.[99] Rapid changes only began in the 1960s. Hence, no relationships with Europeans were established until the 1960s.

### 2.3.7 Linguistic overview

It should be noted that for reasons of history and migration, the Dene band is the most southerly of all Dene Chipewyan-speaking communities in Canada and is geographically isolated from other Dene Chipewyan speech communities. Consequently, the dialect spoken at Cold Lake is particularly conservative and rich in phonological and lexical contrasts that have been lost in more northern dialects. Indeed, many Cold Lake Dene speakers regard their dialect with pride as the purest form of Dene Chipewyan (whatever is left of their language).

Dene features a polysynthetic linguistic system, i.e., bound morphemes constitute complex words or even sentences and the syntactic object of the sentence is incorporated into what may be termed the verb cohort. The general encoding pattern in Dene indicates that the language features a predominant and consistent classificatory verb system including directional prefixes as well as a postpositional inventory creating relational predication cohorts or constructions.[100] Such verbs have different morphological forms depending on the object to be encoded. Cook argues that Dene has about 36 postpositions that morphologically behave like nouns. They inflect with pronominal prefixes.[101] Cook also highlights the fact that the determination of a postposition's meaning is as notoriously difficult in Dene as in English or any other language, making it often impossible to determine the precise meaning out of context. However, these postpositional prefixes are widely acknowledged as modifying the meaning of the verb stem.[102] Their stems change depending on the shape, animacy, and/or physical features of the object being located or handled.[103]

The general focus of this chapter is on the formation of certain semantic construction types and the encoding of the figure-ground asymmetry as modified by the linguistic construction. The language features a predominant classificatory verb system, as do all of the other languages of the same phylum.

> All the Athapaskan languages exhibit an alternation of verb themes that is traditionally called classificatory. The classificatory themes describe the nature of

[99] Smith 1981, 282.
[100] See Li 1946, Kari 1979, Cook 2004b, K. Rice 1989, McDonough 2000, S. Rice 2002 on the general structure of the Athapaskan verb stem system.
[101] Cook 2004a, 92.
[102] S. Rice 1996.
[103] S. Rice 2002, 69.

> an object handled with respect to parameters such as extension and dimension. The verb theme indicates the nature of the object handled, while the type of activity involved is expressed in the prefixes.[104]
>
> The choice of a particular verb stem from the appropriate set of verb stems has the effect of assigning to the noun of the sentence certain qualities of number, shape, texture, or purpose. If these qualities are semantically inappropriate to the noun, another verb stem must be used.[105]

These stems profile existential situations or actions of certain categories of objects.[106] Table 2.1 summarizes the four main classificatory verb types used in Dene.[107]

| | |
|---|---|
| Posture or locative verbs | no movement involved: e.g., 'sit', 'stand', 'lie', 'be in position/location' |
| Verbs of handling, manipulation, continuing manual contact | e.g., 'give', 'hand', 'take', 'put', 'handle', 'bring', 'carry' |
| Verbs of partially controlled action (+ agent) | e.g., 'toss', 'throw', 'hang up', 'set down', 'drop', 'lose', 'push over' |
| Verbs of free movement, independent of agent | e.g., 'fall/tip over' |

Table 2.1: The different classificatory verb types

According to traditional accounts, the Dene verb consists of a verb theme (the basic lexical entry made up of a stem and one or more thematic prefixes; a unit including a verb base plus other morphemes combining to a specific meaning construction); and additional prefixes.[108] The Dene verb construction can be described as a composite construction similar to Navajo.[109] It is claimed here that the Dene verb system is compiled via a string of distinctive elements fused or agglutinated together to form a lexical unit or word, or a sentence. The verb stem is the basic entry or atom derived from a verbal root. The theme profiles the verb base (classifier plus stem construction), i.e., a skeleton of a meaningful lexical unit.[110] The verb stem is assumed to be the content part of the verb, and contains rich semantic information.

The Dene verb shows polysynthetic and fusional characteristics in its morphology and with its rich prefix system.[111] Subject and object prefixes are fused within the verb.[112] These prefixes encode also five modes, and three aspectual forms, person, and number.[113]

---

[104] K. Rice 1989, 779. The concept of 'verb theme' is explained below.
[105] Carter 1976, 24.
[106] Davidson, Elford, and Hoijer 1963; see Senft 2000 on a collection of papers on classification.
[107] Davidson, Elford, and Hoijer 1963, K. Rice 1989, S. Rice 1997, S. Rice 2002, Cook 2004a.
[108] Li 1946, Hoijer 1951, Young and Morgan 1987, K. Rice 1989.
[109] Young and Morgan 1987, Young and Morgan 1992.
[110] Young and Morgan 1987, 99.
[111] Buschmann 1855, Morice 1890, Li 1946, Boas 1977.
[112] S. Rice 2002, 66 ff. Cook 2004a.
[113] The five modes are: the neuter, momentaneous, continuative, customary, and the progressive mode; the three aspectual forms are: the imperfective, perfective, and future aspect; see Li 1946, 404, 409.

The neuter verb refers to the state or the position of the figure. The momentaneous profiles a rapid action or transition from one state to another as in ‘to sit down’, ‘to handle a round solid object’ or ‘to lie down’. The continuative verb profiles an activity that lasts in time such as ‘to stay’ or ‘to own’. The customary verb encodes a repeated action and the progressive encodes an ongoing action.[114] Themes occur as free and bound lexical units. Free themes profile nouns and modifiers, bound themes are verbs and pronouns.[115]

To show the verb stem changes according to the figure to be encoded, an example of stem variation is given in table 2.2. It is evident that different objects to be handed over or handled affect and change the verb stem, i.e., the morphology.[116]

| | |
|---|---|
| *be(3SG.)-gha(to)-n(MOM)-i(1SG.S)-l(CLASS)-ti(STEM)* | ‘I gave animate being to him/her.’ |
| *be-gha-n-i- ?a* | ‘I gave round/hard object to him/her.’ |
| *be-gha-n-i- ta* | ‘I gave sticklike object to him/her.’ |
| *be-gha-n-i-l-chudh* | ‘I gave flat object to him/her.’ |
| *be-gha-n-i-la* | ‘I gave plural objects to him/her.’ |
| *be-gha-n-i-ka* | ‘I gave open container to him/her.’ |
| *be-gha-n-i-chu* | ‘I gave unspecified object to him/her.’ |

Table 2.2: Variations on the theme ‘I transferred **X** to him/her’

The Dene verb stem changes according to the quality of the figure, i.e., differences in shape, size and animacy of the objects to be encoded determine the choice of a verb’s stem.

In the literature on Athapaskan languages it is common to use rather idealized templates as presented above. The number of prefixes varies significantly, e.g., Athna has 23 prefix positions,[117] Slavey 14,[118] and Navajo 10.[119] McDonough divides the verbal complex into a bipartite structure: Positions 1 to 4 are the satellites, and positions 5 to 10 are defined as the pre-stem position.[120] The positions (1 to 4) (= disjunctive prefixes) and (5 and 6) (= pronominal subjects/ objects) are part of the disjunct or lexical zone and largely have a derivational function, while positions (7 to 10) are called conjunct or grammatical zone and include obligatory inflectional categories such as tense, aspect, modality, subject agreement, or valency.[121] Valency classifiers in position 10 indicate the transitivity and voice of the verb, i.e., whether the subject takes a direct object or not. With regard to the data description, the stem plus the positions 8 to 10 as well as 1 are of primary importance.

We have seen some important aspects of the Dene culture and language. The next section presents some background on the anthropological and linguistic aspects in Eipo.

---

[114] Li 1946, 405.
[115] Hoijer 1946, 297.
[116] For the linguistic abbreviations, please consult table 2.13, preceding the Bibliography. Here and in the following, the question mark (*?*) denotes a glottal stop sound in Dene.
[117] Kari 1979.
[118] K. Rice 1989.
[119] Young and Morgan 1987, Young and Morgan 1992.
[120] McDonough 2000.
[121] Li 1946, 409.

## 2.4 Anthropological and linguistic background: Eipo

The Eipo language and culture are members of the Mek group of Trans-New-Guinea-Highland Papuan languages and cultures.[122] The Eipo live at the northern slope of the central cordillera in the valley of the Eipomek River in the central Mek region. (*Mek* is the term for water and river in the Eipo dialect of the Mek languages and was therefore chosen as denominator for this ethnolinguistic group,[123] other dialects use *mak* or *me*.[124]) The Eipo territory is located approximately at a latitude of 140 degrees east and a latitude of 27 degrees south in what is now called *Kabupaten Pegunungan Bintang*, the 'Star Mountains District' of the Indonesian Province of Papua (formerly Irian Jaya). Thus, Eipo belongs to an estimated number of 760 Papuan languages of about 4 to 5 million speakers divided up into sixty language families.[125] Foley presents a comprehensive overview of the Papuan phylum, its location and its historical background.[126] An important aspect, as Foley points out, is that according to his analysis, Papuan languages are not genetically related, i.e., they do not trace their origin back to a single ancestral language.[127]

Quite unlike the Dene, the Eipo preserved most aspects of their way of living until the mid-1970s, when two major earthquakes hit their region and they began to convert to Christianity. The typical Eipo community consisted of hamlets of 35 to 200 people that are settled at around 1,300 to 2,000 m above sea level, but the Eipo hunting area extends up to 4,000 m above sea level. These numbers are compatible with Foley's account according to which New Guinea societies are based on hamlets between 100 and 300 people.[128] His explanation for the small size is that ecological conditions, especially the difficult terrain, prevent people from moving across barriers (see below).

However, Eipo women and men, also children, cross the high mountains frequently and a number of men report having even climbed from their village at 1,700 m to the pass at 3,700 m, i.e. 2,000 m altitude, in darkness. These extraordinary feats usually happened in clear nights with a good moon, but are still a most remarkable performance given that the path is often hardly visible even in bright daylight and that a wrong step could cause death on many of the perilous tracks to be negotiated. These reports and Schiefenhövel's personal experiences of walking long distances at high altitude with Eipo friends demonstrate that they, like other highland Papuans, are adapted to their environment with a perfection foreigners can hardly fathom.

The Mek share some cultural features with their neighbors in the east and in the west.[129] The term *mek*, as mentioned above, stands for 'water', 'river', 'brook', also for 'sweat' and other semantic units, generally for watery liquids (3894).[130]

---

122 Wurm 1982.
123 Schiefenhövel 1976, Schiefenhövel 1979, Heeschen and Schiefenhövel 1983, Heeschen 1990, Eibl-Eibesfeldt, Schiefenhövel, and Heeschen 1991, Schiefenhövel 1991, Heeschen 1998.
124 See also Louwerse 1978 and Louwerse 1988.
125 Wurm 1982, Foley 1986, Bußmann 2008.
126 Wurm 1982, Foley 1986.
127 Foley 1986, 3; but see Heeschen 1992 who argues for the genetic relatedness of all Highland Papuan languages.
128 Foley 1986, 14.
129 The Mountain Ok in the east (cf. Pouwer 1964) and the Yali, a subgroup of the Dani, in the west (Koch 1984).
130 Arabic numbers in parenthesis refer to the entry in the unpublished File Maker corpus of Eipo held at the Max Planck Institute for the History of Science. It is based on the dictionary of Eipo which not only contains words and their translations into German and English, but also features quotes of actually spoken phrases, sections of legends, songs etc. (Heeschen and Schiefenhövel 1983). Those entries exemplify the Eipo terms, with the result that the

*Mek* was an obvious local word to be used as ethnonym to designate the cultures and languages in the Mek area. The relationships between the groups in this region and their linguistic and cultural unity were unknown to the local people until 1975.[131] The Eipo River or Eipomek is the main river of the area where Eipo was spoken by approximately 800 people at the beginning of fieldwork in 1974. The total number of Mek speakers north and south the central range may have been around 15,000 at that time. The number of speakers had risen to at least double this figure in 2009.

Other dialects in the Eipo area were spoken by an additional number of around 700 persons, so that, at the beginning of research in 1974, about 1,500 speakers of Eipo and related dialects lived in the area. As noted above, the villages had between 35 and 200 inhabitants. This figure has also risen greatly due to the dramatic population growth typical of the highlands as well as the other regions in Papua Province and, on the other side of the border, in Papua New Guinea, where the annual percent population surplus is estimated at 1.89% for 2013,[132] other estimates derived from studies in the first years of wide-ranging acculturation place this figure between 2.1%–2.6%.[133] In the past, village communities and political alliances were rather small, following a pattern which was found in many New Guinea Highland Societies, except where wide valleys had brought about a different settlement pattern, e.g., the Balim Valley in the Province of Papua and the Whagi Valley of Papua New Guinea, where much larger populations lived.

The phrases in table 2.3 present the importance of the rivers and similar features *(mek)* as landmarks and origin of mental concepts and metaphors in the Eipo language. (Numerous other semantic usages of *mek*, that do not refer to spatial deixis, have been left out.)

Eipo speakers base their directional system on the river stream system.[134] The spatial terms *ou* 'down the river', *or* 'across the river on same level or below one's own position', *ei* 'up the river', *er* 'across the river above one's own position', and others are river based. Also, as indicated in the list above, many metaphors use river and water as *tertium comperationis*, as in *mek-arye* 'steam' and *mek kate* 'ice'. In addition, some shape forms are based on the morphem *mek*, e.g., the bowl-shaped form that results from water washing out a certain spot, or a cavity made by water *(mek loktena)*.

With respect to natural boundaries it has to be mentioned that it is difficult but usually possible to find ways through the rainforest adjacent to the inhabited areas like those in the Mek region, as well in the montane and alpine regions of New Guinea. The swampland present in some lower altitudes poses greater problems for human mobility and has probably contributed to the very marked cultural and linguistic diversity for which New Guinea is known. As Foley states, the terrain thus poses some genuine barriers to human social interactions and would certainly favor linguistic diversity.[135] It seems likely that the extraordinary variety of languages and cultures in this part of the world is also the product of an aggressive (warrior-like) attitude of one group toward another, even inhabitants of one valley toward the neighboring one. Intergroup warfare increases intragroup cohesion and

---

monograph is more an ethnographic wordbook than a mere dictionary. These entries were transformed into the above-mentioned electronic data file. Additionally, examples of Heeschen's substantial *Ethnographic Grammar of the Eipo Language* (Heeschen 1998) and field notes of Wulf Schiefenhövel are used in this chapter.

131 Schiefenhövel 1976.

132 http://www.indexmundi.com/papua_new_guinea/population_growth_rate.html, accessed 6 December 2013.

133 King and Bathgate s.d.

134 See Brown 1983.

135 Foley 1986, 9.

is very likely to have led, in a process of character enhancement, to the very fragmented cultural and linguistic scene typical for mainland and island New Guinea.[136]

| | |
|---|---|
| *mek burwe* | 'head water region' |
| *mek youkwetam* | 'downstream' (3894/31), 'toward the foothills', 'north' |
| *mek bongbong* | '(narrow) valley' |
| *mek arum* | 'water surface' (191/1) |
| *mek lu* | 'water surface' (3623/2) (*lu* = 'even', 'flat', 'down', 'low') |
| *mek amwe* | 'bed/bottom of a river, a lake' |
| *meke ebrarik* | 'water', 'rivers split up/join', 'river junction' |
| *mek bene* | 'stagnant water', 'swamp' |
| *sisilya arang mek* | 'reddish brown water (e.g. coming from swamps)' |
| *mek kwen* | 'lake', 'pond' |
| *mek bun* | 'bridge' (936) |
| *mek dala* | 'river bank' (3894) |
| *mek denemna* | 'border of a brook' |
| *mek duman* | 'the river shore, along the river': cf. *Eipodumanang* 'we are the ones who live at the shore of the Eipo River (the Eipo)' |
| *mek irikna* | 'river bank' or 'edge of a river' (2220/1) |
| *mek deya* | 'hollowed out river bank' (3894/6) |
| *mek dorobna* | 'small spring' |
| *mek lum* | 'waterfall', lit.: 'water veil' (3894/8) |
| *mek ib* | 'to dam a water' (3894/10) |
| *mek kate* | 'ice', lit.: 'hard water' (2427/9) |
| *mek loktena* | 'hollow/cavity made by the water' (3575) |
| *mek-arye* | 'that which is caused by water', 'steam' |
| *mek burbur anmal* | 'the river swells up' |
| *moke wik meke bo'lunmak* | 'when there is a lot of rain the rivers swell up' |
| *wakna mek* | 'actual course of the water' (3446/2) (as opposed to *wakal kwoten mek* 'old river bed' (5439)) |
| *mekin bal* | '(mythological) snake (which created the land by damming and derouting the water' |
| *basam mek* | 'water from sacred ponds which pigs should drink to grow faster' |
| *beta mekduman mereklamuk* | '(the ancestor) walked the whole way along the river' |
| *mek aleng* | 'the stringbag which people put over their eyes when they commit suicide by jumping into the river' |

Table 2.3: Semantic variation of 'river' in Eipo

[136] Schiefenhövel 2001.

The data suggest that the process of pseudospeciation so typical for New Guinea with its many hundred ethnolinguistic groups set in motion not only by the long history of settlement and the rugged nature of the terrain, but also by the above-mentioned high level of aggression between the groups, thus by a biopsychological factor. Linguistic markers of ethnic identity and the dynamism of languages developing away from a common origin play, of course, an important role in this process as well. Foley's hypothesis may be true for the inundated or swampy sections of the lowlands, but one can safely say that neither very high mountain ranges of close to 4,000 m altitude nor large rivers (like the Idenburg-Mamberamo system north of the Mek area) have kept people from moving across those 'borders'. This is in contrast to what Europeans would assume in view of these formidable barriers.

Our species is an extremely mobile one, as proven by the fact that the ancestors of today's Papuans, after crossing the open ocean at the Wallace line between Bali and Lombok, arrived at the New Guinea coast some 50 to 60,000 years ago[137] and settled throughout the interior. Much later, Papuans, probably initially on the islands and coasts of the Bird's Head area in the westernmost part of New Guinea, mixed with people arriving from Southern China and/or Taiwan (the Protoaustronesians). That Austronesian seafarers made their homes on almost all the islands in Melanesia, Micronesia and in the vast Polynesian Pacific long before James Cook arrived is a truly extraordinary feat of spatial orientation and human expansion across the inhospitable vastness of the Pacific Ocean.

### 2.4.1 Contact history and recent acculturation

The Eipo were first contacted by members of the heroic crossing of West New Guinea, from the south to the north coast, by members of the expedition of Pierre Gaisseau (1961) in 1959, and in 1969 by a group of Indonesian military personnel including Gaisseau, who parachuted into the southern Eipomek Valley,[138] and stayed some weeks in this and the adjacent area in the east. They produced a small amount of good ethnographic and linguistic data and are still remembered by the local people. In the early 1970s a few missionaries of the Unevangelized Fields Mission (UFM) walked through the Tanime, Eipomek and Nalcemak Valleys to check possibilities of building mission stations.

When fieldwork of the interdisciplinary German research team[139] began in 1974 the Eipomek Valley did not have an airfield and a mission station. At that time, the Eipo therefore lived in marked isolation. Moreover, very few metal tools (bushknives, axes) and a few new plants (e.g. *Zea mays, Sechium edule*) had found their way into this area. Schiefenhövel's fieldwork[140] was mainly carried out in the village of Munggona, the cultural and religious center of the southern Eipomek Valley, but also included the neighboring valleys east and west, the Heime Valley south of the central range and regions at the northern fringe of the Mek culture near the Idenburgh River as well as the In Valley around Kosarek (where the westernmost Mek speakers live), and the until then uncontacted area inhabited by the Lauenang north of Kosarek.

In 1979 the inhabitants of the Eipomek Valley accepted Christianity. It is important to note that this acceptance was basically a political, not a religious decision. The Eipo had

[137] Swadling 1981.
[138] Komando Daerah Militer XVII "Cenderawasih" 1969.
[139] Funded by Deutsche Forschungsgemeinschaft.
[140] First period from 1974 to 1976; follow-up visits in 1979, 1980, 2008, 2009, 2010, 2012, 2013, 2014, 2015 and 2016.

realized that they had lived separated from the rest of the world with its astonishing superiority in material goods and technologies and wanted to become part of this world. As in other regions of Melanesia the new religion was seen to hold the promise to connect them to the hitherto almost completely unknown way of life. Until 2016 the strategy to accept Christianity as an avenue to the modern world has worked out well for them. Many Eipo go to school and are doing very well, and some of the young people are students of Cenderawasih University in the provincial capital of Jayapura or in other academic institutions of the Indonesian Republic, even in the capital Jakarta. These remarkable changes were all achieved within one generation. This radical change had, and still has, repercussions on the Eipo culture and language. Movements for religious revival, including the classic cargo-cult type millenarian prophecies, have not affected the Eipo yet. They have, indeed, so far opposed such utopian ideas. It seems they have understood that the only way to move forward and to secure their survival as a cultural and political group is to become as well educated as possible.

Many elements of their traditional lives have changed, but others have remained much the same as in 1974, partly because there is no road for any type of vehicle connecting their region with any of the centers of the province. Walking and the airplane will be the only means of transport for a long time to come.

One of the most dramatic changes in the political field concerns the fact that the Eipo and their neighbors have understood that they form a larger single ethnic group with the same Mek language and very similar cultural traditions and that they should cooperate in the arena of provincial politics. They have thus developed a new spatial-political concept, which is paralleled by their new, much widened horizon: quite a few of them travel by plane to Jayapura, the provincial capital on the north coast (about 200 km in a straight line or one and a half hours' flight time), and other cities, e.g. Wamena, the main hub of the highlands of Papua Province, and some Eipo have visited Germany and other European countries. Walking beyond the formerly rather confined borders of areas where relatives lived is also common now. Quite a number of Eipo, including middle-aged persons, walk to Oksibil, the government center in the east of the Mek region not far from the border with Papua New Guinea, and live there for a while, despite the fact that people in this region speak the Ok language which they do not understand. The lingua franca is Bahasa Indonesia which many Eipo speak quite fluently.[141] Most administrational posts are filled by persons of Papuan origin, including the governor of the province and the rector of the University in Jayapura-Abepura. Eipomek, the name of the airfield and the administrative seat of the upper Eipomek Valley, has a number of public service offices, but no one is working there yet.

### 2.4.2 Material culture and subsistence techniques

Traditional tools were the *ya* 'stone adze', *kape* 'stone knife', *fa* 'bamboo knife', *kama* 'wooden digging stick', *yin* 'large bow', *mal* 'arrow', *aleng* 'string bags' (of various sizes), *towar* 'ratan liana' for binding and fire-sawing and some other, smaller tools plus a range of body decorations.[142] Subsistence techniques were a mix between horticulture, hunting and gathering. Highland New Guinea is the homeland of some important domesticated food plants and thereby one of the very few centers of early agriculture worldwide. Some of the

[141] *Ok* is the term for water and river in this part of the New Guinea highlands.

[142] For a complete inventory of their material culture, see Koch 1984.

main plants are the *am* 'taro' (*Colocasia esculenta*), *kuye* 'sugar cane' (*Saccharum officinarum*), *bace* a related plant eaten as a vegetable (*Saccharum edule; pitpit* in Neomelanesian Pidgin), some protein-rich leafy greens *(mula, Rungia klossii; towa, Abelmoschus manihot)* and probably also *kwalye* 'banana' (*Musa paradisiaca*) belong to these autochthonous foods. Various cultivars of sweet potato (*kwaning*, *Ipomoea batatas*), the arrival date of which (either after the conquista or through early Polynesian transpacific contacts) in New Guinea is still debated, provide the bulk of carbohydrate energy and thus comprise the staple diet. Hunting[143] is not very efficient, as the local species of marsupials[144] are small, yet it played an important role in providing essential amino acids and was held in high esteem by the men. Hunted game is still ritually important (to host special groups of guests, as part of the bride-price etc.). *Basam* 'pig' (*Sus scrofa*) and *kam* 'dog' (*Canis familiaris*) are placental, i.e. non-marsupial animals, possibly introduced by the Austronesians, and thus foreign to the ex-Sahul fauna typical for New Guinea and Australia with kangaroos, wallabies and the like. Dogs are not eaten by the Eipo, whereas the pig was, and still is, a very important source of protein and fat. As pigs are not able to find enough food themselves they are fed, usually sweet potato, and thus represent a luxury food reserved for special occasions. They continue to be very important for ceremonial exchange as well.

Horticulture provides the staple foods of the Eipo. Gardens *(wa)* were usually made in areas which had been cultivated before and allowed to lie fallow for approximately 15 years. This period was determined via a bioindicator: the growth of the *urye*-tree (*Trema tomentosa*). When it had reached a certain height and diameter the soil was seen to have recovered and to be ready for a new round of planting and harvesting. Fallow periods have been shortened for several years now due to the marked population increase and the need for more food. The garden land is owned by patrilineal families. Some clans, those said to have come later in the history of settlement, do not formally own land in the Eipomek Valley but are given plots to grow their food. In this way, there was, in normal situations, neither shortage of suitable land nor of garden produce. Everyone who was physically able to work in the garden could and still can do so and was and is able to provide food for him/herself and the family.

Garden land is sacrosanct. The individual plots are clearly identifiable: at the corners or other crucial spots of the garden's border the sacred *yurye* (*Cordyline terminalis*) is planted. This is a small tree with often reddish, lancet-shaped leaves, of which several cultivars are known. It is also planted at other crucial places, e.g. near the sacred men's house, at meeting places or at the head of the long cane bridges spanning roaring rivers. Interestingly, this particular plant signifies places of religious importance throughout the Pacific, e.g. the entrance of temples in Bali and holy sites in Polynesia.[145] The visual line connecting the *yurye* is the border (*wa wiliba,* literally: 'the garden work-stopper') in Eipo gardens. Failure to respect this border by clandestinely or openly transgressing and planting or harvesting in the land of one's neighbor leads to serious conflict: verbal aggression and, possibly, physical fights. Everyone knows this law and usually respects it. There is, thus, family-owned, not communally owned garden land. The geometry of the gardens, their general shape, slope, geological condition and suitability for particular crops is common knowledge, as is

[143] With bow and arrow, often assisted by specially trained dogs or with snares and traps.
[144] Mice, rats, opossum-type animals of the *Phalangeridae* family.
[145] At this point it remains an open question whether the surprisingly wide distribution of this plant as a religious symbol is pure coincidence or the effect of cultural exchange.

the closer and wider area around the village which is represented by a complex network of place names.

When one walks on a path leading away from the village toward the periphery one crosses from zone to zone, all with defined borders, specific place names with their specific history of what happened there in mythical, remembered, and recent times. Known space is, thereby, meaningful territory, a carpet of culturally encoded signals, enriched with one's own experience, with emotionally and cognitively relevant contexts. Arguably, this might have been similar in any rural environment and in daily contact with its spatial and other features. It is at least similar to the Dene Chipewyan tradition.

### 2.4.3 Social structures

Patrilinear descent and virilocal residence, i.e., the wife moving to the husband's village, are still in place. The marked division of the society into female and male spheres (with men's houses and women's houses, both religiously meaningful, and other cultural institutions), which was present in the Eipo culture, as in that of other Papuan groups in the New Guinea highlands, has been reduced in recent years. Similar to other Papuan societies the leading roles in the public arena were, in the past, taken by the big men (*sisinang*, literally: 'the ones who speak'). They got these positions through a mix of personal characteristics, among which intelligence, vitality, rhetoric and social skills were the most important. This meritocratic system without heritable chieftainship controlled all public affairs, including the decision whether to wage war or make peace with the main enemy in the adjacent western Famek Valley. Today, new leading positions have become available, among them those of church leaders and teachers; incipient forms of election are becoming institutionalized. Clan exogamy was, and still is, the guiding principle for marriage. In the past, 12% of all men were, at one time in their lives, married to more than one, usually two, exceptionally three wives; this optional polygyny was largely abandoned with the acceptance of Christianity. Divorce was common; the woman usually took the separated couple's younger children with her, went back to her own family and usually remarried quickly.

### 2.4.4 Traditional religion

This section provides some ideas on the former animistic religion of the Eipo. Like that of the other highland New Guinean religions or, in fact, Melanesian religions in general, it was based on the belief that the visible and invisible world is filled with beings, i.e., *isa* 'spirits' of various kinds similar to the Dene Chipewyan tradition. Most important were creator spirits, e.g., the *Yaleenye*. Similarly powerful were the sacred pig and several female beings like the *kwaning fatane kil*, the 'spirit woman who is always hungry for food'. Some of them were thought to be still existent and active, interfering in people's lives. *Yaleenye* (literally: 'the one who comes from the east') and other 'creator gods', as one may call them, shaped the Earth, making its formerly swampy surface inhabitable by wedging stones into it and by planting sacred trees. Thereby, they created the kind of soil in which plants, especially food plants, could grow and on which people could live. They also formed the beds of the large and the small rivers and instructed the early people how to lead a proper life. They showed them how to make stone adzes from rocks in the Heime Valley, how to establish men's and women's houses and how to carry out ritual ceremonies.

One mythical account narrates how the first humans dug their way from underground to the surface with their foreheads. *Yaleenye* taught them how to transform their ugly, dirty faces by cleaning them with leaves and pig fat and decorating them with ochre, and thus how to become real humans with beautiful faces. Other *isa* were those of the animals (wild and domesticated), of rivers, conspicuous rocks, trees, certain places (like that of the sacred pig *kwemdina basam*), and of all the dead (*ise dib* 'the true spirits'). These agencies dwelling in the different spheres close to or farther from the abode of humans were able to influence their life, the fertility of their gardens and other important aspects of livelihood. Diseases were thought to be caused either by one of these spirits or by harmful black magic (*kire*). Specific ceremonies (*kwetena*) to improve the condition of the sick person were carried out by male or female healers (*kwetenenang*) thought to be able to communicate with the spirit world. Sorcerers believed to have killed somebody were sometimes 'divined' by a seer (*asing ketenenang,* literally: 'someone whose eyes are sharpened') and then killed by the family of the deceased person.

Religion and secular life were not distinct, but essentially intertwined. Before dancers of the Heime Valley descended from the mountain pass to the village of their hosts, where they would carry out their rather spectacular dance performance,[146] they prayed to *Murkonye*, one of the powerful creator spirits, to make them shine and radiate with beauty and vitality. Moreover, during everyday actions, religious ceremonies were interconnected with what people did. If one were to chop down a tree with one's stone adze, one would first carry out a ceremony designed to safeguard this procedure: the adze should not become damaged, one should remain unharmed and the tree should fall quickly into the right direction. When one approached a rock shelter in the high mountains one would address the spirit believed to dwell there to receive the human visitors kindly and to protect them from the harsh and dangerous surroundings.

### 2.4.5 Physical environment

This section presents some information related to the local topography, and hence spatial coordinates as defined above that are of particular importance in this chapter. The Jayawijaya Mountains, the stretch of the central cordillera separating the northern and southern Mek groups are, like the rest of the Trans-New Guinea mountain chain, a formidable alpine massive. The lowest passes to cross from north to south or vice-versa are at about 3,700 m altitude; the highest summit of the Province, the Puncak Jaya or Carstensz Top, reaches 5,000 m, while the highest peaks in the country of the Eipo (e.g. Abom, Mt. Juliana, Gunung Mandala) are about 4,700 m high. The geological situation is such that the northern slope is much more gradual than the one on the southern side, where often very steep cliffs make human access very difficult. Still, these high ranges with their threatening cold temperatures and lack of food are commonly traversed by the local people. Their survival then depends on finding suitable rock shelters where one can build a fire and a makeshift windshield of branches, grass and bushes in the narrow, rain-protected strip under overhanging rocks. The Eipo and their neighbors undertook, and still make these potentially dangerous trips for a number of reasons, mostly for visiting trade and marriage partners on the other side of the range or for snaring or otherwise hunting the small marsupial rats and mice which live in this altitude. People actually die up there, the most feared form of death, *moke baybubuk*

[146] See Simon and Schiefenhövel 1989, which is a film on *mote* 'visiting feast'.

‘he/she died out there in the rain without protection’. The loneliness and exposure to the forces of nature is perceived as horrible rather than death as such, which was, and usually is, accepted with a fatalism produced by the normative power of the factual: around each individual there is a lot of dying, plants, animals and humans die and (apart from religious, i.e. psychosomatic forms of medical treatment) there was never a chance to do anything about this. Besides hunting and trapping, the region of the mountain forest above the regularly inhabited areas was utilized to provide building material for the houses and collect wild foods. The most important of these was *Pandanus brosimos*; the nut-like seeds of the large compound fruits have a high fat content, otherwise very rare in the Eipo diet. Other edible plants, like berries and mushrooms, were also gathered in this region.

The radius of firsthand geographic knowledge of the Eipo (and the other peoples in this part of highland West New Guinea) was about three days (fast) walking. They did not venture any further as there were no relatives on whose assistance they could count for food and protection. Walking was and is the only form of getting from one point to another. Today, a small number of airstrips facilitate travel to some extent, provided one has the money for the ticket. Small children very soon acquire amazing skills in mastering difficult terrain with bare feet. It is impressive to see the relative ease with which everyone, including old persons, walks on slippery narrow logs, wades through deep swamp and finds a footing in stretches of vertical walls. None of the informants ever complained about the necessity of walking to distant gardens, hunting grounds or villages.

### 2.4.6 Relationship to neighboring groups

The Heime River runs southwards in a kind of mirror image of the Eipomek River which runs northwards. Here, near the village of Langda, are two quarries of Andesit stones, the material from which high quality stone adze blades can be knapped. The next such place is about 150 km away (Balim Valley). The relationship of the Eipo to the Heime was, therefore, of vital importance: without stone adzes, neolithic life was impossible. Apart from this trading relationship (the Eipo paid for the unpolished stone adze blades with stringbags and food stuffs less frequent in the Heime Valley) marriage partners were often found in the two valleys across the dividing range. It is, therefore, not surprising that such trips were regularly made, either in larger groups invited to dance and feast[147] or in smaller groups of a few family members, despite the fact that the journey involves climbing from 1,700 m (the altitude of Munggona, the central village of the upper Eipomek Valley) to 3,700 m (the pass) and then approximately 2,000 m down again to Langda and the other villages on the southern side. Sometimes this 4,000 m feat was performed by the locals in a single day. The mountain range was therefore, as mentioned above, not a ‘natural border’ for these Papuan groups.

Relations with the neighbors in the Tanime Valley east of Eipomek were not as close, but good, whereas the neighbors in the Famek Valley to the west were the traditional enemies.

Warfare *(ise mal, male fey bin-)* was common (11 months during the first fieldwork period from 1974 to 1976) and caused many deaths, as did intragroup fighting *(abala)* in the village or political alliance: 25% of the men were victims of armed conflict.[148] There was no system of conflict resolution through a third party, therefore revenge and the consequent

[147] Simon and Schiefenhövel 1989, Eibl-Eibesfeldt 1995.
[148] Schiefenhövel 2001.

spiraling escalation of aggression were the cause of the high blood toll and, as mentioned above, for the high degree of cultural pseudospeciation so typical for New Guinea. Cannibalism *(ninye dina)* occurred exclusively in the course of warfare; when an enemy had been killed in a situation where his body could not be defended by his own group, it would be cut up, carried to the village of the enemy and prepared there, in the traditional earth oven, for a ritual meal. It is interesting that some persons declined to participate in these ceremonies which were, as the informants said, designed to destroy the slain enemy completely and utterly with one's teeth.[149] Since 1979 the *pax christiana* has so far stopped warfare between the Eipomek and the Famek Valley and drastically reduced intragroup homicide.

### 2.4.7 Linguistic overview

The Eipo language features predominantly a subject-object-verb order.[150] Object-subject-verb structures are frequently used as well. Compounding is the main source to denote or construe word meaning. Nouns are generally not inflected and not morphologically marked. They are morphologically simple and case marking is pragmatically handled, i.e., the actual discourse marks the subject and object of a sentence or situation. In transitive propositions the noun is profiled as the direct object, things and living beings are acted upon, they undergo actions, manipulation and creation by human beings. Gender (only for animals) is profiled by ways of compounding and derivation, e.g., using *yim* for 'male' or *kil* for 'female' to classify the noun, if needed for particular reasons. In normal speech, gender is not specified in verb conjugation. Number is expressed either by context or via the verb morphology. Nouns are modified by adjectives. More specifically, adjectives denote dimension, distance, and position in geographical and social space. They also denote color, age, value, and properties of human beings, animals, plants, and objects. The class of adverbs profiles verbs, adjectives, pronouns, adverbs, and sentences. Eipo differentiate between various adverb types such as temporal ('day', 'time') local ('down there', 'in the middle', 'into the direction of', (see the lists of terms for spatial deixis, tables 2.7 and 2.8), and modal adverbs, degree adverbs ('very'), and focus or conjunctional adverbs ('also', 'too'). Verbs denote actions and processes.

According to cognitive linguistics, verbs, as opposed to nouns prototypically profiling landmarks and objects located in space, denote motion events between such landmarks, actions, processes, or conditions.[151] Verbs designate a process unfolding in conceived time.[152] Langacker calls a verb a 'symbolic expression' whose semantic pole (a symbolic structure consists of a semantic and a phonological pole) profiles a process.[153] The quote below summarizes the idea of a process in connection to the verb as a symbolic expression unfolding in time.[154]

> A process is defined as a sequence of configurations (states) conceived as being distributed over a continuous series of points in time. Usually the separate con-

[149] Heeschen 1990.
[150] The following outline is based on Heeschen 1998, 197–287. Note that Heeschen claims also that Eipo is a noun plus verb language with the possibility that further nouns are basically treated as a free units, i.e., associated constituents are freely moved around this basic unit (Heeschen 1998, 286).
[151] Bußmann 2008, 773.
[152] Langacker 1987, 244.
[153] Langacker 1987, 244.
[154] Langacker 1987, 143–144.

> figurations are distinct, i.e. a verb typically designates a change through time; a normal verbal predication is therefore highly complex, for it incorporates as many separate conceptual situations as there are recognizable different states in the designated process.

Adapting Langacker's definition, verbs in Eipo profile various processes such as aspect and tense, but also person, number, and mood. The morphemes are suffixed to the verb. Syntactically, verbs profile predicates, and person-number suffixes agree with the subject noun phrase (NP). Note that NPs in Eipo can be constructed out of a noun or a pronoun. The grammatical suffixation of the verb can be parallel to Eipo proper nouns, which can take suffixes for human beings indicating gender. It is also important here to mention that the number of nouns is inferred either from the context or profiled by the verb's morphology and its respective suffix.

More important for the discussion of spatial language is the lexicalization process of compound verbs. This process of the formation of lexical units (as opposed to grammar) is a characteristic typological feature of Mek languages. With respect to position in space, Heeschen argues that *buk-* 'to sit' and *tek-* 'to stand/stay' are the main lexemes in profiling space.[155]

As such these verbs behave like posture verbs in most Germanic languages and, more specifically, they are similar to the above described classificatory verbs. It is not argued here that the Eipo language features a classificatory noun/verb system. Nevertheless there is a tendency for classification, albeit a weak one which is not comparable to the other Papuan languages or Dene.[156]

Foley gives an example from Waris, a Papuan language spoken in Sandaun Province, Papua New Guinea, in which morphemes are prefixed to the verbs encoding objects found inside a container (*vela*), spherical objects (*put-*), food cooked and distributed in leaf wrappers (*ninge-*), leaf-like objects with a soft stem or no stem (*lé*), leaf-like objects with hard stem (*pola-*), etc.[157]

As opposed to the rather limited encoding possibilities of position described above, in Eipo "reference to direction is systematically made more precise […]." in Eipo.[158] This implies that the main semantic function of Eipo verbs is the denotation of motion in space.[159] Hence, it is not so much a static location of the figure in a certain place but rather the trajectory of the figure with respect to the ground which has a higher degree of specificity.[160]

With respect to the assumption that Eipo features classificatory verbs, we have seen that in Dene Chipewyan various verbs encode different characteristics of the handled objects. Examples were verbs of handling, manipulation, continuing manual contact, e.g., 'give', 'hand', 'take', 'put', 'handle', 'bring', 'carry'. We see a system of verbs that encode different aspects of actions. Partially controlled action, for instance, includes an agent (e.g., 'toss', 'throw', 'hang up', 'set down', 'drop', 'lose'), while verbs of free movement are independent

155 Heeschen 1998, 234.
156 Heeschen argues for such a tendency in Eipo (personal communication); for Papuan languages in general see Heeschen 1998, Wurm 1982.
157 Foley 1986, 95.
158 Heeschen 1998, 234.
159 Heeschen 1998, 231.
160 Thiering 2013.

of an agent (e.g., 'fall' or 'tip over' in Dene).[161] This system enables the language user to profile exactly the semantic features of the object and its manner of motion to be encoded.

Finally, it has been noted that Papuan languages have a complex morphology especially in the verb system. In particular, the morphology features agglutinative patterns. The complexity of the verb makes the languages interesting especially in comparison with First Nation languages of the Americas such as Dene Chipewyan,[162] Hopi,[163] Navajo[164] Slavey,[165] all supposedly polysynthetic languages. With respect to polysynthesis Boas indeed claims that[166]

> a large number of distinct ideas are amalgamated by grammatical processes and form a single word, without any morphological distinction between the formal elements in the sentence and the contents of the sentence.

Cook notes that in Dene a verb stem cannot alone constitute a word as opposed to a noun stem.[167] He claims that the internal structure of a verb is equivalent to a full sentence in English.[168] Arguably, such grammatical amalgamation processes are also found in Eipo to a certain extent.

We conclude this subsection with some comments on tense-aspect marking. The Eipo language possesses six tense-aspect suffixes and six sets of tense-mood-person-number suffixes.[169] With respect to tense-aspect the Eipo language distinguishes today's past (past.i), near past (past.ii) and remote past (past.iii). The same applies to the future aspect, i.e., immediate (fut.i), near (fut.ii), and far future (fut.iii).[170] The following example from Eipo presents the fine-grained structure of aspectual marking.[171] It is a typical construction using the deictic morpheme *a-* 'here'.

| | | | |
|---|---|---|---|
| *aik* | *a-bu-lam-se*, | *bai* | *a-ba-lam-se*. |
| hut | here-sit-HAB-1SG.PAST.III | outside | from/here-go-HAB-1SG.PAST.III |

'I lived in this hut, I was going from here into the forest.'

The speaker of the quoted phrase, first person singular, explains that s/he lived in a specific house that was the point of departure for several trips into the garden land and the forest. The deictic marker relies on the speaker's intended orientation in which 'here' means a close proximity.

The next section presents some fundamental cultural concepts, especially in Eipo, showing some interesting culture-specific practices such as building a house. Additionally, some environment-based topographies will be presented.

161 Cook 2004a, Thiering 2006, Thiering 2009a.
162 Thiering 2009b.
163 Malotki 1979, Malotki 1983.
164 Young and Morgan 1987.
165 K. Rice 1989.
166 Boas 1977, 74.
167 Cook 2004a, 85.
168 Cook 2004a, 86.
169 Heeschen 1998, 246.
170 Heeschen 1998, 257, table 47 gives an overview of the tense-mood-person-number suffixes.
171 Heeschen 1998, 143.

## 2.5 Center, periphery and distance in Eipo

This section presents specific spatial concepts of Eipo only. This is because the data concerning these aspects are much more comprehensive than for Dene.

### 2.5.1 Building an Eipo house

Building an Eipo house is an interesting example in which an old tradition, an old practice becomes visible. This is a tradition based on joint action rather than orally transmitted knowledge. The community's center of life was the men's house (*yoek aik*), a most important point of reference. Sometimes two or three of these sacred houses existed in a community. All socially meaningful structures were usually situated concentrically around the sacred men's house, radiating out of that center. Hamlet, garden, and forest created quasi circular rings around the *yoek aik* and the sacred village ground, *asik kata*. Every place or location in the garden area is owned by someone, be it a hillside or a knoll. There is a fine grained network of place names represented in mental maps which are already very well developed in children and juveniles, who give accurate accounts of this aspect of local geography.

The mountains above the garden land, used for collecting and hunting, are connected to specific clans, but can be utilized by others as well. Sacred places can be found all around the living space, i.e., there is a sacred matrix or topology of exactly determined locations based on sacred arrays in the area.

One of the major points of departure for orienting oneself in Eipo culture was the house, either the men's house or the women's house (*bary aik*) or one of the family houses (*dib aik*). The men's house signified the center, while the women's house was at the periphery of the village. The house as a general concept of shelter can be understood as a universal place for protecting human beings from the environment, and as a place of safety and comfort, a place in which the family unit functions as a small-scale community in itself. It is interesting to survey more specifically the various usages in which 'house' appears as a location, either as a point of departure or as a place of an event in the life of the Eipo (cf. the entries under *aik* in the dictionary[172]). The house has crucial locational functions in other cultures as well. This should be of no surprise as it is a shelter and place of ritual habits in Western cultures as well. Moreover, the concept of 'house', signifying the place where a family or similar group lives, is primarily psychological, not architectural.

The following summary is based on Koch's work, specifically the section on building family and men's houses.[173] It introduces not only the technique and the different steps for building a house in the Eipo culture, but also the central significance of houses, including the various sacred objects. Moreover, several semantic structures extracted from the Eipo dictionary will be presented, if possible with their language contexts.

The noun *aik* encodes 'house' and various usages imply its importance or significance for the Eipo community. The entry alphabetically first in the above-mentioned Eipo corpus beside *aik* itself is *ninye aik bun berekilbin* 'people are meeting in the core of the house'. The entry for *aik* contains a number of related expressions specifying the function and importance of the house. First and foremost *aik asin* means the 'fireplace in the house'. Further, *aiktam* designates 'in the house', 'inside'; note the locational construction N + suffix to encode

[172] Heeschen and Schiefenhövel 1983.
[173] Koch 1984, 38–56.

'inside' based on the interior of the house. The way home or to the house is encoded as *aik bisik*. The term *aik* is also used for a sickness caused by a spirit as in *aika* or *aik mek dikmal* 'a sickness caused by a spirit': a severely sick person does not leave the house any more, often until he or she dies.

The basic form of an Eipo house was round with a cone roof, while less well built houses were either round or rectangular with a ridge roof; today quite a large variety of shapes and sizes are found in the Eipo villages. The average diameter of a family house was between 2 and 3 meters and the height about 2 meters. The average men's house had a diameter of between 5 and 6 meters with a height of about 4 meters.[174] Most of the houses had an elevated ground floor at a height of about 40 cm to 100 cm above the actual ground. The space underneath (*ambonga*) was sometimes used as a hog house, to store firewood, and to keep the ashes. Hence, it was a kind of a stockyard for all sorts of things in general. In the case of the men's houses it was also where the spirit houses, *isa aik*, were placed. The living space measured about 1 to 2 square meters per person.[175] These close quarters were not perceived as a disadvantage by the Eipo, but as a welcome means to literally stay in direct contact with each other. Building a house is primarily men's business and the process of building a house is classic group and assembly work. All the necessary construction material, including the planks for the walls which are hewn where the specific trees grow in the mountain forest, is gathered weeks beforehand, i.e., the actual process of building the house is similar to assembly work on a construction site. Women participated, even in the building of a sacred men's house, by carrying building material to the storage places or the actual building site. They still do this today. Reusable material from old houses was, and is, incorporated into the new building.[176]

The Eipo mainly used one universal tool, the adze *ya* with a blade made of stone. This specific kind of well-made hatchet was used to fell trees, to split up logs and to shape posts and other building material, including rattan for binding. One could say that the stone adze was some kind of 'leatherman' or 'swiss knife' for the Eipo in terms of a universal tool. The different stages in constructing and building a house will be described below with respect to the former tradition of building a men's house. This socially, politically and religiously most meaningful building was the most important anchor in the Eipo community. Its continuity was granted by keeping the same location and the same sacred objects and by using parts of the old building material for the new building. Koch and Schiefenhövel (2009) documented the reconstruction of the old men's house of the village of Munggona, called the *Binalgekebnaik*. It had a diameter of approximately 6 meters. Planning took place far in advance and some of the sacred rituals were already carried out in the forest. To start off, the men removed the sacred digging stick *kwemdina kama* (a relic from mythical times, the beginning of creation) and placed it against another men's house during reconstruction work. Normal digging sticks, *kama*, were used as tools, e.g., to dig, to harvest, to weed, and to level the ground.[177] The *kwendina kama* was the most important sacred object in the southern Eipomek Valley, a holy grail, so to speak. Then the men took off the cone roof and placed it beside the building site. The ensuing demolition of the old house was accompanied

[174] This difference in size already indicates the significance of the men's house.
[175] Röll and Zimmermann 1979.
[176] Usually the roof of an old house is used again, along with planks for the walls and other pieces that are still of good quality.
[177] Michel 1983, 66.

by sorting out usable material; phrases describing this are *aik nonge ulobuka dobnab* 'we take away/pull down the house (except the roof)', *aik nonge duk'namab* 'we will take the house apart', *aik kolubrabuk* 'one broke down the house/the house is destroyed'. During this process the spirit houses *isa aik* became visible. After leveling the ground the men brought *ayukumna,* long house posts, which provided the main structure of the house. This stage was orchestrated like a procession and performed in an ecstatic, rhythmic dance, accompanied by the typical inspiratory whistling which provides the basic rhythm during Eipo dance feasts. The *ayukumna* were driven into the ground to a depth of about 40 cm. Rolls of bark from a specific confer were brought into the circle of posts to check whether they fitted the diameter of the house. This was the only type of measurement done; all the other pieces were placed intuitively. The bark would later cover the floor, providing a soft, even top layer (*amsona*).

The next step was to set the four slim poles *ateka* to delimit the fire place (*ukwe asin* 'fire place in the house'). Two of these are called *mem ateka* (taboo poles) and have a sacred meaning. They were covered with fern leaves to protect the men's hands from being burnt by the hot poles. When the men brought them, they again danced and chanted rhythmically. Several layers of circular transverse struts *afanya* were then carefully bound to the *ayukumna*. They held the house posts in place and provided a horizontal rim supporting the floor. Later another ring of *afanya* was fixed at the upper end of the posts to stabilize them and provide support for the roof.

In building a men's house or other houses the next step was to place, in a criss-cross fashion, long flexible sticks on the horizontal rim provided by the *afanya*. This created a flexible floor which slightly slanted toward the middle as an interesting feature which helps utilize the heat of the central fireplace more efficiently. To give more stability to the floor layers (30 or more men may be inside the men's house at a given time), crossbeams *wanun yo* were squeezed horizontally underneath. For family houses reed *(Miscanthus floridulus, fina)* was sometimes used instead of wooden sticks as it is easier to come by. Short planks, *abelenga*, reaching from the ground to the level of the floor, were fixed with rattan, the classic material for all bindings. This first circle of short planks typical for men's houses blocked the view of the space below the floor where new little spirit houses were placed in the meantime. The planks forming the wall of the men's house above the floor and reaching to the roof were gradually fixed as well. Even today these planks are still cut from a tree (*Galbulimima belgraveana, lue*) which easily splits so that flat, even boards can be produced. Today, although Christianity has superseded their belief in spirits, the Eipo still seal the walls of their houses as securely as possible: Little openings, cracks etc. could provide an entrance for spirits or other harmful agents, and in former times also for arrows.

The following language examples reflect the importance of spirits in the old Eipo tradition: *aika* 'sickness' (caused by house spirits); *isa kum angnulamak* 'the spirits come up to the neck (i.e. they eat the person, make him/her fall sick)', *aik mek dikmal* 'water is stuck to the house/(metaphorically) the spirits are catching them (the inhabitants)', *isenang* 'the spirits, (met.) the enemies', *kingkin bisik keniklamak* 'they are caulking the clefts (between the boards of the wall of the house as protection against arrows and spirits)'. Especially the last example indicates how important it was to protect the house from the spirits. In the small, roughly built houses underneath the ground floor of the men's houses they had an official abode and, at the same time, were contained so that they did not come into direct contact with people.

The most devastating events, believed to have been caused by a giant spirit *(Memnye)* living deep down underground, were the two earthquakes in June and October 1976, both measuring above 7 on the Richter scale. Throughout, the whole ritual connected to building a men's house and various kinds of sacred ritual practices were thought to be necessary to calm down or appease the ghosts. It should be noted that the Eipo regarded earthquakes as well as sickness, accident or other mishap as punishment for broken taboos or disrespect toward the spirits. The massive earthquakes, in the course of which several Eipo died and which completely destroyed the whole village of Munggona and its sacred men's houses, including many sacred objects, had a deep impact on the people. This facilitated the transition to Christianity and thereby initiated the very rapid process of acculturation. As a consequence, the transmission of cultural knowledge passed on orally via myths was partly interrupted.[178]

Returning to the description of the sacred *Binalgekebnaik* men's house's construction, the next step was to construct the support to hold the conical roof, the main weight of which was resting on a short central pole which was attached to the four poles, *ateka*, delineating the fireplace. The outer rim of the roof was resting on the upper end of the house posts (*ayukumna*) stabilized by the top ring of *afanya*. Finally, the old roof was carefully put in place; many men, and sometimes women even, participated in this final climax of sacred actions.

### 2.5.2 Natural limitations in Eipo

Mountains and the sky mark the limits of the Eipo world. The place where the mountain and the sky meet is called *motokwe ime ebrarik* 'mountain (or land) and sky, the two meet'.[179] Beside the sky as an obvious visible limitation, the mountainous region has its repercussions on the Eipo culture and language in terms of places, and natural limitations. See the following examples, all indicating the importance of environmental landmarks such as mountains and their function in Eipo culture. Table 2.4 presents various semantic differentiations of the concept 'mountain' in Eipo.

Clearly and not surprisingly, the mountainous region has a culture-specific and central meaning for the Eipo, as it has in any other region with such environmental specificities.[180] Hence, mountains have several functions in Eipo. Beside the above meanings, some related concepts are discussed below.

The Dakul and the Lyene are particularly important mountains formerly believed to be the 'mythical abode of Sun and Moon' (1143, 3732). The direct connection between the Moon and the Eipo region is expressed in the term *Yaburye* 'mythical river attributed to Moon and Sun' (5683). Both the Sun and the Moon have specific cultural values as in *ketinge-ton wale-ton Dukuramduweik a-kururak* 'Sun and Moon, the two of them created the Dukuramduweik-men's house here' (3038), or *im maka* 'secretion of the sky (code for: Sun and Moon)' (3776/4).

In table 2.5 there are some descriptions of the various stages and some metaphorical expressions relating to the various positions of the Moon, which is also connotated as female: *wale are kil* 'the Moon is a woman' (2641/1). The examples present various metaphors of the Moon in its different stages in Eipo.

[178] Heeschen 1990, 143.
[179] Which might be translated as the concept of a 'horizon' (1692).
[180] Berthele 2006.

| | |
|---|---|
| *motokwe aryuk-* | '(mythologically) to pile up' or 'create the mountain' (194) |
| *motokwe berengne* | 'a world of emptiness' or 'solitude, i.e., without any plants' (475) |
| *motokwe akonum bereksingibuk* | 'the land lay bare, nothing grew' (476) |
| *motokwe cange wik* | 'mountain is spacious' or 'big' (1050) |
| *motokwe dandoble* | 'the mountain' or 'the area is uninhabited' (1176) |
| *motokwe kon dinib'mak* | 'they go round the ridge of the mountain (in order to avoid climbing it)' (1442) |
| *motokwe dok* | 'flank of a mountain' (1502) |
| *motokwe dub* | 'top of a mountain' (1592/2) (*bebengdina, bebengdin* = mountain top (a mountain range is often the border between two regions, e.g., between the Eipo and the Marikla, who were enemies; the same metaphor is used for the border between the world of man and the world of the spirits) |
| *motokwe seringsarang fabminyak* | '(magically) the empty earth shall bear flowers' (1797) |
| *motokwe filibable* | 'the mountain becomes smooth' or 'flat' (metaphorically for 'to faint', 'to become unconscious') (1962/1) |
| *motokwe kwakwa lakabdanamle* | 'the world (= mountains) will be transformed into a butterfly (when praying to the ancestors it is asked that the leaves of all food plants should move in the wind like the wings of a butterfly)' (3102/1) |
| *doa motokwe-dam lelelamle* | 'the clouds are piling up at the mountain there' (3425/7) |
| *loun motokwe* | 'an area or a mountain not under taboo where everybody is allowed to go' (3620/1) |
| *marman, motokwe marman* | 'transverse (path) under a cliff' (3867) |
| *motokwe kon* | 'mountain top ridge' (4087/4) (*sin* 'mountain top', 'high plateau' |
| *motokwe tob-nang* | 'those who know about the world are able to explain the world' (4087/6) (*toba* = 'it is there'; 'is/are present', 'continuous') |
| *motokwe yim* | 'mountain (ridge edge)' (4087/7) (*bisik wamumna* 'ridge') |
| *tarekna motokwe* | '(lit.) cold mountains' or 'high mountains' (4087/9) |
| *motokwe erelamle nun gum ob* | 'the mountains arose at a time when we weren't yet there' (4448/2) |
| *sik motokwe* | '(this is) their mountain' or 'area or hunting ground' (4708/2) |
| *motokwe tilibak* | 'places or areas where the trees grow densely or where there is a lot of growth' (5181) |
| *motokwe yupa* | 'pass' (5920) (*Tekiltakalyan* 'to climb up and meet', 'to meet on a mountain top, a pass' (5103) |

Table 2.4: Semantic variation of 'mountain' in Eipo

| | |
|---|---|
| *wal su eleklamle* | 'the Moon is wrapped in leaves/can no longer be seen' (5450/6) |
| *wal yulamle* | 'the Moon is cooking (in the earth-oven)', 'new moon' (5450/7) |
| *wale yang kelamle* | 'the Moon is or becomes like a tusk', 'crescent moon' (5775/2) |

Table 2.5: 'Moon' in Eipo

As is apparent, the Moon in its different stages is encoded via figurative usages that intuitively make sense to a Western speaker as well. The general importance of the Moon for fertility is evident also in Eipo. The Eipo interpreted the waning and waxing of the Moon as phases of its menstruation. In particular, the New Moon was thought of as residing in a heavenly women's house, just as women during their menstruation reside in the women's house for about 3–4 days. The Moon marks, additionally, the connection between a mythical spirit and the bare landscape, in particular the high surrounding mountains.

### 2.5.3 Distance in Eipo

The data from the dictionary and Heeschen's grammar, the various ethnographic films, and the myths[181] suggest that the Eipo do not possess abstract terms for distance, area, and volume. In one instance, an interesting observation was made. Work at the airstrip, carried out by the local people under supervision of Wulf Schiefenhövel and an assistant from Ilu, a mission station in Dani country west of the Mek area, had been going on for many weeks. The general shape of the landing field was visible. It was delineated by longitudinal ditches which were dug to drain off the substantial amount of daily rain water at the sides. The width was thus determined, as well as the lower and upper end. When it was announced that Wulf and Grete Schiefenhövel would walk to Bime, the nearest mission station which had been opened two years previously and from where the advance group of the German Research Team had started its five-day walk to Eipomek, several men and boys said they would like to come along. As soon as the group had arrived in Bime, some men looked for string, i.e., long sections of bast and other fibres and similar material. They connected many pieces by knots and when the string was long enough, measured the width of Bime airstrip, marking its size before the string was rolled up and stored in one of the men's bags. Schiefenhövel was quite surprised by this activity and asked what they were doing. They answered: We are comparing (*kiklib-*) the 'axillary wing' (*ke fol)* of the airplane. We know that the plane can land here and we want to check whether the *ke fol* of the airstrip we are building with so much effort in Eipomek has the same size so that the plane can also land there. 'Stone-age' Eipo were checking the job of the white fieldworkers as they wanted to be sure that the engineering was done according to standard. This is quite a scientific procedure. They were happy when, on return from Bime, they found that the *ke fol* of their future landing field had the proper width. This measuring was not done by counting steps or feet, but by a quasi holistic act of comparing. This act of comparing reflects the idea that distance and length is preserved regardless of place and direction, a central cognitive structure in the gestalt-like mental model of space.[182]

[181] Heeschen 1998, Blum et al. 1979–1996, Heeschen 1990, and Heeschen and Schiefenhövel 1983.
[182] Thiering 2014, see also Chapter 1 of the present book, in particular sections 2 and 3.

| | |
|---|---|
| *boltak-, boltakab-* | 'to keep distance from someone or something' (732) |
| *yanyane faye bin-* | 'to leave foot-prints (song and dance texts for) to walk long distances' (1874) |
| *inib-, enib-* | '(to make see) to search, to invite over a long distance' (2190) |
| *karen, karin* | 'unoccupied, keeping distance' (2395) |
| *karenkaren balamak* | 'they go separately, keep distance' (2395/2) |
| *aik kwakne bisik* | 'the path through / in between the houses' (3098) |
| *lukfara ban-* | 'to look out, to look out into the distance' (3647) |
| *nisin diberen-* | 'to look into the distance' (4395) |
| *onob-* | 'to refuse, to turn down, to keep at a distance' (4527) |
| *yan onolbin-* | 'to make a big step (on the day when the sacred men's-house is built one is not allowed to walk a long distance. The taboo is apparently nullified by taking a big step over a puddle or a small pond.)' (4528/1) |
| *tamublabdongob-* | 'to gain a greater distance to someone who is following, to keep a distance when walking' (5000) |
| *tekisib-* | 'to keep a distance' (5107) |
| *tekisibnin balamak* | 'the women keep a distance (to the men while walking)' (5107/1) |
| *usamkila* | 'clouds rising in the distance' (5411) |
| *webrongob-* | 'to follow closely, to be attracted' (5526) |
| *winilkidik-* | 'to wander about, to walk big distances (said of the ancestors)' (5627) |
| *bisik* | 'way, path, direction' (612) |
| *bisik dukuble* | 'the path/entrance is just wide enough (to be able to carry s.th. through)' (612/5) |
| *bisik kwangdanya* | 'fork in the road' |
| *bisik lebarikna* | 'the circumventing' or 'avoiding of a steep part of the path' |

Table 2.6: Various expressions of distance in Eipo

A somewhat similar idea in terms of using straight lines, but without comparing lengths, is implied, as already mentioned above, in the practice of delineating garden lands. *Wa* (usually old gardens reused after approximately 15 years of lying fallow, sometimes newly cleared primary forest) is divided into individual plots without employing fixed units of distance. The borders of the plots are commonly marked by small trees *(yurye, Cordyline terminalis*, a sacred plant in many regions of the Pacific) in such way that the line connecting the *yurye* is defining the end of one plot and the beginning of another owned by families and passed on in the patriline. To encroach into the land of another family is considered a serious offense and leads to open conflict.

Some morphemes indirectly represent ideas of distances such as 'in between', i.e., a specific distance between two landmarks. They are presented in table 2.6.

The most common word to express distance is *fera, fere* = 'distant', 'far away', requiring a long walk. The term *fera* as well as the various phrases presented above do not, of course, entail a specific, precise measure of distance, as steps, miles or kilometers. But for an adult member of the Eipo society, who knows her or his territory extremely well and has also walked to places further away, this term is sufficient. The problems arise when foreigners, like white researchers, hope they can extract some measurement of distance or time from their informants: *fera* can be quite close, but also very far away. Hence, it can be stated that there is no technical term for distance in Eipo, but a variety of context-dependent phrases and words, for which one can use the term 'distance' as a translation.

Nevertheless, with respect to building houses, traps or bridges the Eipo are able to conceptualize the exact structure and architecture and order of actions necessary to assemble various materials to build the different types of houses, the technically advanced traps (as in Dene Chipewyan) or a bridge. It is apparently not necessary to have an explicit and abstract measurement to construct buildings or even the rather sophisticated cane bridges spanning across wide rivers, examples of neolithic high-tech. Similarly, abstract terms for distance are not necessary for constructing stable buildings and functioning devices whose stability and functioning we would today explain using the principles of physics. It is not necessary to know, e.g., the abstract concept of the number $\pi$, i.e., it is not important to know and apply the idea of a circle in a strictly geometrical sense. The Eipo and other peoples have developed practices and ritual actions which fulfilled their purpose more effectively than others and thus became part of their culture.

## 2.6 Representations of spaces in Eipo and Dene Chipewyan

In this section, the two languages under survey are compared with respect to their spatial concepts, ways of spatial categorization, and use of spatial markers of environmental landmarks.[183] As stated in the introduction, our interpretation of Eipo and Dene spatial concepts is guided by the fine-grained analysis of Hopi ideas of space (*Raumvorstellungen*).[184] Malotki's survey seeks to present the various facets of this language in their function of encoding spatial relations in specific detail.[185] Eipo and Dene Chipewyan present specific environment-dependent encoding patterns mirrored in the languages.[186] The mountains and rivers as important limitations in Eipo and Cold Lake in the Dene culture show their repercussions in the language patterns and the carving-up of spatial concepts on the language level. In the following sections we will present a variety of examples from Eipo and Dene showing various ideas of space.

### 2.6.1 Orientation in Eipomek

The following summary on Eipo structures presents some firsthand data.[187] As we have described above, in the Eipo religious tradition humans appeared on Earth from the underground and gathered in groups. Their most important place became the men's house. It was

[183] Mark et al. 2011.
[184] Malotki 1979.
[185] Svorou 1993, Thiering 2013.
[186] Thiering 2014.
[187] Heeschen 1990, Koch 1984, Koch and Schiefenhövel 2009. A further source are Schiefenhövel's recent fieldnotes 2008–2010.

a crucial place for securing the life and prosperity of the hamlet. It was hence the center both as a real location and as a spiritual place. From the center to the periphery there was a network of paths and additionally of arrangements and limitations that began inside the men's house, e.g., with a specific seating arrangement and positioning of the sacred objects. It has to be added that each Eipo village had one or two women's houses, which were also sacred and taboo for the men. In some respects they are the equivalents of the men's houses for the women. This social organization following a marked gender dichotomy, and the specific environmental conditions, are well established in the language structure and religion, i.e., many points of orientation are semantically filled with culture-specific entities or landmarks. The following two examples show this specificity.[188]

| | | | |
|---|---|---|---|
| *a-kame* | *ara* | *lulukene* | *mem*. |
| here-stick | THEME | shake/make(VN) | forbidden |

'As to this sacred digging-stick, it is forbidden to cause it to be shaken.'

| | | | |
|---|---|---|---|
| *am* | *bob-m-ik-ine*, | *ou-Dek* | *bob-ik.* |
| Taro | carry-DUR.-3PL./PAST.III-SCENE. | down/there-Dek | carry-3PL./PAST.III |

'They were carrying the taro, and then they carried them to the Dek River down there.'

The examples present some important and relevant objects in Eipo, e.g., the sacred digging-stick *kama,* sometimes pronounced *kame*, which was kept as the most important religious item, and the ritually important ancient food plant *am* 'taro', or specifically meaningful places, e.g., the Dek River, or the Northern lowland area. Moreover, the examples indicate the importance of cultural-specific habits relying on specific practices, e.g., the digging-stick as a sacred object is also responsible for a certain order or ritual as in *kama bukwotebnin yanamuk*, which can be translated as 'the primeval digging-stick came putting everything in order and smoothing everything'. As the stick of creation it was kept in a specific place, some kind of shrine in the men's house.[189] Interestingly, in all cases a deictic marker (*a*) is used to indicate the exact position of the place, the direction or the event.

Eipo speakers orient themselves in their mountainous environment by a finely meshed network of names for mountains, hills, slopes, rivers, and plains.[190] Heeschen describes the use of this environmental topology:[191]

> Eipo speakers mainly use the spatial deictics as a condensed and abbreviated structure in face-to-face-communication: here the deictics are accompanied by a pointing gesture.

Basic orientation in space for the Eipo is, as has been mentioned above, provided by five deictic points of reference based on the speaker's position, 'here', 'there', 'up-valley',

[188] Heeschen 1998, 270.
[189] Koch and Schiefenhövel 2009; Heeschen 1990, 85.
[190] Foley 1986; Heeschen 1998, 143.
[191] Heeschen 1998, 143.

‘down-valley’ ‘across (the valley)’.[192] The basic set of deictic markers consists of the following morphemes, taken from the dictionary.[193]

| | |
|---|---|
| *a-* | ‘here’ |
| *ei-* | ‘up there’ (see below for further examples) |
| *ou-. u-* | ‘down there’ |
| *or-* | ‘across here’, ‘across the valley’, ‘on the other side’, ‘the other slope (but not upwards)’ (4536) |
| *or-asik* | ‘the hamlet over there’ (4536/1) |
| *or-deibsilyam* | ‘put it there (at the same height)’ |
| *ortiba* | ‘it’s over there’, ‘across the valley, spot across the river’ |
| *er-* | ‘across the valley/the river’, ‘upward of own position’ |

Table 2.7: ‘Here’ and ‘there’: General deixis in Eipo

These examples exemplify the various usages of the dual distinction between ‘here’ and ‘there’, i.e., the horizontal distance and place of a speaker being ‘here’ and the vertical ‘up’ and ‘down’ distinction. All of the usages are rather unspecific in terms of metrical distance between the speaker and a potential hearer. We also see the importance of orientation depending on the environment, e.g., ‘river’ and ‘valley’. The prefix *d-* is added to deictic morphemes to form longer distances or sharper contrasts. The above data set presents a more detailed semantics of the basic deictic markers. The added prefix increases the spatial semantic detail in the encoding of proximal, medial and distal distances. In addition, vertical specification or specification of altitude is given in greater detail than in the examples above.[194]

| | |
|---|---|
| *da-* | ‘here’ (in a wider area around the speaker and hearer, here and there) |
| *dei-* | ‘very far up there’ (across the mountains) vs. *fera* = ‘far way’, as opposed to *dam*)’; *dam* = ‘close by’, ‘short (way)’ |
| *dam banmarak* | ‘the two of them are coming closer’, ‘they are approaching’ |
| *dou-* | ‘very far down there’ (‘very far down the valley’) |
| *dor-* | ‘very far across the ridges in the next valley’; ‘at same level or lower than own position’ |
| *der-* | ‘very far across the ridge in the next valley’; ‘higher than own position’ |

Table 2.8: ‘Here’, ‘there’ and ‘far across’: Specified deixis in Eipo

[192]Note that the three valley-related orientations function in Eipo just like cardinal directions in Eipo. Hence, the frame of reference is in a transition from a relative to an absolute frame. For a survey of frames of reference, see Levinson 2003, Levinson and Wilkins 2006.
[193]Heeschen 1998.
[194]Heeschen 1998, 144.

These examples indicate that Eipo rely on a topographical system which includes, in these last cases, distances in various metric situations, i.e., proximal, medial and distal. In the example below, the deictic marker refers to a distance between speaker and another group of people.[195]

| | | | |
|---|---|---|---|
| *Marikle-nang* | *lukenyan* | *or-yan-ma-se-ak,* | *a-mab-ma-lam-buk.* |
| Marikle-people | night | from/across-come-DUR-us-3PL.PRES, | here-sleep-DUR-2SG.PRES-when(DS.) |

'During the night the Marikle people come to us from across (the valley) there.'

Syntactically the deictic markers are bound morphemes that combine with other parts of speech such as verbs, nouns, postpositions, and predicating suffixes.[196] Here, the deictic marker encodes the trajectory of the figure (the Marikla people, i.e., the enemy living across the valley, are coming) and their transition from their home location (the unspecific 'from across the valley') to an implied speaker or vantage point ('us').

An interesting example in terms of an imagined location is given below.[197]

| | | | |
|---|---|---|---|
| *a-kil* | *ara,* | *a-yanga-lam-lye-ak-da* | *a-tek-am-lul.* |
| Here-woman | THEME | here-come-HAB-3SG.MED-at-but | here-stand-PERF-3SG.HORT |

'As to the woman here, she may have come to the place where he might have been standing.'

The deictic marker *a-* used in the above example encodes, in the first instance, a particular place. In the second and third instance, it encodes an imagined or abstract space that is removed from the speaker to a distance in which 'here' (depending on the speaker) is not the location of the speaker in a real context. The locational marker removes the scene from the actual speaker/discourse. Heeschen argues that the Eipo are imagining a place they do not know. From a morphosyntactic point of view it is interesting that the deictic marker is used repeatedly. Every possible location is marked for each location of the figure and the ground, thereby identifying the places at which the actions of the two phrases take place.

The example below gives a flavor of the encoding of imagined things that a speaker describes to a hearer who does not know the spatial landmarks.[198]

[195] Heeschen 1998, 143.
[196] Heeschen 1998, 143.
[197] Heeschen 1998, 144.
[198] Heeschen 1998, 144.

| | | | |
|---|---|---|---|
| *Aike* | *irikna* | *a-ub-ma-le-to-ak,* | *ou-tonun* |
| hut | edge | here-be-DUR-3SG.PRES-as-at | down-as |
| *li-am-ik-ye-ak* | *aik* | *dike* | *ou-deli-lam-ak.* |
| put.into-PERF-3PL.MED-and-at | hut | food (ritual) | down-put-HAB-3PL.PRES |

'They put away the food at one edge of the hut, at a place which is similar to this one here (the speaker points to something), in a similar way they have put down there (things into a stringbag).'

It is apparent that this last example can only be understood in its real speech act context since the speaker is actually pointing at some place. As outlined above, another interesting aspect is the delimitation via mountains and thus a seemingly unspecific distance.[199]

| | | | | | | |
|---|---|---|---|---|---|---|
| *An* | *yuk* | *asik* | *a-ub-na-lyam,* | *nun-da* | *der-motokwe* | *bi-nam-ab.* |
| you | alone | hamlet | here-be-FUT.II-2SG.HORT | we-but | very/far/across/up/there-mountain | go-FUT.III-1PL. |

'You alone should stay in this hamlet here, but we will go to the mountain very far across there.'

The idea of 'very far across there' seems rather unspecific for a speaker unfamiliar with the environment, but for the Eipo speaker the distance to the central range in the south is very well known. Moreover, it seems evident that the future tense marker encodes a distance in space as well. Note that future.iii is used for long-distance journeys, while future.ii is used to designate 'staying here'. The hortative (mode of the verb specifying an act of collective action) construction *lyam* encodes the mode of the verb specifying a collective action, i.e., the part of the English translation introduced by 'You should stay' and ending with 'we will go'. Both utterances are related to specific places, the 'hamlet' and the 'mountain'. The opposite of *asik* 'village' or 'hamlet' is *bay* meaning 'outside' and thereby carrying the notion of 'wilderness', 'uncontrolled', 'dangerous' (cf. *bure, budu* 'outside'; *bure ketib-* 'someone who stays outside, comes back to the village late'; *bure* is purely deictic, i.e., not used metaphorically to signify danger, threat etc.). *Motokwe* has several additional meanings such as 'land', 'landscape', 'region', 'place', and 'world' (see table 2.4). The prefixed bound morpheme *a-* has, as already shown, several meanings depending on the context as summarized in table 2.9.

The prefixed deictic marker *a-* encodes two possible locations depending on the speaker's intention to indicate a specific direction, i.e., 'here' and 'there'. Note that the morpheme *ortam (or-tam)* encodes, as mentioned above, 'over there'; 'across the valley'; 'across the river' (indicating direction) (4544).

The next section presents some general ideas of space in Dene, in particular concerning delimitations and limits, that are mirrored in the language.

[199] Heeschen 1998, 144.

| | |
|---|---|
| *a-* | 'here', 'there' (as opposed to 'over there') |
| *a-tam* | 'here', 'this way' (indicating direction and place; *-tam* = 'side') (cf. *u-tam* = 'down there', 'down the valley' (indicating direction); *u-tiba* = 'it is down there', 'down the valley/the river') |
| *a-teba* | 'here it is' (*-teba* = predicative particle with deictic pronouns) |
| *a-tebuk* | 'here', 'this here' (*-tebuk* = predicative particle with deictic pronouns, pointing to something which is past or which had been mentioned before; what has been mentioned in the past or in the preceding conversation and is thus known to the speaker) |
| *a-binmal* | 'here'/'there he/she/it comes' |
| *a-bisik* | 'this way', 'along here' |
| *a-motokwe* | (lit:) 'this mountain here', but also: 'here', 'with us', 'in our place' |
| *a-nirya* | 'all this' |
| *a-yo* | 'the wood'/'the tree here', 'this tree'/'this wood' |

Table 2.9: Deictic expressions in Eipo

### 2.6.2 Orientation in Dene Chipewyan

The previous section provided some basic spatial concepts in Eipo based primarily on environmental landmarks. This section presents some data from Dene Chipewyan and neighboring languages. It is based on Thiering's field work.[200] This language has interesting spatial terms such as 'up above' (*yudaghe* 'above, at a certain place above'); *betthiye* 'above it (current, wind)', 'down below', 'upstream' or 'up river' (north), 'downstream' (south), 'up from shore', 'down toward shore', 'out to sea or forward' (into or out to open sea), 'inside', 'outside'. This set of terms are very similar to the corresponding ones in Eipo. Most of the concepts are related to lakes or rivers, more precisely, particularly those around Cold Lake. Related languages such as Carrier, Eyak, Hupa, Koyukon, Navajo, Slavey, and Tlingit also encode spatial concepts based on the immediate environment, such as rivers they traveled to, e.g., for fishing.[201] As we shall demonstrate below, Dene behaves similarly to its neighbor cousins. Table 2.10 presents some of the affiliated languages, Tlingit,[202] Carrier, Koyukon, and Hupa and some of their spatial concepts that are similar to those in Dene.[203]

[200] Li 1946, Cook 2004a.
[201] Leer 1989.
[202] See Thornton 2011, 275–289.
[203] Leer 1989, 613, 622, see also Kari 2011, 239–260; the following abbreviations are used: all = allative; loc = locative; abl = ablative case, suf = suffix.

| Tlingit | Carrier (all, loc, abl) | Koyukon (all, areal) | Hupa (loc, suf) |
|---|---|---|---|
| *ké-* 'up above' | *-do*, *-doh*, *-des* 'up above, over' | *-dege*, *-degu* 'up above' | *-dah*, *-de* 'up' |
| *ye-, ya-* 'down below' | *yo-*, *-yoh*, *-yes* 'down, underneath' | *-yege*, *-yegu* 'down below' | *-yah*, *-ya* 'down' |
| *naka*(north) 'upstream (north-east)' | *-nu?*, *-nud*, *-nuz* 'upstream, away up (from the outlet of a lake)' | *-na'e*, *-nuye* 'upstream, back behind, to the rear' | *-nage*, *-nah-* 'upstream (south-east)' |
| *-?ix-ka* 'downstream (south)' | *-da?*, *-dad*, *-daz* 'downstream' | *-do'*, *-duye* 'downstream' | *-de?*, *-da-* 'downstream (north-west)' |
| *-dag* 'up from shore, interior' | *-no*, *-noh*, *-nes* 'north' | *-nege*, *-negu* 'up from shore, up on or above shore (from water), toward back (of house)' | *-dage*, *-dah* 'away from the stream (north-east)' |
| *yeg, ?ig* 'down toward shore' | *-cen*, *-cid*, *-ciz* 'down toward a body of water' | *-ene*, *-uye* 'down to shore, toward front (of house)' | *-ce?ne*, *-sen-* 'toward the stream, downhill (south-west)' |
| *de-ka* 'out to sea, out into open' | *-nes*, *-nes* 'forward' | *-nela*, *-nelye* 'ahead, out on open water' | |
| *yan* 'across, on the other side (of water)' | *-ni?*, *-nid*, *-niz* 'behind, in the rear, away from a body of water' | *-nane* 'across, on the other side (of water)' | *-mane*, *-?an-* 'across the stream (south-west)' |
| *-nel* 'inside' | *-yan*, *-yad*, *-yaz* 'on the opposite side (of the water)' | | |
| *gán* (north) 'outside' | *-?en*, *-?ad*, *-?az* 'away, off' | *-?ene*, *-?uye* 'off to the side, away' | *-?a*, *-?a* 'beyond, on the other side' |

Table 2.10: Environmental spatial concepts in Tlingit, Carrier, Koyukon, and Hupa

It is not necessary to present a detailed analysis of every spatial morpheme in the different languages here. What is evident, and striking, with respect to the subject of this chapter is that in all these languages, spatial marking is aligned to some environmental landmark, i.e., house ('toward or back to the house'), water or river (*des* in Dene; up- or downstream). In addition, the direction of the water is paralleled with cardinal directions as in Hupa. The examples further indicate a striking similarity to the Dene data. Like the affiliated languages, Dene bases its orientation also on environmental landmarks, but additionally uses the cardinal system (*sayesi* 'East from under the Sun'; *-da, yethda* 'The Great Bear constellation'). For example, the North, *yatthé*, profiled also 'up' (cf. *tthi* 'in the north'; *yatthi* 'to the north'; *ghadhe* 'the West'; *dási* 'west', 'from down river', 'to the west of'). The direction of the wind (*betthiye* [up current], above it (current, wind)) is also marked by the cardinal direction, i.e., *tthísníltsi* is 'wind from the North' and *nasniltsi* encodes the 'wind from the South'. The Dene today even possess the concept of North and South poles (*yatthé néné laghil* and *níl holaghe*, respectively).

More precisely, the Dene Chipewyan territory was strictly limited by the water systems, i.e., large streams and numerous lakes, but also by extensive swamps, prairies, barren areas, and forest.[204] The main limitations were the water systems as can be seen in the following expressions in Dene delimiting the territory. *Kechagha-hotínne* 'down-stream they-dwell' is placed west and south-west of Great Slave Lake, near the mouth of Hay River along Mackenzie River, and the lower course of Liard River.[205] The expression *Kaí-theli-ke-hotínne* means something like 'willow flat-country up they-dwell'. This region is centered around the western end of Athabaska lake at Fort Chipewyan and extends northward to Fort Smith on Slave River and southward to Fort McMurray on Athapaskan River.[206] *Kes-ye-hotínne* 'aspen house they dwell' encodes a place near the head of the Churchill River system (Lac Isle la Crosse, Portage la Loche, Cold Lake, Heart Lake, Onion Lake). *Háthé-hotínne* 'lowland they-dwell' is the region of Reindeer Lake draining southward into Churchill River. *Sa-yísí-dene* 'Sun under (the eastern) people' is in the barrens between Reindeer lake, Hudson Bay, and Chesterfield Inlet. *Tandzán-hotínne* is on the northern shore of Great Slave Lake along the Yellowknife River (*Dení-nu-eke-tówe* 'moose island up lake-on'). The *Hli-chá-dene* are the 'dog flank people' (Dogrib) between Great Slave Lake, Great Bear Lake, and La Martre and Coppermine River.

With respect to deictic information, as seen for Eipo above, Hopi as a very distant cousin language of the Athapaskan language family presents for all three distances 'here', 'there', 'over there' in the example below, but expands the deictic system into a more refined pattern including medial information (which is known from Dene as well).[207] Note that the basic space structure in Hopi is based threefold on the following case system: a locative, a destinative, and an ablative determine the place or site, destination, and point of origin.[208] Hence, a clear linguistic division via spatial deixis markers and general orientation is encoded as in Eipo and Dene Chipewyan. This is clearly an indication of a high degree of specificity in spatial semantics.[209] Hopi separates this deictic space into a four-way matrix such as *ya-ng* 'here' (proximal), *a-* (medial), *e-p/pa-* 'there' (distal), and *ay* 'over there' (extreme-distal).[210] Note that the morpheme *da-* means something like close to the respective 'here', but not as far away as 'there', 'here and there' (cf. *deira, doro, doura* in Eipo).[211]

[204] Curtis 1976, 3.
[205] Curtis 1976, 5.
[206] Curtis 1976, 3.
[207] Thiering 2006.
[208] Malotki 1979, 23,84.
[209] Thiering 2013.
[210] Malotki 1979, 27, 59, 145.
[211] Malotki 1983, 16.

> Central to any analysis of spatial configuration are the linguistic coordinates that dissect the area taken up by the speaker (first person), the hearer (second person), and the persons or things other than the speaker and hearer (third person). English basically structures the terrain occupied by these entities into 'here' and 'there'. Formally adverbs, the semantic thrust of 'here' and 'there' is deictic, with 'here' indicating a point in the immediate vicinity of the speaker and 'there' selecting one further removed from him.

A more detailed account of Dene will reveal even more about the interaction between environmental landmarks and its representation in language as we have just shown above in Eipo. The following examples present very basic directional locative markers in Dene.

| | |
|---|---|
| *(ne)ja* | 'here' |
| *?eyer* | 'there' |
| *yughé* | 'over there' |
| *ekozi* | 'near there' |
| *hoch'a zi* | 'away from there/it (time, place)' |
| *-k'ezi* | 'over'; 'out on' (lake, hill, prairie, flat surface) |
| *nizi* | 'in presence of' (close proximity) |
| *yuwé nigha* | 'go (over there)' (verb) 'You go over there.' |
| *-thethe* | 'above', 'over' |
| *nadaghe* | 'in front of' |
| *náhésja* | 'go' (start across) 'I started across' |
| *náhédel* | 'go' (start across) 'They (plural) started across.' |
| *náhélgé* | 'go' (start across) (animal) 'He has started across.' |
| *nalé* | 'in sight of' (person, at a distance) |
| *nidhá* | 'far'; 'It is far.' |
| *nidháíle* | 'near', 'close by' |
| *nu tedhe* | 'over us' (dual and plural) |
| *-thethe* | 'above', 'over' |
| *ho tedhe* | 'unspecified area' |
| *be tedhe* | 'person'; 'thing over a person or something' |
| *se tedhe* | 'over me'; 'above', 'over my head', (metaphorically) 'I do not understand.' |
| *nu tedhe* | 'over us' (dual and plural) |
| *ni dúe* | 'standing close together' |
| *-gáh* | (literally) 'close', 'near' |
| *hube tedhe* | 'over them' (plural) |
| *t'ázi* | 'behind' ('going the other way'); 'leaning against'; *ne-t'azi* 'behind your back' |
| *tanizi* | 'center', 'middle' |
| *tajáhai* | 'in the middle of the lake.' |
| *t'abábel* | 'near the shore line' |

Table 2.11: Basic directional locatives in Dene Chipewyan

These selected examples indicate that Dene Chipewyan (and also Hopi) exhausts a large range of spatial concepts, depending also on environmental landmarks, e.g., lakes in these examples. Additionally, distances are specified, as mentioned before, in a threefold system encoding proximal, medial, and distal relationships between the figure and ground. Those are only approximate distances not relying on exact geometrical or mathematical concepts.

Beside these obvious spatial concepts profiling certain spatial configurations, the next data set presents a case that focuses on truly environmental landmarks. An initial word count in the Elford dictionary of the noun 'water' and related constructions presents 199 hits for water alone. The aggregate 'ice' yields about 70 hits.[212]

| | |
|---|---|
| *ten* | 'ice' |
| *ten deteni* | 'ice' (thick) (noun/verb) 'The ice is thick.' |
| *ten déch'el* | 'cracked ice' (verb) 'The ice is cracked (with one big crack).' |
| *ten dzíré líi* | 'ice' (drifting) (noun) |
| *ten elt't'aghidzeghi* | 'iceberg' (noun) |
| *ten héltál* | 'cracked ice' (verb) 'The ice is cracked (with one small crack).' |
| *ten hóeni* | 'dangerous' (verb) 'The ice is dangerous.' |
| *ten húlár* | 'float' (verb) 'Ice floated past.' *ten nádhilteni* 'icicle' (noun) |
| *ten nádénitthel* | 'chop ice (to carve a way)' (verb) 'He chopped ice away.' |
| *ten nágheltal* | 'crack (ice)' (verb) 'The ice is cracked (with many small cracks).' |
| *ten náthelá* | 'float' (verb) 'Ice lifted or floated up.' |
| *ten nithelár* | 'float' (verb) 'Ice (large pan) floated to shore and out again.' |
| *ten táthedzegh* | 'float' (verb) 'Ice floated to shore.' |
| *táthela; ten táthelar* | 'float' (verb) 'Ice (large pan) floated to shore.' |
| *ten táthelár* | 'float' (verb) 'Ice lifted or floated up.' |
| *ten táthi* | 'float' (verb) 'Ice is floating (to shore).' |
| *ten ts'et'ani* | 'ice (thin)' (noun) not accessible |
| *ten tsele* | 'ice (fall)' (noun) not accessible |
| *ten ts'íli* | 'ice (spring)' (noun) not accessible |

Table 2.12: Variation of 'ice' in Dene Chipewyan

The above set of examples of various linguistic constructions encoding different qualities of 'ice' neatly complements the Eipo data on 'river' as an important landmark. During the fishing season, the Dene needed to know the specific qualities of ice, e.g., its thickness. Ice fishing necessitated the exact knowledge of a location where the ice was thin enough to drill a hole and which was at the same time above the fish grounds. Note that in Dene most of the above-quoted linguistic constructions are, nowadays, used only by a few fluent elders. It can be assumed that in a generation from now, most of the constructions will be gone.[213]

[212] See also 'river' = 22, 'lake' = 31 (as opposed to 'mountain' = 3), 'land' = 37; 'shore' = 6; 'fish', 'fishing' = 106.
[213] Thiering 2009a, Thiering 2010.

## 2.7 Conclusion

This chapter has presented aspects of spatial cognition reflected in two unrelated languages and cultures. The interrelation of culture, environment, and language has been shown for Dene Chipewyan and Eipo. Some aspects of spatial cognition turned out to be culture specific, being shaped, for instance, by practices of spatial orientation and organization. Thus, depending on the practical and environmental contexts, we found differing degrees of specificity in the different cultures.

Arguably, language here plays a double role as an external representation or semiotic system, on the one hand throwing light on structures of cognition and on the other shaping cognition and influencing its structure. On the basis of the study of a sample of two unrelated cultures and utterances in their languages, the chapter attempted to distinguish aspects of spatial cognition. Some might be candidates for universals although they may find different expressions in different languages. It is impossible, of course, to draw inferences from a sample of just two languages, but in the broader context of this book, it appears obvious that certain non-linguistic features of spatial thinking shape spatial language and practice in all societies. An example is the figure-ground asymmetry as a fundamental structure of spatial cognition.[214]

Aspects of spatial topography have been shown that are truly culture specific in the sense that different cultures develop different cognitive structures. Examples have been provided by deixis and other references to and conceptualizations of space. Moreover, the current chapter presented cultural and language specific *ideas of space* of Eipo and Dene Chipewyan (and some selected from other languages such as Hopi). Such spatial concepts have been shown to be of crucial importance in the two ethnic groups and related cousin languages. People in both cultures lived in complex environments, traveled long distances into dangerous terrain and usually made their way back safely. Survival in their habitats depended on evolved capacities typical for our species to efficiently manage orientation in space. Moreover, it depended on ontogenetic learning about the geography of the environment with its many specific features and on a culturally transmitted, linguistically encoded spatial reference system sufficiently precise to foster the process of forming mental maps of their land.

We have provided linguistic information about the encoding of such spatial concepts. These concepts are topography-based and related to environmental landmarks. These landmarks can be mountains or rivers and lakes. The concepts are also based on one's own experience when walking to and returning from various distant places. These individual experiences are made in the context of culture-specific practices and techniques which therefore shape the spatial concepts. Examples of such practices and techniques are the making of gardens, hunting and snaring in high altitude and the partly ritualized process of building a men's house in the society of the Eipo or hunting in the society of the Dene. These practices embody culturally shared knowledge and, at the same time, reflect the environmental affordances.

As for other cultures, spatial classification in Eipo and Dene involves locating the objects, i.e., defining places is basically deliminating, based on the environment. Speakers parse up their environment into an important and necessary topography or spatial matrix. This is represented in the language via mountain, river, and place names. The description

[214] See, however, Thiering 2011.

of such components, as Malotki rightfully points out, should include anthropological and cultural aspects of the language.[215] One of the main empirical sources we have for Eipo and Dene are the oral traditions as transmitted in their myths. These myths function as a chronological topology of places.

Certain practices, habits, and environmental landmarks clearly have repercussions on language (as shown in some selected linguistic examples). Hence, our research on Amerindian and Mek languages corroborates some insights from early nineteenth and twentieth-century scholars such as Franz Boas, Edward Sapir, and Benjamin Lee Whorf (and contemporary scholars such as Helmut Gipper and Ekkehart Malotki). Those insights were built on Humboldt's idea of *Weltansichten* 'world perspectives', i.e., the idea that the structure of language influences the thought process. In North America, this concept is known as the linguistic relativity principle or Sapir-Whorf theory. We subscribe to the idea that languages differ in the way they shape our world perspectives, but believe that non-linguistic information has its impact upon spatial language and categorization. Hence, our current research aimed to show the ideas of space ('Raumbilder') as a web of intertwined interaction of language, culture, and cognition.

The following quote by Heeschen summarizes the impact of non-linguistic, e.g., environmental, cultural etc., information on language, in this case the Mek language.[216]

> The importance of reference to space, the social context of giving and taking, and references to non-verbal communication shape the content of the vocabulary. The characteristics and peculiarities of everyday interaction and speech follow from the fact that speech is complemented by, and related to, other semiotic systems.

We subscribe to Heeschen's point of view with respect to the reference to space and its relation to semiotic systems. We have presented language data showing the influence and constructive process of environmental landmarks and cultural heritage on shaping of spatial categorization in the two languages.

Finally, we hope we have shown that spatial knowledge is embedded in cultural and linguistic practices. This was outlined above as our guiding principle, i.e., that spatial knowledge is not only encoded in mental concepts, but also embodied in the lived histories of human beings. These histories are represented by cultural and linguistic practices. Hence, our concept presented at the beginning of this chapter arguing in favor of an influence of non-linguistic information upon spatial language and categorization has been shown to apply. The points taken from the selected empirical data indeed indicate the influence and even constructive process of environmental landmarks and cultural heritage on the shaping of spatial categorization in the two languages.

[215] Malotki 1979, 301.
[216] Heeschen 1998, 381.

**List of linguistic abbreviations**

| | | | |
|---|---|---|---|
| ADV | adverb | MOM | momentaneous |
| CLASS | classifier | PKT | punctive |
| DIF | diffusive | PL | plural |
| DUR | durative | PP | post position |
| FUT | future | PRES | present |
| HAB | habitual | PRON | pronoun |
| HORT | hortative or optative | S | subject |
| INCORP | incorporation | SG | singular |
| ITER | iterative | VN | verbal noun |
| MED | sentence medial verb | | |

Table 2.13: List of linguistic abbreviations used in this chapter

**Bibliography**


Ameka, F. (2006). Elements of the grammar of space in Ewe. In: *Grammars of Space*. Ed. by D. Wilkins S. Levinson. Cambridge: Cambridge University Press, 359–399.

Bechtel, William and Adele Abrahamsen (1991). *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*. Cambridge, MA: Blackwell.

Berthele, Raphael (2006). *Ort und Weg. Die sprachlichen Raumreferenzen in Varietäten des Deutschen, Rätoromanischen und Französischen*. Berlin: de Gruyter.

Birbaumer, Niels and Robert Schmidt, eds. (1993). *Lernen und Gedächtnis. Neuro- und Sinnesphysiologie*. Springer.

Blum, Paul, Irenäus Eibl-Eibesfeldt, Volker Heeschen, Klaus Helfrich, Gerd Koch, Thomas Michel, Grete Schiefenhövel, Wulf Schiefenhövel, Artur Simon, Franz Simon, and Sven Walter (1979–1996). *Eipo (West-Neuguinea, Zentrales Hochland)*. DVD, IWF Wissen und Medien, Göttingen.

Boas, Franz (1977). *Handbook of American Indian Languages, Part I*. reprint of the 1911 edition. London: Routledge/Thoemmes Press.

Bodley, John H. (1999). Hunter-gatherers and the colonial encounter. In: *The Cambridge Encyclopedia of Hunters and Gatherers*. Ed. by Richard B. Lee and Richard Daly. Cambridge: Cambridge University Press, 465–472.

Brody, Hugh (1982). *Maps and Dreams*. New York: Pantheon.

Brown, Cecil H. (1983). Where Do Cardinal Direction Terms Come From? *Anthropological Linguistics* 25(2):121–161.

Buschmann, J. C. E. (1855). Der athapaskische Sprachstamm. In: *Königliche Akademie der Wissenschaften zu Berlin, Abhandlungen aus dem Jahre 1855*, 144–319.

Bußmann, Hadumod (2008). *Lexikon der Sprachwissenschaft*. Stuttgart: Kröner.

Carlson, Laura (1999). Selecting a reference frame. *Spatial Cognition and Computation* 1(4): 365–379.

— (2000). Object use and object location. The effect of function on spatial relations. In: *Cognitive Interfaces Constraints On Linking Cognitive Inform*. Ed. by Emile Van Der Zee and Urpo Nikanne. Oxford: Oxford University Press, 94–115.

— (2003). Using spatial language. In: *Psychology of Motivation: Advances in Research and Theory*. Ed. by Aaron S. Benjamin and Brian H. Ross. San Diego, CA: Academic Press, 127–161.

Carlson, Laura and Gordon Logan (2001). Using spatial terms to select an object. *Memory and Cognition* 29:883–892.

Carlson-Radvansky, Laura (1993). Frames of reference in vision and language: Where is above? *Cognition* 46:223–244.

Carlson-Radvansky, Laura and Gabriel Carlson-Radvansky (1996). The influence of functional relations on spatial term selection. *Psychological Science* 7:56–60.

Carter, Robin M. (1976). Chipewyan Classificatory Verbs. *International Journal of American Linguistics* 42(1):24–30.

Cook, Eung-Do (2004a). *A Grammar of Dene Suline (Chipewyan)*. Algonquian and Iroquoian Linguistics. Winnepeg, Manitoba: Algonquian and Irquoian Linguistics.

— (2004b). Athapaskan classificatory verbs. *Amerindia, Revue d'Ethno-linguistique amérindienne*( 11):11–24.

Coventry, Kenny R. and Simon C. Garrod (2004). *Saying, Seeing and Acting. The Psychological Semantics of Spatial Prepositions*. Hove/New York: Psychology Press.

Curtis, Edward S. (1976). *The North American Indian. Vol. 18, The Chipewyan. The Western Woods Cree. The Sarsi.* reprint. New York: Johnson.

Davidson, William L., William Elford, and Harry Hoijer (1963). Athapaskan classificatory verbs. In: *Studies in the Athapaskan Languages*. Ed. by Harry Hoijer et al. Berkeley: University of California Press, 30–41.

Downs, Roger and David Stea, eds. (1973). *Image and Environment: Cognitive Mapping and Spatial Behavior*. Chicago: Aldine.

Edelman, Shimon (2002). Constraining the neural representation of the visual world. *Trends in Cognitive Science* 6:125–131.

Eibl-Eibesfeldt, Irenäus (1995). *Eipo (West-Neuguinea, Zentrales Hochland). Männertanz 'sang mote' als Kinderspiel*. DVD. Produktionsjahr: 1979 IWF Bestellnumer/Bandzählung: E 2686.

Eibl-Eibesfeldt, Irenäus, Wulf Schiefenhövel, and Volker Heeschen (1991). Film- und Tonaufnahmen bei den Eipo, Neuguinea. Zur Methodik der Datenerhebung und Auswertung. *Wissenschaftlicher Film* 42.

Foley, William A. (1986). *The Papuan Languages of New Guinea*. New York: Cambridge University Press.

— (1997). *Anthropological Linguistics: An Introduction*. Oxford: Blackwell.

Fowler, Catherine and Nancy Turner (1999). Ecological/Cosmological Knowledge and Land Management among Hunter-Gatherers. In: *The Cambridge Encyclopedia of Hunters and Gatherers*. Ed. by Richard B. Lee and Richard Daly. Cambridge: Cambridge University Press, 419–425.

Geertz, Clifford (1999). *The Interpretation of Cultures. Selected Essays*. New York: Basic Books.

Gentner, Dedre and Albert Stevens, eds. (1983). *Mental Models*. Hillsdale: Erlbaum.

Gladwin, Thomas (1974). *East is a Big Bird. Navigation and Logic on Puluwat Atoll*. Cambridge, MA: Harvard University Press.

Heeschen, Volker (1990). *Ninye bún. Mythen, Erzählungen, Lieder und Märchen der Eipo (im zentralen Bergland von Irian Jaya, West-Neuguinea, Indonesien)*. Mensch, Kultur und Umwelt im Zentralen Bergland von West-Neuguinea 20. Berlin: Reimer.

— (1992). The Position of the Mek Languages of Irian Jaya among the Papuan Languages: History, Typology, and Speech. *Bijdragen tot de Taal-, Land- en Volkenkunde* 148(3): 465–488.

— (1998). *An Ethnographic Grammar of the Eipo Language. Spoken in the Central Mountains of Irian Jaya (West New Guinea), Indonesia*. Berlin: Reimer.

Heeschen, Volker and Wulf Schiefenhövel (1983). *Wörterbuch der Eipo-Sprache. Eipo-Deutsch-Englisch*. Mensch, Kultur und Umwelt im zentralen Bergland von West-Neuguinea 6. Berlin: Reimer.

Hillert, Dieter (1987). *Zur mentalen Repräsentation von Wortbedeutungen. Neuro- und psycholinguistische Überlegungen*. Tübingen: Narr.

— (1992). *Sprachprozesse und Wissensstrukturen*. Opladen: Westdeutscher Verlag.

Hofstadter, Douglas R. (1980). *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Vintage.

Hoijer, Harry (1946). Introduction. In: *Linguistic Structures of Native America*. Ed. by Cornelius Osgood. Viking Fund Publications in Anthropology 6. New York: Viking Fund, 9–29.

— (1951). Cultural implications of some Navaho linguistic categories. *Language* 27(2): 111–120.

Hunt, Earl and Franca Agnoli (1991). The Whorfian hypothesis. A cognitive psychology perspective. *Psychological Review* 98(3):377–389.

Hutchins, Edwin (1983). Understanding Micronesian Navigation. In: *Mental Models*. Ed. by Dedre Gentner and Albert L. Stevens. Hillsdale: Erlbaum, 191–225.

— (1996). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Johnson-Laird, Philip Nicholas (1983). *Mental Models. Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, MA: Harvard University Press.

Kandel, Eric and Richard Hawkins (1994). Molekulare Grundlagen des Lernens. In: *Gehirn und Bewußtsein*. Heidelberg: Spektrum, 114–124.

Kari, James (1979). *Athabaskan Verb Theme Categories: Ahtna*. Alaska Native Language Center Research Papers 2. Fairbanks: University of Alaska.

— (2011). A case study in Ahtna Athabascan geographic knowledge. In: *Landscape in Language: Transdisciplinary Perspectives*. Ed. by David M. Mark, Andrew J. Turk, Niclas Burenhult, and David Stea. Amsterdam: Benjamins.

King, David and M. Bathgate (s.d.). Population of Papua New Guinea. In: *Papua New Guinea Atlas: A Nation in Transition*. Ed. by David King and Stephen Ranck. R. Brown and Associates in conjunction with the University of Papua New Guinea, 20–21.

Knauff, Markus (2013). *Space to Reason: A Spatial Theory of Human Thought*. Cambridge, MA: MIT Press.

Koch, Gerd (1984). *Maligdam. Ethnographische Notizen über einen Siedlungsbereich im oberen Eipomek-Tal, zentrales Bergland von Irian Jaya, West-Neuguinea, Indonesien.* Mensch, Kultur und Umwelt im Zentralen Bergland von West-Neuguinea 15. Berlin: Reimer.

Koch, Gerd and Wulf Schiefenhövel (2009). *Eipo (West-Neuguinea, Zentrales Hochland) - Neubau des sakralen Männerhauses in Munggona*. DVD. Produktionsjahr: 1974, first published in 1979; IWF Bestellnummer/Bandzählung E 2475.

Koffka, Kurt (1935). *Principles of Gestalt Psychology*. New York: Harcourt.

Komando Daerah Militer XVII "Cenderawasih" (1969). *Laporan expedisi ilmiah lembah-X, 2 Okt. sampai dengan 20 Des. 1969 di Irian Barat. [Report on the scientific expedition into the Valley of the X-River from 2 October 20 December 1969 in West-Irian]*. Mimeograph.

Kuipers, B. (1982). The "map in the head" metaphor. *Environment and Behaviour* 14:202–220.

Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar. Vol. I: Theoretical Prerequisites*. Stanford: Stanford University Press.

— (2008). *Cognitive Grammar: A Basic Introduction*. New York: Oxford University Press.

Leer, Jeff (1989). Directional systems in Athapaskan and Na-Dene. In: *Athapaskan Linguistics. Current Perspectives on a Language Family*. Ed. by Eung-Do Cook and Keren Rice. Berlin: Mouton de Gruyter, 575–622.

Levinson, Stephen C. (2003). *Space in Language and Cognition. Exploration in Cognitive Diversity*. Cambridge: Cambridge University Press.

Levinson, Stephen C. and David Wilkins, eds. (2006). *Grammars of Space*. Cambridge: Cambridge University Press.

Lewin, Kurt (1936). *Principles of Topological Psychology*. London. McGraw Hill.

Li, Fang-Kuei (1946). Chipewyan. In: *Linguistic Structures of Native America*. Ed. by Cornelius Osgood. Viking Fund Publications in Anthropology 6. New York: Viking Fund, 398–423.

Louwerse, Jan (1978). A tentative Una phonology. *Irian* 7(3):43–90.

— (1988). *The Morphosyntax of Una in Relation to Discourse Structure: A Descriptive Analysis*. Pacific Linguistics 100. Canberra: Australian National University.

Lucy, John A. (1992a). *Grammatical Categories and Cognition: A Case Study of the Linguistic Relativity Hypothesis*. Studies in the social and cultural foundations of language 13. Cambridge: Cambridge University Press.

— (1992b). *Language Diversity and Thought: A Reformulation of the Linguistic Relativity Hypothesis*. Studies in the social and cultural foundations of language 12. Cambridge: Cambridge University Press.

Malotki, Ekkehart (1979). *Hopi-Raum: Eine sprachwissenschaftliche Analyse der Raumvorstellungen in der Hopi-Sprache*. Tübingen: Narr.

— (1983). *Hopi Time. A Linguistic Analysis of the Tempral Concepts in the Hopi Language*. Berlin: Mouton de Gruyter.

Mark, David M., Andrew J. Turk, Niclas Burenhult, and David Stea, eds. (2011). *Landscape in Language: Transdisciplinary Perspectives*. Amsterdam: Benjamins.

Marr, David (1982). *Vision: A Computational Investigation in the Human Representation of Visual Information*. San Francisco: Freeman.

McDonough, Joyce (2000). On a bipartite model of the Athabaskan verb. In: *The Athabascan Languages*. Ed. by Theodore B. Fernald and Paul R. Platero. Oxford: Oxford University Press, 139–166.

Michel, Thomas (1983). *Interdependenz von Wirtschaft und Umwelt in der Eipo-Kultur von Moknerkon. Bedingungen für Produktion und Reproduktion bei einer Dorfschaft im*

*zentralen Bergland von Iran Jaya, West-Neugiunea, Indonesien*. Mensch, Kultur und Umwelt im zentralen Bergland von West-Neuguinea 11. Berlin: Reimer.
Miller, George A. and Philip Nicholas Johnson-Laird (1976). *Language and Perception*. Cambridge: Cambridge University Press.
Morice, A. G. (1890). The Déné Languages (Considered in Themselves and Incidentall in their Relations to Non-Amercan idioms). In: *Transaction of the Canadian Institute*. 1, 170–212.
Murre, Jacob M.J. and Rainer Goebel (1996). Connectionist modeling. In: *Computational Psycholinguistics*. Ed. by Tom Dijkstra and Konrad De Smedt. London: Taylor and Francis, 49–81.
Nunberg, Geoffrey (1978). *The Pragmatics of Reference*. Bloomington: Indiana University Linguistics Club.
Penrose, Roger (1991). *Computerdenken: Die Debatte um künstliche Intelligenz, Bewußtsein und die Gesetze der Physik*. Heidelberg: Spektrum.
Piaget, Jean and Bärbel Inhelder (1956). *The Child's Conception of Space*. London: Routledge & Kegan Paul.
Portugali, Juval, ed. (1996). *The Construction of Cognitive Maps*. Dordrecht: Kluwer.
Pouwer, J. (1964). Social System in the Star Mountains, towards a Reorientation of the Study of Social Systems. *American Anthropologist* 66(4):133–161.
Rice, Keren (1989). *A Grammar of Slave*. Berlin: Mouton de Gruyter.
Rice, Sally (1996). Prepositional prototypes. In: *The Construal of Space in Language and Thought*. Ed. by Martin Pütz. Berlin: Mouton de Gruyter, 135–165.
— (1997). Giving and taking in Chipewyan: The semantics of THING-marking classificatory verbs. In: *The Linguistics of Giving*. Ed. by John Newman. Philadelphia: Benjamins, 97–134.
— (2002). Posture and existence predicates in Dene Suline (Chipewyan): Lexical and semantic density as a function of the 'stand'/'sit'/'lie' continuum. In: *The Linguistics of Sitting, Standing, and Lying*. Ed. by John Newman. Philadelphia: Benjamins, 61–78.
Ritter, H., T. Martinetz, and K. Schulten (1991). *Neuronale Netze. Eine Einführung in die Neuroinformatik selbstorganisierender Netzwerke.* Bonn: Addison-Wesley.
Röll, Werner and Gerd Zimmermann (1979). *Untersuchungen zur Bevölkerungs-, Siedlungs- und Agrarstruktur im zentralen Bergland von Irian-Jaya, West-Neuguinea, Indonesien*. Mensch, Kultur und Umwelt im zentralen Bergland von West-Neuguinea 1. Berlin: Reimer.
Rubin, Edgar (1921). *Visuell wahrgenommene Figuren*. Copenhagen: Gyldendalske.
Sapir, Edward (1915). The Na-Dene Languages, a Preliminary Report. *American Anthropology* 17(3):534–558.
Sarfert, E. (1911). Zur Kenntnis der Schiffahrtskunde der Karoliner. *Korrenspondenzblatt der Deutschen Gesellschaft für Anthropologie, Ethnologie, und Urgeschichte* 42:131–136.
Schade, Ulrich (1992). *Konnektionismus: Zur Modellierung der Sprachproduktion*. Opladen: Westdeutscher Verlag.
Schiefenhövel, Wulf (1976). Die Eipo-Leute des Berglands von Indonesisch-Neuguinea: Kurzer Überblick über den Lebensraum und seine Menschen. Einführung zu den Eipo-Filmen des Humanethologischen Filmarchivs der Max-Planck-Gesellschaft. *Homo* 26(4):263–275.

— (1979). The Eipo: Members of the Mek group in the Highlands of Irian Jaya. *Irian, Bulletin of Irian Jaya Development* VII(2):47–67.
— (1991). Die Mek und ihre Nachbarn. Internationale Konferenz über Neuguinea-Forschungen. *Zeitschrift für Wissenschaft und Forschung* 91(1):21–25.
— (2001). Kampf, Krieg und Versöhnung bei den Eipo im Bergland von West-Neuguinea: Zur Evolutionsbiologie und Kulturanthropologie aggressiven Verhaltens. In: *Begegnung und Konflikt eine kulturanthropologische Bestandsaufnahme*. Ed. by Wolfgang Fikentscher. Bayrische Akademie der Wissenschaften, Philologisch-Historische Klasse, Abhandlungen 120. München: Beck, 169–186.
Schreuder, Robert and Giovanni B. Flores d'Arcais (1989). Psycholinguistic issues in the lexical representation of meaning. In: *Lexical Representation and Process*. Ed. by William D. Marslen-Wilson. Cambridge, MA: MIT Press, 409–436.
Senft, Gunter, ed. (2000). *Systems of Nominal Classification*. Cambridge: Cambridge University Press.
Siegel, Alexander W. and Sheldon H. White (1975). The Development of Spatial Representation of Large-Scale Environments. In: *Advances in Child Development and Behavior*. Ed. by Hayne W. Reese. New York: Academic Press.
Simon, Franz and Wulf Schiefenhövel (1989). *Eipo (West-Neuguinea, Zentrales Hochland). 'Mote', ein Besuchsfest in Munggona*. DVD. Produktionsjahr: 1976 IWF Bestellnumer/ Bandzählung: E 2803.
Sinha, Chris and Tanya Kuteva (1995). Distributed spatial semantics. *Nordic Journal of Linguistics* 18:167–199.
Smith, James G. E. (1981). Chipewyan. In: *Handbook of North American Indians. Vol. 6: Subarctic*. Ed. by June Helm. Washington: Smithsonian Institution, 271–284.
Strube, Gerhard, ed. (1996). *Wörterbuch der Kognitionswissenschaft.* Stuttgart: Klett-Cotta.
Svorou, Soteria (1993). *The Grammar of Space*. Philadelphia: Benjamins.
Swadling, Pemela (1981). *Papua New Guinea's Prehistory*. Port Moresby: Gordon & Gotch.
Sweetser, Eve (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge University Press.
Talmy, Leonard (1978). Figure and Ground in complex sentences. In: *Universals of Human Language*. Ed. by Joseph Harold Greenberg, Charles A. Ferguson, and Edith A. Moravcsik. Stanford: Stanford University Press, 627–649.
— (1983). How to structure space. In: *Spatial Orientation: Theory, Research, and Application*. Ed. by Herbert Pick and Linda Acredolo. New York: Plenum Press, 225–282.
— (2000). *Towards a Cognitive Semantics, Vols. I and II*. Cambridge, MA: MIT Press.
Thiering, Martin (2006). *Topological Relations in an Athapaskan Language*. Papers in Experimental and Theoretical Linguistics (Department of Linguistics Working Papers) 1. Edmonton: University of Alberta.
— (2009a). Language loss in spatial semantics: Dene Suliné. In: *Variation in Indigenous Minority Languages*. Ed. by James N. Stanford and Dennis R. Preston. Amsterdam: Benjamins, 485–516.
— (2009b). *Linguistic Categorization of Topological Spatial Relations*. Preprint 373, Max Planck Institute for the History of Science, Berlin.
— (2010). *Intralingual Variation of Spatial Concepts in an Athapaskan Language*. LAUD Paper No. 742, Linguistic Agency University of Duisburg-Essen.
— (2011). Figure-Ground Reversals in Language. *Gestalt Theory* 33(3/4):245–276.

Thiering, Martin (2012). *Topographical Coordinates and Spatial Language*. LAUD Paper No.770, Linguistic Agency University of Duisburg-Essen.

— (2013). Degrees of specificity in spatial semantics. In: *Variation in Language and Language Use. Linguistic, Socio-Cultural and Cognitive Perspectives*. Ed. by Monika Reif, Justyna A. Robinson, and Martin Pütz. Frankfurt a. M.: Peter Lang, 367–420.

— (2014). Cognitive Maps of Landmark Orientation. In: *Multilingual Cognition and Language Use: Processing and typological perspectives*. Ed. by Luna Filipović and Martin Pütz. Amsterdam: Benjamins, 151–182.

Thornton, Thomas F. (2011). Language and Landscape among the Tlingit. In: *Landscape in Language: Transdisciplinary Perspectives*. Ed. by David M. Mark, Andrew J. Turk, Niclas Burenhult, and David Stea. Amsterdam: Benjamins.

Tolman, Edward C. (1948). Cognitive Maps in Rats and Men. *The Psychological Review* 55(4):189–208.

Watzlawick, Paul (1981). *Die erfundene Wirklichkeit: Wie wissen wir, was wir zu wissen glauben? Beiträge zum Konstruktivismus*. München: Piper.

Wender, Karl Friedrich (1980). *Modelle des menschlichen Gedächtnisses.* Stuttgart: Kohlhammer.

Whorf, Benjamin Lee (1956). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.

Wurm, Stephen A. (1982). *Papuan Languages of Oceania*. Tübingen: Narr.

Wygotski, Lev S. (1964). *Denken und Sprechen*. Berlin: Akademie Verlag.

Young, Robert and William Morgan (1987). *The Navajo Language: A Grammar and Colloquial Dictionary*. Albuquerque: University of New Mexico Press.

— (1992). *Analytical Lexicon of Navajo*. Albuquerque: University of New Mexico Press.

Zell, Andreas (1994). *Simulation neuronaler Netze*. Bonn: Addison-Wesley.

Zlatev, Jordan (2003). Holistic spatial semantics of Thai. In: *Cognitive Linguistics and Non-Indo-European Languages*. Ed. by Eugene H. Cassad and Gary B. Palmer. Berlin: de Gruyter, 305–336.

# Chapter 3
# The Impact of Notation Systems: From the Practical Knowledge of Surveyors to Babylonian Geometry
*Peter Damerow*

## 3.1 Introduction

Knowledge about the early development of human cognition can be gained from two sources:

1. Archaeologically excavated artifacts can be interpreted by elaborating the minimal cognitive preconditions that made the invention and use of these artifacts possible.
2. If archaeological findings suggest that social and economic settings in an early human community are similar to those known from indigenous peoples untouched by modernization processes, it can be concluded that their members possessed similar cognitive abilities required to perform similar activities.

For the conditions under which prehistoric humans lived, abilities reconstructed by methods of this type are, as a rule, perfectly sufficient to master the problems to which they were exposed in their environment. Compared to later historic developments, however, such abilities show characteristic differences.

Max Wertheimer showed that indigenous cultures do not share our context-independent concept of number, but solve problems involving the judgement of quantities with context-specific mental constructions he called *Zahlgebilde*.[1] Counting sequences, if there are any, remain rudimentary. They are often linked to the specific classes of objects to which they are applied. No elaborated instruments for measuring sizes are used. Consequently, the languages of non-literate peoples lack words designating abstract concepts of number and size.

The situation is similar in the case of the concepts of time and space. Norbert Elias has argued convincingly that the metric concept of time is a product of modern economy and technology and neither required nor functional in premodern societies.[2] The classical concept of space, and more so the concept of space as an aspect of the integrated concept of spacetime in contemporary non-classical physics, are also dependent on conditions provided by modern science and technology. By contrast, the previous chapter has shown how tasks involving abilities of spatial orientation were solved among non-literate peoples. The development of spatial knowledge in these communities is linked closely to external knowledge representations, in particular to coordinated actions, landmarks for orientation, and lexical and grammatical properties of spoken language that were used to express spatial

[1] Wertheimer 1925.
[2] Elias 1984.

relations. Such abilities are not based on standardized measures representing spatial relationships quantitatively. No symbol systems exist which would allow for calculation techniques like those required, for instance, to determine the sizes of areas. The abilities remain context-specific and cannot be applied universally to spatial magnitudes and relations.

This raises the interesting question about the historical circumstances under which the nature of the conceptualization of space changed and what brought about these changes. Obviously, this development is somehow related to the invention of writing. Fortunately, in the case of what is possibly the earliest writing system, that is, cuneiform writing invented in Mesopotamia, an abundance of preserved sources from the time of its invention well into the time of Classical Antiquity provides information about the development of conceptual patterns and corresponding technologies and methods dealing with spatial relations.

The focus of the following analysis on this development will be concentrated on its early phase. The earliest sources that indicate innovations related to the conceptualization of spatial relations, in comparison to what is known from non-literate cultures, are administrative tools, in particular proto-cuneiform administrative tablets written between 3200 and 3000 BCE, which triggered the invention of writing. The analysis will cover the following time period of about 1500 years into the middle of the second millennium, when the so-called Babylonian mathematics was fully developed. It will be shown under which circumstances knowledge about spatial relations in this period developed from the basic forms known from non-literate societies into an esoteric art of formulating complex geometrical problems and solving them using sophisticated arithmetical tools applied to geometrical intuition. It will be shown that this development was not a direct consequence of the invention of writing. Rather, it has to be conceived as a coevolution of glottographic and non-glottographic symbol systems[3] ending up with the dissociation into writing, arithmetic, and geometry.

This development originated in the context of the bureaucratic administration which necessarily evolved together with the concentration of economic and military power in the early Mesopotamian cities and the emergence of socially stratified, centralized states in the time period from the late fourth to the late third millennium BCE. Measurement technologies had to be developed and externally represented in written or other symbolic form in order to control the acquisition and redistribution of unprecedented amounts of resources. The quantification of spatial relations by length, area, and volume measures, in particular, was based primarily on the growing knowledge of surveyors and the reflection on their means and practices. The resulting mental constructions remained implicit, but can partly be reconstructed from the arithmetical operations documented by administrative records and by Babylonian mathematics. One of the surprising results of the analysis will be that the outcome of this development differs from the geometry we know from the Euclidean tradition. The kind of geometry that is part of Babylonian mathematics shows 'non-Euclidean' peculiarities such as the neglect of the role of angles, which result from the practices of the surveyors they reflect. It will become clear that the Babylonian case demonstrates that there is no canonical path to a single universal geometry.[4]

[3] The term *glottographic* denotes the dependence on spoken language. For a detailed discussion of the classification of early writing systems and their relation to spoken language (phonology, linearity, etc.), see Hyman 2006.

[4] The present paper is heavily based on my inaugural lecture 1994 at the University of Konstanz and its extended publication (Damerow 2001).

## 3.2 The origin of notation systems in Mesopotamia in the third millennium BCE

Rural communities as we know them from indigenous cultures apply simple techniques of quantifying such as using the breadth of fingers or the length of the forearm to control productive activities of knitting, weaving, or braiding. Long distances may be distinguished by counting the days of travel. However such quantifications remain embedded in the context in which they were used and do not give rise to their integration into a comprehensive concept of length. Accordingly, the assessment of areas or capacities by some kind of estimation or tool-based quantification does not cause any mental construction of a three-dimensional space that integrates lengths, areas, and volumes. It is reasonable to assume that the rural communities of early Mesopotamia applied similar techniques of rudimentary quantification to those known from extant indigenous cultures.

In the fourth millennium at the latest the nomadic tribes and rural communities in Mesopotamia were complemented by larger settlements and even fortified cities.[5] This development was connected with unprecedented cognitive constructions. The first and foremost type of such constructions in the early phase of the development of new cognitive abilities triggered by administrative challenges was the integration of context-specific measures into comprehensive and standardized systems of measures represented by corresponding systems of symbolic notation.[6]

What is the evidence for such a cognitive development? Archeological finds from this period show that the new challenges of the situation required an extensive use of existing traditional methods to control quantities. This is indicated especially by the expansive use of certain geometrically shaped clay tokens whose functions were long unclear. Sealed spherical envelopes made of clay and containing combinations of such tokens finally made evident that the tokens were used as counters to document quantities of resources that were the subject of administrative transactions.[7] Some of these envelopes bear impressed markings on their surfaces that – in view of the type of impression – can be identified as precursors of later numerical notations. At about the same time, such markings were pressed into the surface of sealed clay tablets, known as numerical tablets, the appearance of which immediately predates the invention of proto-cuneiform writing with its developed numerical notation systems.[8]

A close analysis of these archaeological findings and the results of deciphering the earliest proto-cuneiform administrative tablets strongly suggest that the invention of writing was in fact coupled with the emergence of completely new cognitive constructions, among them incipient forms of numbers.

Many prehistoric numerical notations preceding proto-cuneiform and cuneiform numerical notations consist of a few vertically or obliquely impressed marks similar in shape and size. They are comparable to one or more series of notches or dots as known from much older artifacts such as bones, tools, or cave paintings. It is likely that such repetitions of signs represent the quantity of the objects represented by the individual signs. Other notations on the prehistoric numerical tablets show a more complex structure. They often consist of sev-

[5] For an example, see the excavation of Habuba Kabira (Strommenger 1980).
[6] Nissen, Damerow, and Englund 1993.
[7] Schmandt-Besserat 1992.
[8] For an example see the numerical tablets of Jebel Aruda written in the second half of the fourth millennium BCE (Driel 1982).

Figure 3.1: Prehistoric sealed numerical tablet from Jebel Aruda with four series of impressions, repeated according to the quantities they represent

eral series of impressions that differ in size and shape. Generally, the symbols are arranged symmetrically and placed as if they would represent a hierarchy of quantities or numerical units.

Without knowing the context which might explain the function of such repeated symbols, it is impossible to decide definitively whether they represent units in the framework of an already existing number concept, or still represent objects such as animals, containers, or pieces of cheese. While tablets that display more than one type of repeated signs show them arranged in a way that resembles the structure of later numerical notations, they nonetheless share properties with simple types of numerical impressions, suggesting that they still represent objects and not units of a general metrological or numerical system.[9]

The most conspicuous properties indicating a non-numerical meaning of the individual signs are:

1. The *low degree of standardization of their shapes*. These shapes differ from tablet to tablet. Since we do not know the context of their usage it is for the most part impossible to decide whether signs on different tablets are variants of the same sign or denote different units with different values. Their meaning seems to be determined by the context of their use rather than by internal relations within a formal system of numeration, which would it make possible to identify their meaning independent of any knowledge about the context of their use.
2. Their *unlimited repeatability*. The numerical signs are frequently repeated more than ten times, even if they are placed in the middle of a complex numerical notation. The numerical notations of the prehistoric numerical tablets obviously lack the typi-

[9]This is true, for instance, for several of the above mentioned numerical tablets from Jebel Aruda, published by Driel 1982.

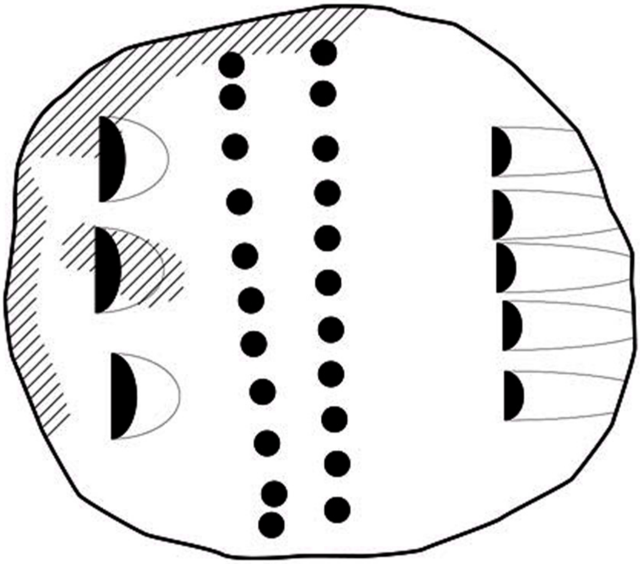

Figure 3.2: Numerical tablet from Jebel Aruda with a sign repetition exceeding any known relation between numerical or metrological units

cal bundling structure of all known notation systems of numbers. Even when repeated small units represent an amount greater than the next higher unit, they were nevertheless not converted. This makes it at the very least likely that these symbols represent real objects and not units of a numerical or metrological system.

These properties of prehistoric numerical notations strongly suggest that the numerical signs in fact still represent real units such as containers of different sizes, not values within a standardized system of context-independent measures. Their precise meaning was determined by the context in which they were used. This makes deciphering of the prehistoric numerical notations virtually impossible and, consequently, also the determination of the meaning of the combinations of clay tokens, which was transferred to the new medium of numerical tablets. In particular, it is impossible to find out whether prehistoric numerical notations existed which represented geometrical measures such as measures for length, area, and volume.

At the end of the fourth millennium, the situation changed radically. As a consequence of the invention of the proto-cuneiform writing system, the numerical notations were complemented with iconic graphs.[10] The shapes of the signs became largely standardized. With few exceptions, repetitions of small units were systematically converted into higher units as soon as the repeated units exceeded their size. Several numerical notations qualified by

[10]Usually called ideograms. On the precise meaning of the term *iconic graph*, see Boltz 2006. Later, when the cuneiform writing system had developed into writing in the proper sense, that is, a system representing spoken language, such ideograms developed into a non-syllabic notation within a syllabic writing system. The same signs are then rightly called logograms or Sumerograms (i.e., adapted from Sumerian writing).

iconic graphs were now written on one and the same tablet. Sequences of such entries were frequently complemented by their total, using the conversion procedure determined by the size relationships of the applied notation system.

The internal consistency of the application of these rules to the numerical notations made it possible to decipher their meaning in spite of the systematic ambiguity which is a consequence of a persisting context-dependency of the meaning of the numerical signs.[11] In fact, the numerical signs had no fixed numerical values. They were arranged in different, but internally coherent systems of notations with changing numerical values depending on the system in which they were used. These systems were applied consistently in specific areas of application. In order to facilitate the identification of the system, the numerical signs were partly modified by additional strokes or impressions produced with the tip of a stylus. Generally, however, the context indicated by the iconic graphs accompanying the numerical notations, and the extremely different numerical values of the signs when they were used in different systems, were sufficient to identify the notation system used. Thus, the relation of the sign representing the unit in the counting system to the next higher unit in this system changes between one to six, one to ten, and one to eighteen, depending on the context of application. The numerical signs frequently changed even the order of their numerical values. The signs with the values of 60 and 3,600 units in the counting system applied to grain measures, for example, assume the values 180 and 60 respectively. Thus the first sign, which represented $^1/_{60}$ of the second sign, assumes a value 3 times higher than the second sign.[12]

It is obvious that – in spite of this still existing context dependency of the numerical signs' values on the context of their application – the consistent notation systems of the administrative proto-cuneiform tablets document a new kind of cognitive construction different from those of their prehistoric precursors. They represent the earliest well-documented notation systems of measures that are independent of the actual measuring tools, constituted by internal relations between units which determine their sizes. Among them, systems can be identified which implicitly define the dimensions of space: length, area, and volume.

The structures of the proto-cuneiform systems of numerical notations allow for some inferences about their origins. Most of them consist of a core group of measures with numerical relations between them that lack a systematic pattern. At the upper and the lower ends, these irregular patterns tend to become systematic, based on the sexagesimal pattern of one of the two counting series also documented by the administrative proto-cuneiform tablets. This distinction within the systems suggests that the core of a notation system represents inherited prehistoric measures which in a first phase of development had to be brought into standardized numerical relations taking into account the relative sizes determined by traditional measuring tools. In a second phase the core notation system was artificially expanded in order to cover the new challenges of a central administration of goods and resources. Since such an expansion was no longer determined by existing measures, this structure could be defined much more systematically than the structure of the core of the system.

In the case of the system of length measures, the core, which consisted of measures such as the finger and the cubit, is known only from later cuneiform sources. The reason

[11] For the results of deciphering of the proto-cuneiform numerical systems and an overview of the limited results of earlier attempts, see Damerow and Englund 1987. For a comprehensive account of the social context, see Nissen, Damerow, and Englund 1993 and the part written by R. Englund in Bauer, Englund, and Krebernik 1998, 15–233.
[12] See the overview of proto-cuneiform sign systems in Nissen, Damerow, and Englund 1993, 28–29.

may be that these measures were unimportant for the administrative activities recorded on proto-cuneiform tablets. It seems that in the context of these activities length measures were used predominantly for surveying.

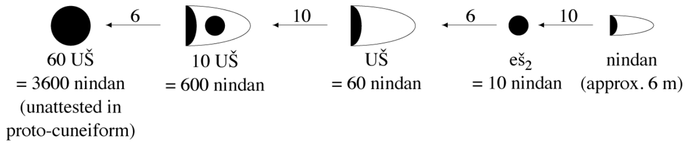


Figure 3.3: Relations between length measures on proto-cuneiform tablets

The length measures on the administrative proto-cuneiform tablets show a strict sexagesimal structure, suggesting that they were artificially created by officials of the central administration and played no role in prehistoric times. The basis of the system was the *nindan*, with a size of approximately six meters, probably derived from a measuring rod of the surveyors. This unit *nindan* was represented by the sign used in the sexagesimal counting system for the unit one. The higher units of the notation system for length measurements strictly follow the counting systems with signs for 10, 60, and 600, the last representing a length of about 3,600 meters. Contrary to what is known from the end of the third millennium BCE, such a distance still seems to extend beyond the field sizes at the beginning of the third millennium when the proto-cuneiform documents were written. At least, the size of 600 *nindan* is attested to on proto-cuneiform tablets only by so-called school texts.[13]

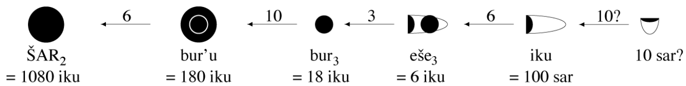


Figure 3.4: Relations between area measures on proto-cuneiform tablets

The situation is different in the case of area measures. The core notation system starts with a unit representing a size of about 0.36 hectares or 3,600 square meters, that is, 100 times of a unit later written with the cuneiform sign *sar*, designating a garden.[14] This basic unit of the core system of 100 *sar* was again written with the sign used in the sexagesimal counting system for the unit 1. It was later designated as *iku*, a term representing a field of arable soil. The unit *iku* was followed by a unit six times greater, *eše*$^3$, and a unit 18 times greater, *bur*$^3$. This core group of area notations was obviously derived from prehistoric measures of fields, the sizes corresponding to what one would expect from rural communities where the sizes of fields are related to the work to be invested and the yield required to support a family or a clan. With the creation of the area notation system, the measures forming the core group must have been redefined by the numbers that most closely approximate their originally imprecise relations to each other, resulting in the unsystematically organized core

[13] The number of proto-cuneiform tablets documenting length measures is relatively small. They contain length measures of fields up to approximately seven kilometers. See, for example, the school text discussed in Nissen, Damerow, and Englund 1993, 50.

[14] There is one tablet containing a sign possibly representing ten *sar*; see Nissen, Damerow, and Englund 1993, 57.

group of units. This core group of area notations was expanded, again following the structure of sexagesimal counting. Accordingly, the proto-cuneiform administrative tablets contain signs for 10 and 60 *bur*$^{3}$, which were artificial extensions of the core system to meet the requirements of controlling larger areas of arable land in a stratified society as compared to rural communities.

While the notation systems of length and area measures documented by proto-cuneiform administrative tablets were dependent on each other (as will be shown in the next section), notation systems of volume measures were still completely independent at that time. The reason is that they were used in completely different contexts, probably representing containers ranging from small ration bowls up to huge silos of about 260,000 liters.[15] The main area of application of this notation system was to represent amounts of different types of grain or liquid grain products. Different types of grain and grain products such as grain, barley groats, or malt were indicated by modifying the numerical signs.

The core notation system for volume measures consists of five signs with values ranging from 4.8 to 4,320 liters. Taking into account that, as a staple food, grain had to be stored for a period from one harvest to the next even in small-scale rural communities, it is plausible that all six of these measures go back to containers already used in prehistoric times. This core notation system was extended to higher units again by artificially adding sexagesimally structured signs, documented for ten and 60 times of the highest unit of the core system. The system was extended to smaller units by using unit fractions from $^1/_2$ to $^1/_6$, and probably also $^1/_{10}$ of the smallest unit of the core system.

The creation of coherent notation systems for length, area and volume was a major step towards a context-independent concept of space. While it is true that all three systems were still closely related to certain areas of application, that is, surveying in the case of length and area measures, and storage facilities in the case of volume measures, they nevertheless changed the direction of the determination. The handling of real objects no longer determined the methods of their quantification, and thus the perception of the measures. Instead, the notation system determined how the objects had to be quantified, thereby turning them into realizations of amounts determined by the notation system as a tool of administrative control.

### 3.3 The problem of multiplication and of the calculation of areas of fields

In addition to the creation of coherent notation systems there was another, even more influential innovation made possible by the creation of these systems. Most of these systems were constructed independently, based only on the different quantities of the same type of measure. However, in the case of length and area measures it seems that they depended on length measures from the very beginning. This connection must have influenced the construction of the notation system for areas and made it dependent on the system of length measures. This dependency probably reflects the fact that the sizes of fields were determined by surveyors on the basis of length measurements.

How precisely did surveyors assign values of sizes to the fields? While most of the entries of proto-cuneiform tablets concerning fields give their sizes without any information

[15] On the determination of the absolute sizes of the signs for capacity measures, see Damerow and Englund 1987, 153–154.

on how the surveyors determined them, a small group of extant tablets contains measurements of the lengths and widths of fields together with the resulting areas. These examples show that the areas were determined implicitly by the relation that an area unit has the size of an area of 1 unit length and 1 unit width. We know from the later tradition that 1 *sar* was the size of an area of 1 *nindan* length and 1 *nindan* width. Accordingly, 1 *iku* was the size of an area of 10 *nindan* length and 10 *nindan* width. Although these relations are not directly attested to in proto-cuneiform surveying documents, the relations between higher units makes evident that the area notations in these documents were already based on these relations. In accordance with the basic relation between *nindan* and *iku* known from later documents, proto-cuneiform documents show, for instance, that the area of 2 *bur*$^{3}$, which is 360 *iku*, was the size of a field where length and width both measure 60 *nindan*.

How were the sizes of fields calculated on the basis of these relations? From a modern point of view and assuming that the fields were rectangular their areas have to be calculated by multiplying length and width. For this purpose, the measures of length and width would have to be converted into the same unit, preferably into the smallest unit *nindan*. The numbers of these units would have to be multiplied by some kind of algorithm, followed by a conversion into the standard area measures. But the evidence on the extant tablets precludes this interpretation.

1. The calculation of areas by multiplying the number of units of the length by the number of units of the width presupposes that the rectangular area is conceived as being covered with unit squares the number of which is defined as the size of the area. But the idea to determine the size of an area by covering it with squares was an unfamiliar idea to the whole Babylonian mathematical tradition. This is because the surveyors who coined the basic geometrical concepts created what Solomon Gandz described as a *geometry of lines*,[16] that is, a geometry without any concept of an *angle*. The size of an area was conceived as being determined by the length and width of the area independent of its shape. Any area with equal length and width was considered to have the same size. It will become clear in the following that this kind of geometry was perfectly adapted to carry out the task of the surveyors to determine the sizes of fields making use of only length measurements.
2. In proto-cuneiform documents the length measures were recorded with a precision down to 1 *nindan*. The area measures, however, were never recorded with a precision down to 1 square *nindan*, that is, to the level of 1 *sar*. While the lowest unit in area notations is usually the *iku*, areas are not usually recorded at this level of precision, but use *iku* as the smallest unit. It is unlikely that the units of *sar* which would automatically result from a multiplication of numbers of *nindan* were simply suppressed.
3. The two next higher units, the *eše*$^{3}$ with 600 *sar* and the *bur*$^{3}$ with 1800 *sar*, can no longer be expressed as squares of a length recorded with the applied notation system. But in any case they were too big to justify the precision of 1 *nindan* used by the surveyors of that time as the basic unit of length measurements. One must acknowledge that the sizes of areas were not perceived as the result of the multiplication of sexagesimal numbers to yield the number of square *nindan*, followed by a conversion into the irregularly structured system of area measures. Obviously, the notations representing

[16]Gandz 1929.

highly differentiated metrologies did not encourage the development of methods to perform abstract multiplications.

The neglect of algorithms for performing the multiplication of arbitrary sexagesimal numbers is characteristic until the end of the third millennium. The third millennium administrative tablets document instead three different kinds of implicit multiplication already evident in proto-cuneiform administrative documents.

1. Similar to the multiplication designated by the English term *times*, quantitative entries could be multiplied by small whole numbers, a technique which can easily be performed as repeated addition. From a modern point of view this arithmetical technique is the application of a dimensionless numeric operator to a qualified magnitude.
2. Linear relations between two sets of quantities such as the numbers of workers and corresponding amounts of rations to feed them can easily be extended to tables of corresponding pairs of values by applying the first type of multiplication to both of the corresponding values. In this case the factor by which the amounts of rations could be calculated from the numbers of workers, usually a fraction, remains completely implicit. From a modern point of view this arithmetical technique is the application of a dimensionless numeric operator to a linear function.
3. Areas of fields are, as we have seen, somehow calculated from length measures. From a modern point of view this arithmetical technique determines values by a bilinear function from one type of a specific kind of qualified magnitude, represented by length measures, into another type of qualified magnitude, represented by area measures.

The last type of multiplication is the subject of the discussion here. It is the only type where it is not obvious what kind of operations were performed in order to obtain the result. However, an atypical early Dynastic tablet excavated at Fara provides some critical information for the reconstruction of this procedure since it bears on its reverse some scribbled numerical notations that seem to be intermediate results of the calculation procedure.[17]

In the following entries on the obverse of the tablet the measures of a field and its area are registered:[18]

1. The *second entry* contains its length, measuring 1 UŠ 1 eš$^2$ 2 nindan.[19]
2. The *third entry* contains its width, measuring 1 UŠ 2 eš$^2$ 2 nindan.[20]

[17] See the interpretation of this tablet in Damerow 2001, 260–261. A rough drawing of the tablet was published by Jestin 1937, TSŠ 51. According to a photo and an unpublished copy of the tablet by Kazuya Maekawa discussed by Jöran Friberg at the *Third Workshop on Concept Development in Babylonian Mathematics* in December 1985 in Berlin, the drawing Jestin published contains serious errors. In particular, this drawing does not distinguish between a numerical sign written at the left edge of the tablet and the scribbled notation on the reverse nearby. The reconstruction of the numerical notations presented here is based on the photo of the tablet and Maekawa's copy.

[18] Since the first and the last entry on the obverse and the first line on the reverse are unrelated to the calculations discussed in the following, they are not taken into account here. The first entry on the obverse which is nearly illegible, probably due to an erasure by the ancient scribe, seems to contain an area notation of 1 eše$^3$ 3 iku. This area equals the notation in the clearly legible first line on the reverse. The relation to the rest of the tablet is unclear. The colophon in the last entry of the obverse seems to contain a geographical designation and a title.

[19] The value seems to have been corrected by the ancient scribe from 1 UŠ 2 eš$^2$ 2 nindan to 1 UŠ 1 eš$^2$ 2 nindan.

[20] In addition to the metrological notation of the width and the sign for the width of a field, the entry contains three oblique strokes which might be interpreted as representing three units of smaller measure. Since such a small length measure does not occur on any other tablet related to surveying from the Fara period, the meaning of the three strokes remains obscure. Smaller units known from later times have various sizes. 1 nindan equals 4 ni-kaš, 6 kuš-numun, or 12 kuš, see for instance the tables of calculated areas on the tablets OIP 14, 70 Edzard 1969 and

3. The *fourth entry* contains an area notation of 3 bur$^3$ 5 iku. Assuming that the readings of the measures of length and width are correct, the area of the field was 3 bur$^3$ 5 iku 4 sar which is extremely close to the area of 3 bur$^3$ 5 iku registered in the *fourth entry*.

How was this area calculated? The answer is provided by four of the five scribbled numerical notations on the reverse of the tablet:

1. The *second line* contains the notation 2 bur$^3$ 1 eše$^3$ 1 iku 20 sar, which is precisely the area of a field with a length of 1 UŠ 1 eš$^2$ 2 nindan and the width of 1 UŠ.
2. The *third line* contains the notation 2 eše$^3$ 2 iku which is close to the area of a field with the same length of 1 UŠ 1 eš$^2$ 2 nindan and the width of 2 eš$^2$.[21]
3. The *fourth line* contains the notation 3 bur$^3$ 3 iku, which is the sum of the second and the third line, thus representing the area of a field with a length of 1 UŠ 1 eš$^2$ 2 nindan and a width of 1 UŠ 2 eš$^2$. Again, the precise value of 3 bur$^3$ 3 iku 60 sar was rounded by dropping the 60 sar.
4. Finally the *fifth line* contains the area of a field of $1\frac{1}{2}$ iku(?) which is the rounded value of a field with the same length of 1 UŠ 1 eš$^2$ 2 nindan and the width of 2 nindan.[22]

Thus three of the scribbled numerical notations on the reverse of the tablet represent the results of the calculation of partial areas of the field with a length of 1 UŠ 1 eš$^2$ 2 nindan, and the three parts of the width 1 UŠ, 2 eš$^2$, and 2 nindan. Their sum, 3 bur$^3$ $4\frac{1}{2}$ iku, is the approximate area of a field with a length measuring 1 UŠ 1 eš$^2$ 2 nindan and a width measuring 1 UŠ 1 eš$^2$ 2 nindan as registered in the second and third entry on the obverse of the tablet. Rounded up to 3 bur$^3$ 5 iku, the calculated area equals the fourth entry on the obverse of the tablet.

This suggests that the ancient scribes used a calculation procedure based on the knowledge of the area for each pair of units of length and width measures:

If the length equals 1 nindan and the width 1 nindan then the area equals 1 sar.
If the length equals 1 nindan and the width 1 eš$^2$ then the area equals 10 sar.
If the length equals 1 nindan and the width 1 UŠ then the area equals 60 sar.
If the length equals 1 eš$^2$ and the width 1 eš$^2$ then the area equals 1 iku.
If the length equals 1 eš$^2$ and the width 1 UŠ then the area equals 1 eše$^3$.
If the length equals 1 UŠ and the width 1 UŠ then the area equals 2 bur$^3$.

A field with a length or or a width that is a multiple of a length unit has an area which is the same multiple of the corresponding *unit area*, that is, the area grows linearly with its length or width. Thus, partial areas such as those represented by the scribbled notations on the reverse of the tablet discussed above could easily be calculated from the *unit areas* by using the well-documented techniques to operate with linear relations.

This procedure was necessarily specific for the calculation of areas and was never adapted to other problems that, from a modern point of view, involve multiplication. It was based only on the arithmetical techniques described above. Furthermore, it was implicitly assumed that the sum of the sizes of partial areas equals the size of the total area.

CUNES 50-08-001 Friberg 2007a, 419–425, 499–500. For all three units the additional area (between 18 and 54 sar) the additional area would not substantially change the result but bring it closer to the total area recorded in the fourth entry of the obverse.

[21] The precise value would have been 2 eše$^3$ 2 iku 40 sar. The 40 sar have obviously been dropped.

[22] The precise value would have been 1 iku 42 sar.

Besides these conditions the procedure did not presuppose any arithmetical techniques that are unattested by the extant sources. In particular, the sizes of areas could be determined without making a detour through an intermediate sexagesimal result.

The precondition for effective application of such a procedure was, however, the knowledge of the sizes of *unit areas* for all combinations of length and width units. Tablets dating back to the first half of the third millennium, in fact, document that learning the sizes of such *unit areas* was an important part of the training of the scribes. From the Early Dynastic II period around 2700 BCE onward, school tablets survived containing tables or problems which could be calculated or solved by applying simple linear operations to the *unit areas*. Tables of this kind usually do not contain the *unit areas* directly but rather areas calculated from single-digit sizes of length and width, that is, sizes represented by multiples of a single sign.

The earliest tablet containing such a table, dates back to the Fara period.[23] Three columns list, first, the length of a field, followed by its width (equal to the length) and the calculated area. The lengths and widths go down from 10 to 1 UŠ, which equals 6 eš$^2$, continuing from 5 to 1 eš$^2$, and ending up with 5 nindan, which is half of 1 eš$^2$.

Such tables do not necessarily contain lengths and width with equal sizes. A recently published tablet[24] is quite similar to the tablet from the Early Dynastic II period, but starts with a length of 5 nindan and a width of 5 UŠ, continuing by doubling both sizes to 1 eš$^2$ and 10 UŠ, respectively, subsequently adding 1 eš$^2$ and 10 UŠ to length and width, respectively. This operation goes up to 5 eš$^2$ and 50 UŠ.[25]

Of special interest concerning the learning of the sizes of *unit areas* as a precondition for calculating areas is an Old Akkadian tablet some one hundred years younger, containing a problem and its solution.[26] The problem concerns small Old Akkadian length and area measures.

The length measures used are:

| | |
|---|---|
| 2 zipaḫ | = 1 GIŠ.BAD (= a cubit of approx. 50 cm) |
| 2 GIŠ.BAD | = kuš$^3$-numun |
| 6 kuš$^3$-numun | = 1 nindan |
| 10 nindan | = 1 eš$^2$ |

The area measures used are:

| | |
|---|---|
| 60 gin$^2$-tur | = 1 gin$^2$ (= approx. 0.6 square meter) |
| 60 gin$^2$ | = 1 sar |
| 100 sar | = 1 iku |

[23]The tablet VAT 12593 was originally published by Deimel 1923 and has been discussed several times in the literature, see e.g. Powell 1976, 429–431, Nissen, Damerow, and Englund 1993, 136–139, Damerow 2001, 262–263, Friberg 2007a, 149–150.
[24]Tablet MS 3047 of the Schøyen Collection, published by Friberg 2007a, 150–153, 484.
[25]This is followed by an entry with huge sizes that are difficult to interpret.
[26]Limet 1973, no. 36; see Powell 1976, 426–427.

The tablet contains the problem to calculate the area of a quadrilateral with equal sides the lengths of which are the sum of the involved length units. Thus, each side has the length:

1 eš$^2$ 1 nindan 1 kuš$^3$-numun 1 GIŠ.BAD 1 zipaḫ

If, in fact, the calculation method was based on the knowledge of the *unit areas*, this seemingly strange problem finds a simple explanation: The solution of the problem requires adding up the 25 *unit areas*, each combining two length units. With this problem the teacher therefore was able to check whether his pupils knew these *unit areas* or were able to calculate them.

| | eš$_2$ | nindan | kuš$_3$-numun | GIŠ.BAD | zipaḫ |
|---|---|---|---|---|---|
| eš$_2$ | 1 iku | 10 sar | 1 sar 40 gin$_2$ | 50 gin$_2$ | 25 gin$_2$ |
| nindan | 10 sar | 1 sar | 10 gin$_2$ | 5 gin$_2$ | 2 gin$_2$ 30 gin$_2$-tur |
| kuš$_3$-numun | 1 sar 40 gin$_2$ | 10 gin$_2$ | 1 gin$_2$ 40 gin$_2$-tur | 50 gin$_2$-tur | 25 gin$_2$-tur |
| GIŠ.BAD | 50 gin$_2$ | 5 gin$_2$ | 50 gin$_2$-tur | 25 gin$_2$-tur | 12 gin$_2$-tur $^{30}/_{60}$ |
| zipaḫ | 25 gin$_2$ | 2 gin$_2$ 30 gin$_2$-tur | 25 gin$_2$-tur | 12 gin$_2$-tur $^{30}/_{60}$ | 6 gin$_2$-tur $^{15}/_{60}$ |

Factor 10
Factor 6
Factor 2
Factor 2

Figure 3.5: *Unit areas* that had to be known to solve the problem on an Old Akkadian school text (Limet 36)

This procedure remained difficult to perform, however. In the third millennium, tablets containing length and area measures often show errors in the calculation of the areas, and partial areas including small measures were often treated negligently. Thus, in this case as well, the given solution is not correct.

The tablet contains the solution:

$1\frac{1}{4}$ (iku) gan$_2$ $2\frac{1}{2}$ šar 6 gin$_2$ 15 gin$_2$-tur

The correct solution would have been:

$1\frac{1}{4}$ (iku) gan$_2$ $2\frac{1}{2}$ šar 6 gin$_2$-tur $^{15}/_{60}$ gin$_2$-tur

The cause of the error is obvious.[27] The smallest *unit area* has the size:

6 gin$_2$-tur $^{15}/_{60}$ gin$_2$-tur

[27] Powell 1976, 427 has argued that the error proves that, in spite of the lack of any direct textual evidence, the principles of the sexagesimal positional system were already used in the early third millennium. Friberg 2005a, § 4.3 argued convincingly that other interpretations of the cause of the error are possible. He gives an geometrical reason different from the interpretation favored here.

It was erroneously interpreted as:

6 $gin_2$ 15 $gin_2$-tur

Given that no designation is known for $^1/_{60}$ of the small unit $gin_2$-tur (about 1 square decimeter) the error may well have been caused by the scribe's unfamiliarity with such small units, which, in fact, played no role in the contemporary practice of surveying.

## 3.4 The invention of the *surveyors' formula*

The determination of the size of an area from its measured length and width met only some of the challenges with which the surveyors of the administration of the early states were confronted. The fields they had to measure were rarely regular enough that their sizes could be determined on the basis of just measuring a length and a width. They usually were quadrilaterals with four unequal sides.

In order to solve the problem to calculate the size of such fields, the surveyors of the third millennium Mesopotamia already used what is called the *surveyors' formula*. This formula is well-known, in particular from later Greek and Roman surveying technology. The formula determines the area of an irregularly shaped polygon with four edges as the result of the multiplication of the means of opposite sides.

Determining the mean of two measures was a simple problem given the notation of measures used by the ancient scribes of Mesopotamia. As a rule, the opposite sides of a field did not differ too much so that the number of higher units of the notation of their sizes was equal. These units could simply be neglected. Only the differing parts of lower units had to be added for each pair, and then half the result had to be appended to the strings of higher units in order to get an area with equal opposite sides.

This procedure is well documented by sources from the second half of the third millennium. However, a favorable circumstance allows us to trace the roots of this formula back to the end of the fourth millennium, that is, to the origins of cuneiform writing. Among the earliest proto-cuneiform tablets from the Uruk IV period there is a group of school tablets containing exercises in area calculations without giving the result. One of these tablets turned out to provide crucial information about the early history of the *surveyors' formula*. The tablet is heavily damaged, but the geometry of the two surviving fragments made it possible to completely reconstruct its content. Each side contains two columns with two entries each. On both sides, the entries in the first column are qualified by an archaic sign for the length of a field, known from other proto-cuneiform tablets containing results of surveying. Accordingly, the entries in the second columns of both sides are qualified by an archaic sign for the width of a field. Obviously, each side contains the measures of an irregularly shaped field. The tablet does not give any information on the areas of the fields, but the application of the *surveyors' formula* results in both cases in the same 'round number' of 600 $bur_3$ or 10 $šar_2$.

There are several indications that the tablet contains a school text:

1. The text contains an exercise, giving the primary data, but not the solution.
2. The field area to be calculated is unrealistically large, corresponding to an area of a field with a length of about 7 km and a width of about 5 km. At the beginning of the third millennium, such a huge field would not fit into what is known about sizes

of individual fields at that time. Numerical area notations of this order of magnitude occur only as totals of several fields.

3. The solution would have to be written with a sign for 10 *šar*$_2$ which is not well attested by other documents of the time. Usually, this sign is the same sign as the sign for 10 *bur*$_3$, excluding misreadings only by its position in front of the sign *šar*$_2$. It seems that the exercise was composed with the aim to demonstrate the difficulty of denoting large areas with the proto-cuneiform means of the time.

In addition to the attestation of the earliest known use of the *surveyors' formula*, the tablet provides further important information. Starting from a predetermined result, the exercises are constructed accordingly on both sides of the tablet to produce the desired result. In the present case, the desired result is 10 *šar*$_2$. One of the possibilities to arrive at this solution is to choose a field of 1,200 nindan length (written as two times the sign for 600 nindan) and 900 nindan width (written with one sign for 600 nindan and five signs for 60 nindan). From this basic data the problem on the obverse of the tablet is constructed by adding 30 nindan to the first width and subtracting 30 nindan from the second width. The problem on the reverse of the tablet is constructed by subtracting 220 nindan from the first length and adding 220 nindan to the second length, and furthermore adding 380 nindan to the first width and subtracting 380 nindan from the second width. Calculating the means of opposite sides then results in the simple exercise to calculate the area of a field of 1200 nindan length and 900 nindan width.

This exercise tablet thus throws some light on the origins of the *surveyors' formula*. The ancient surveyors apparently assumed that the area of a field remains equal if they subtract some part from one side of a field and add a part of the same size to the opposite side.

From a modern point of view, the *surveyors' formula* yields only an approximate value of the 'real' area, that is, the procedure approximates the area as it is defined in the tradition of Euclidean geometry. But this cannot be the way the ancient surveyors of the third and second millennium BCE perceived the values they calculated. As long as the surveyed fields do not deviate too much from a rectangle, the *surveyors' formula* does in fact yield a very close approximation to the Euclidean area. Even if the surveyors had had any idea of an Euclidean area concept, it would have been difficult for them to find any empirical evidence of a difference from what they calculated using the *surveyors' formula*.

Moreover, for the time before the first millennium BCE there are no indications of any alternative concept of area that might have taken into account not only length measures but also the angles between the sides of an area. For the ancient surveyors an area must have been implicitly defined as the result of what they had calculated. There is not even a term for 'area' distinct from the numerical notations representing specific quantities. Only the term *iku* (sign form GANA$_2$) developed from a specific area unit into a general term for fields, used as a semantic marker of area notations. Furthermore, the term *a-ša*$_3$ designating a field, also written with the iconic graph GANA$_2$, ultimately developed into a general term for a calculated area in the Babylonian mathematics of the second millennium BCE. But, as will be shown in the following, the term was always identified with the result of the *surveyors' formula*.

## 3.5 Sophisticated surveying techniques in the Ur III period

The second half of the third millennium BCE was the era of the rise of big empires in Mesopotamia, culminating in the empire of the *Third Dynasty of Ur* (known as the Ur III period) which integrated southern and northern Mesopotamia and even parts of southwestern Iran for somewhat less than 100 years.[28] This integration was accompanied by reforms in the techniques of social control initiated by the powerful emperor Šulgi and, in particular, in the bookkeeping techniques used by the centralized administration.

One of these innovations concerned the administration of arable land. Šulgi created a new category of royal domains that were distributed to high-ranking officials in exchange for their services. Fields such as those of the royal domains could take on considerable size and an irregular shape.

This may be the reason for the emergence of a new type of cuneiform tablets documenting the work of surveyors. These documents do not simply give the measures of the surveyed fields but also drawings of their shape. The fields are drawn as polygons approximating the shapes of the real fields. These polygons were divided into triangles and quadrilaterals. More precisely, mostly one or more quadrilaterals were used to roughly approximate the field. The area of this group of quadrilaterals was designated by the Sumerian term *temen*. The term is usually translated as 'fundament' since it was primarily used to designate the foundation of a temple or a similar building, indicated by pegs nailed into the ground before the building was erected. The sketched *temen* was then further adapted to the shape of the field by adding small triangles and quadrilaterals to the edges where they did not reach the real border of the field and by subtracting such small areas where the edges went beyond them.

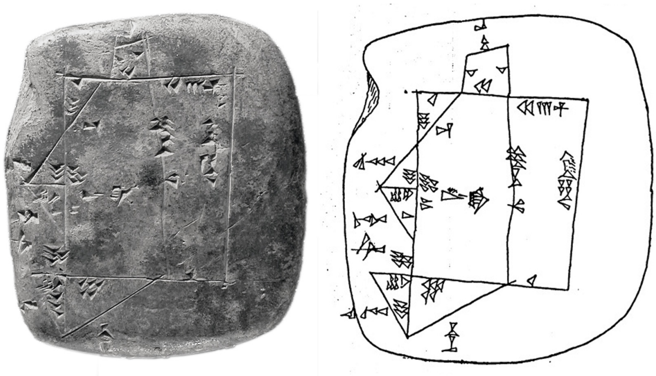

Figure 3.6: Field map of the Ur III period (MVN 10, 214)

Triangles were considered to be half of a quadrilateral with the longest side as the common side of both halves. Accordingly, the size of their area was calculated by halving the product of their two smaller sides. The size of the area of quadrilaterals was calculated using the *surveyors' formula*. In this way, first the size of the *temen* was calculated by

[28]Gibson and Biggs 1991.

adding up the partial areas in the case of complex shapes of the *temen*. The measuring and calculation of the sizes of the small appended or cut-off areas was often simplified by measuring only three of the four sides of a quadrilateral, leaving out the measurement of the outer border. The results of the calculation of the sizes of the appended or cut-off areas were then added to, or subtracted from, the size of the area of the *temen* in order to obtain the size of the total area of the irregularly shaped field.

It is characteristic of the field maps of the Ur III period that they were never drawn to scale. The shapes of the drawn fields always turn out to be heavily distorted when compared with the real fields that were surveyed, as far as their shapes can be approximately reconstructed from the given measures.[29] This characteristic of the field maps can easily be explained by their function. For the administrators who needed to know the sizes of the fields because they had to decide on the amount of required seed grain, the required work force, or the expected harvest, and consequently for the surveyors who prepared the necessary data for them, the geometry of the fields was completely irrelevant. Using the iconic representation of the topology of the irregularly shaped fields, the drawings rather served to determine how the measurements had to be processed in order to arrive at their correct total sizes.

## 3.6 From context-dependent to abstract notations of quantities

Besides the use of maps to document complex shapes of fields and to organize the calculation of their areas, the Ur III period with its centralized administration provided the setting for what can be considered the most important mathematical invention of the ancient Near East, that is, the invention of the sexagesimal positional system of numerical notation.

Before the Ur III period numerical notations were based without exception on signs or sign combinations representing the absolute value of metrological units or, in the case of counting, on specific signs for each base unit, that is, specific signs for 1, 10, 60, 600, 3600 etc. The innovation consisted in repeating the signs for 1 and 10 for the higher units, that is, using 1, 10, 1, 10, 1 etc. for the same sequence of units, distinguishing them only by their position. Eleanor Robson characterized this innovation by arguing that the sexagesimal place value system

> ... changed the status of numbers from properties of real-world objects to independent entities that could be manipulated without regard to absolute value or metrological system.[30]

[29] According to the figures given on the field map MVN 10, 214, presented here as an example, the large quadrilateral in the center of the tablet has a length of 170 nindan (= 2 UŠ 50 nindan, written along the left side), an upper width of 53 ½ nindan (= 10 + 20 + 23 ½ nindan, written along the upper side), and a lower width of 60 nindan (= 50 + 10 nindan, written along the lower side). If the map were drawn to scale it should thus be approximately 3 times as long as it is wide. But the drawing on the tablet is close to a square. The reverse of the tablet (not depicted here) lists the essential figures required by the administration in three lines. The first line contains the sum of the small areas to be added to the large quadrilateral. Accordingly, the third line contains the sum of the areas to be subtracted. For some reason, the area to be subtracted consists of the whole right part of the large quadrilateral (damaged soil?) and the small triangle at its upper left corner. The second line contains the remaining area of the large quadrilateral, designated as *core temen*. The fourth line contains the name of the field, a fifth line the name of the official who authorized the document. The interpretation of the figures in the first three lines differs slightly from the figures given on the map, but comes close enough to them not to exceed the typical lack of accuracy frequently found in surveying documents. For details on the tablet discussed here, see Fuÿe 1915; Dunham 1986.

[30] Robson 2008, 78.

Although the missing zero and the missing separation between whole units and fractions (comparable to the decimal comma in the modern decimal positional system) caused difficulties in applying the new system, the purpose of the innovation is clear and its inherent potential is obvious. Its invention was a precondition for the development of efficient calculation methods, in particular of universally applicable algorithms for multiplication and division.

Unfortunately, the origin of the new system is difficult to trace. For a long time the question was simply neglected, assuming that dealing with quantities is in any case a universal human faculty and the symbolic representation of numbers merely a circumstantial realization of a common human numeracy. The fact that all known sources documenting the sexagesimal positional system come from the first dynasty of Isin, the short-lived dynasty of Larsa, and the first dynasty of Babylon (the so-called Old Babylonian period), or later periods, that is the time from approximately 2000 BCE onwards, did not provoke any investigation of the origin of this system.

This situation changed over the last decades, along with a growing interest in texts of the third millennium that somehow indicate mathematical knowledge, but do not belong to the corpus of Babylonian mathematics in the narrow sense as it was understood by the pioneers of the decipherment of mathematical cuneiform texts, in particular Otto Neugebauer and Françoise Thureau-Dangin.[31] In this context the search for the earliest attestations of numerical notations written in the Ur III period, that is in the last century of the third millennium, became an issue for research.

In contrast to the high number of cuneiform tablets documenting the use of the sexagesimal positional system in the following periods, attestations of the use of this system in the Ur III period are extremely scarce.[32] There is, at least, evidence that the system was developed enough to establish tables of reciprocals which became the main tools for performing division.[33] However, neither is it possible to decide on the basis of current research results whether the invention of the new system was a unique creation ‘out of the box’ or the result of a gradual extension and improvement of some basic ideas, nor is it possible to determine the location of its origin, the paths of its dissemination, or by whom and to what extent it was used in this period.

However, the inherent potential of the sexagesimal positional system of numerical notations is obvious. The new system provided a basis for the development of efficient algorithms for performing multiplication and division independent of the context of application. In particular, the system made it possible to merge the different forms of multiplication into one unique numerical operation. The cumbersome method the surveyors used to determine the areas of fields, for example, could be replaced by such a multiplication algorithm. The precondition was, of course, that the traditional numerical notations be converted into the new system.

The abundance of extant early second millennium tablets dedicated to the use of the new system shows unambiguously that this potential was in fact realized shortly after its invention. The majority of these tablets contain tables which were probably produced in the context of scribal schools to teach them to perform the operations of the new system. A brief

[31] Neugebauer 1935/1937; Thureau-Dangin 1938; Neugebauer and Sachs 1945.
[32] Robson 2008, 75–84.
[33] See the tables documented by Robson 2008, 81–83 and by Oelsner in his contribution to Høyrup and Damerow 2001, 53–59.

survey of the most frequent types of such tables elucidates how systematically the potentials of the new system were explored and transformed into techniques to handle quantitative data:[34]

1. *Metrological tables* containing lists of traditional metrological notations, in particular notations of capacity measures, weight measures, length measures, and area measures, ordered according to the sizes of the values they represent in the first column, and corresponding notations in the new sexagesimal positional system.
2. *Multiplication tables* containing the multiples 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, and 60 of certain fixed numbers, for example 6: 1 times 6, 2 times 6, etc.
3. *Tables of reciprocals* containing, as a rule, the reciprocals of 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 16, 18, 20, 24, 25, 27, 30, 32, 36, 40, 45, 48, 50, 54, 60, 1 4 (= 60 + 4 = 64), and 1 21 (= 60 + 21 = 81).
4. *Tables of coefficients* used for converting values such as, in the context of architecture, to convert a volume into the number of bricks required to fill it; in the context of metal work to convert an area to be gold-plated into the required amount of gold; in the context of geometry to convert the length of the side of a square into its diagonal.[35]

This enormous corpus of hundreds of tablets from the short period of approximately 500 years after the invention of the sexagesimal positional system represents an unprecedented type of tablets with an arithmetical content. While these tablets demonstrate a radical break with a tradition of 1000 years of dealing with quantitative data, they also exhibit a strong continuity, which explains the rapid development of the new system.

1. The *Metrological tables* document the basic continuity between the traditional metrological systems and the sexagesimal positional system. The underlying technique of transformation assured the smooth conversion of traditional data into the new format and, conversely, of sexagesimal positional data into the traditional form.
2. The *Multiplication tables* were created using the method familiar from treating linear relations between two sets of quantities by applying repeated additions to both sets.
3. The *Tables of reciprocals* were created by using and further developing a technique that had been developed in the context of the strictly sexagesimally structured weight measures to replace unit fractions such as $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, $\frac{1}{5}$, etc., with a combination of multiples of smaller units, that is, 30, 20, 15, 12, etc. units.[36]
4. The *Tables of coefficients* make the implicite factors of linear relations explicite. These tables show that the invention of the sexagesimal positional system had the immediate consequence of merging the three former types of multiplications (*n* times *x*, linear

[34] Comprehensive lists and editions of such tables are supplied by Hilprecht 1906, 57–70 and Plate 1–30, Neugebauer 1935/1937, Vol. 1, 4–82, Neugebauer and Sachs 1945, 11–36, Robson 1999, 193–207, and Proust 2008, 15–73.

[35] There are further types of tables which are, however, less common. Such tables contain, for example, squares and square roots, cubes and cube roots, exponentials and logarithms, and, to mention a unique finding, columns of figures based on Pythagorean numbers, that is, whole numbers *a*, *b*, and *c* that satisfy the condition $a^2 + b^2 = c^2$. On the various, partly incompatible interpretations of this extraordinary tablet (Plimpton 322) see Neugebauer and Sachs 1945, 38–41, Schmidt 1980, Friberg 1981, Damerow 2001, 231–232 and 294, and Robson 2001.

[36] Since this method fails if the denominator has prime factors other than two, three, and five, which are the only ones contained in the base sixty of the system, such values are simply omitted in the tables of reciprocals. Furthermore, the calculation of reciprocals of notations with several digits is much more complicated. See Sachs 1947.

relations with an implicit factor, and area calculations) into one context-independent operation of multiplication.

The extent to which the sexagesimal positional system was used for bureaucratic everyday tasks is an open question, as is the extent to which such tables were used by administrators as reference tables and not merely as exercises in scribal schools. But, whoever used them must have adopted a concept of numerical notations that was based on cognitive constructions which were independent of specific contexts to a greater degree than those underlying the traditional metrological systems. This is particularly obvious in the case of length, area, and capacity measures. When represented by sexagesimal positional notations, they all turned out to be merely specific kinds of quantities, so that the traditional methods of calculation could be replaced by unified methods of addition, subtraction, multiplication and division. Only the last of these operations, namely division, remained problematic since it was not universally applicable.

There are further differences from the modern decimal positional system. *First*, the sexagesimal positional system lacked any indication of the absolute size. For example, the vertical wedge represents 1, but likewise also $^1/_{3600}$, $^1/_{60}$, 60, $60^2$, $60^3$, etc. Thus, additions and subtractions required an independent control of the absolute sizes. *Second*, the system lacked a sign for zero so that notations that required an inner zero became ambiguous.

Division was performed by multiplying by the reciprocal of the divisor, but the calculation of reciprocals as they were listed in the *tables of reciprocals* remained difficult and, as mentioned above, could not be performed if the prime factorization of the divisor contained prime factors other than two, three and five, namely the prime factors of sixty.[37]

In spite of the integration of metrological notations into a closed system of abstract numerical operations, this system evidently differs from the modern number concept. It is, in fact, difficult to imagine how, long before the creation of modern mathematical formalism, a numerical system without an inherent absolute value of its entities could be conceived. In fact, neither in the Sumerian, nor in the Akkadian language did a word exist that can be convincingly interpreted as a designation of such an entity. In particular, a word for *number* was missing in these languages.[38]

The conceptual change from metrological to sexagesimal positional notation had its most dramatic consequences for cognitive constructions related to spatial relations. While in the context of traditional metrological notations length measures, field areas, and capacities of containers were associated with only certain external domains of human activities, and connected only through these activities, they now became entities internally linked within the system of the operations that constituted the sexagesimal positional system. The following section will show the consequences of this epistemic shift, resulting in the unparalleled knowledge system of Babylonian geometry.

[37]Two, three, and five have the reciprocals 30, 20, and 12 respectively. The prime factorization of any number containing only these prime factors has a reciprocal which is the product of the corresponding reciprocals. No other prime number has a finite representation by units of the sexagesimal system. Thus, any number containing a prime factor other than two, three, and five also has no finite representation in this system.

[38]Frequently the Sumerian word 'šid' with the Akkadian translation '*minûtu*' or the Akkadian word '*mānu*' are considered as candidates for representing the term 'number'. This attribution is, however, an anachronistic projection of modern concepts, neglecting the contexts in which the terms were used, as they usually related to operations and not to entities.

## 3.7 The heritage of the surveyors in Babylonian mathematics

The types of knowledge discussed so far were more or less forms of the knowledge of practitioners. It is characteristic of such knowledge that it is determined by the goals of its practical use. These goals limit the range of applications and at the same time the potential for further developments. However, social institutions, in which such knowledge played a different role, developed along with the accumulation and specialization of the knowledge of practitioners. This was the case with institutions of education and training, which fulfilled the function of providing the qualifications the practitioners needed. While it is likely that basic qualifications of practitioners were transmitted to apprentices by participation and imitation, as is known from non-literate indigenous cultures,[39] the development of writing was surely accompanied by a growing institutionalization of the transmission of knowledge.[40]

This new type of knowledge transmission changed the character of the knowledge itself. Knowledge transmitted by institutionalized education was no longer structured and limited by the actual goals of the practitioners but rather, to a certain extent, determined by the aim of teaching how to use the available tools and elaborating their inherent potentials. Consequently, together with the establishment of institutionalized teaching, new types of texts arose, which usually are designated as *school texts*, showing one or more of the following characteristics:

1. The texts were *neither signed by a responsible official nor dated*.
2. The *scope of the exercises was unrealistically extended* beyond any practical needs.
3. The *values of exercises were systematically varied* and often documented in the form of tables representing relations between conditions and results, for instance, tables of length measures together with the calculated areas.
4. *Conditions and results of administrative procedures were exchanged* in order to produce new types of exercises: Results of typical operations of practitioners such as a calculated area together with, for instance, the length of a field were given from which the width had to be calculated.

While the operations with traditional metrological notations were not substantially influenced by school traditions, the impact on operations with the sexagesimal positional system was outstanding. Beyond the simplification and integration of arithmetical operations, the arithmetization of geometrical entities such as length or area measures made unrealistic operations such as the addition of lengths and area measures possible. Following the habit of teachers to construct exercises systematically by modifying the conditions and by exchanging the conditions and results of administrative procedures, completely new problems were generated. A typical example consisting of twenty-four problems with more or less systematically varied conditions is provided by a frequently discussed Old Babylonian tablet written in the first half of the second millennium BCE.[41] In an anachronistic modern translation, the first four problems and their solutions recorded on this tablet read as follows:

1. The area and the side of a square I have added: 45 is it.
2. The side of a square I have subtracted from its area: 14 30 is it.

[39] Alt 1956.
[40] Sjöberg 1973; Sjöberg 1975, Robson 2008, 31–33, 40–53, 97–106, 115–136, and 192–198.
[41] BM 13901, see Neugebauer 1935/1937, Vol. 3, 1–14.

3. The third of the area of a square I have subtracted. The third of its side I have added: 20 is it.
4. The third of the area of a square I have subtracted. The side I have added: 4 46 40 is it.

In all four problems, the task is, of course, to determine the length of the side. Keeping in mind that the notations in the sexagesimal positional system are determined except for a power of sixty as a factor, the construction of these problems is straightforward and simple using the multiplication tables of the sexagesimal positional system. In all cases the side of the square to be determined is 30 (times $60^n$). The given areas are calculated as follows:

1. 30 times 30 is 15 (times 60). 15 plus 30 is 45 (the given area).
2. 30 times 30 is 15 (times 60). 15 (times 60) minus 30 is 14 30 (the given area).
3. 30 times 30 is 15 (times 60). The third of 15 is 5. 15, the area, minus 5 is 10. The third of 30, the side, is 10. 10 plus 10 are 20 (the given area).
4. 20 times 20 is 6 40 (6 times 60 plus 40). The third of 6 40 is 2 13 20 (times $^1/_{60}$). 6 40, the area, minus 2 13 20 (times $^1/_{60}$) is 4 26 40 (times $^1/_{60}$). 4 26 40 (times $^1/_{60}$) plus 20, the side, is 4 46 40 (times $^1/_{60}$, the given area).

While the arithmetical operations of multiplication and division involved are quite simple and their results independent of the order of magnitude with regard to the power of sixty, there is obviously a difficult problem if additions and subtractions are involved. Adding or subtracting sexagesimal positional notations produces different results if the relative position of the notations is different with regard to the power of sixty. The ambiguity of the notation system thus causes an ambiguity in the operations to be performed. It was therefore always common sense that the ancient scribes must have kept track of the absolute values of their notations without, however, ever attempting to find a notation to express these values.

This common-sense interpretation is supported to a certain degree by recent philological work.[42] It turns out that the concepts represented by the terms used are partly arithmetical and partly geometrical, representing the traditional distinction between different kinds of metrologically determined operations, in particular different kinds of multiplication.[43] Thus, in the present example the term for multiplication is identical with the term for the area of a field, the terms for addition and subtraction seem to have an arithmetical origin, even though the values to which they are applied, the sizes of lengths and areas, also have a clear geometrical meaning. It is therefore plausible, that the ancient scribes kept track of the absolute values of the arithmetical notations by keeping in mind their metrological meanings.

In contrast to the simplicity of problems based on the operations of the practitioners, reverse operations led to more sophisticated problem solutions. Albeit recorded on many of the tablets, they are still poorly understood.[44] In modern terms, the examples given above represent second-degree equations:

---

[42] Høyrup 1990; Høyrup 2002.

[43] Høyrup 1990, 45–69, Høyrup 2002, 18–49.

[44] Several or even the majority of historians of science will probably not agree with this statement. Translated in terms of modern mathematics most of the problems are in fact understandable. The solutions appear to be either correct or somehow flawed, allegedly due to an insufficient qualification of the ancient scribes. From the viewpoint of historical epistemology, however, such errors shed some light on the inherent mental operations of Babylonian mathematics. In spite of the progress made in the interpretation of Babylonian mathematics after the *geometric turn*, the present understanding of this ‘non-Euclidean’ kind of mathematics is deficient since there is still

$$x^2 + x = a \tag{3.1}$$

$$x^2 - x = a \tag{3.2}$$

$$x^2 - \alpha x^2 + \beta x = a \tag{3.3}$$

$$x^2 - \alpha x^2 + x = a \tag{3.4}$$

Solutions to such problems, if featured on the tablets, describe only the operations to be performed with the given sexagesimal positional notations. Any reflection on the reasons for the operations performed is usually missing. Accordingly, the first step to understand the performed operations was to rewrite them in generalized form using modern algebraic notations with the result that the operations follow roughly the algebraic solution formula of second-degree equations. Despite the fact that the origins of the solution procedure remained obscure, Babylonian mathematics seemed to be some kind of Babylonian algebra.

The new geometric interpretation of key terms of Babylonian mathematics undermined this understanding. For the first time, some Babylonian problem solutions, at least, became plausible on the basis of hypothetical geometrical figures, which might have existed on scrap pads or as memorized mental images.

As a consequence of this *geometric turn* in the interpretation of problems posed in Babylonian mathematics, there is a tendency to supplement them systematically with Euclidean-style figures. Given that there is a common basis of practitioners' knowledge in Babylonian and Euclidean geometry, it does not come as a surprise that the Euclidean-style figures offer some explanation for the solution procedures. However, this approach disguises the close connection with the surveyors' tradition. Babylonian mathematics still did not account for angles and thus for the similarity of shapes.[45] The incompatibility of this characteristic, which was inherited from the surveying of fields based exclusively on length measurements, is particularly obvious in cases where tablets themselves illustrate problems with drawings that contradict a Euclidean interpretation. This is the case, for instance, when triangles were drawn which cannot be right-angled, but the interpretation has to treat them as if they had a right angle.[46]

The geometrical problems of Babylonian mathematics were determined by

1. the length of straight lines, which may approximate curved lines,
2. the assumption that the size of a figure which consists of partial areas equals the sum of these partial areas,

no comprehensive reconstruction of the mental models and operations that provide a coherent image of the partly successful and partly erroneous solutions of the problems documented on the tablets of Babylonian mathematics. For a critical reassessment of some interpretations of tablets allegedly dealing with the Pythagorean theorem see Damerow 2001.

[45] Høyrup 2002, 228 discussing the role of angles and of similarity justifies the use of Euclidean-style figures:

> Exactly *how* it [i.e. the right angle] was understood we do not know – the texts speak too little about it to allow us to understand – but it is beyond doubt that a (probably intuitive) concept of similarity or 'same shape' was at hand.

[46] See for instance the tablet Strssbg. 364 Neugebauer 1935/1937, volume 1, 248–356 and volume 3, plate 11. See also the overview of similar problems provided by Friberg 2007b, 244–268. Friberg 2005b, 46–50 has convincingly shown that Egyptian mathematics dealt with problems such as those on Strssbg. 364 in the same way.

3. the assumption that the area of a quadrilateral is determined by the *surveyors' formula*.

With the Euclidean concept in mind, these assumptions could lead to counterintuitive consequences. On one of the Old Babylonian mathematical tablets,[47] for instance, a quadrilateral is cut into two equal pieces. The scribe asked for the length of the dividing line.

From a Euclidean perspective and its concept of the size of an area this question is pure nonsense, because the length of the dividing line depends on the angle of the cut. From the perspective of the Babylonian concept of the size of an area, however, the problem makes perfect sense. The condition that the size of the whole area calculated according to the *surveyors' formula* equals the sum of the sizes of the cut parts, also calculated according to the *surveyors' formula*, is a strong condition contradicting the Euclidean concept of the size of an area. Under this strong condition, the length of the dividing line is, in fact, mathematically determined by a simple relation between the lengths of the two opposite sides of the quadrilateral, which are not cut by the dividing line, and the length of the dividing line itself. If $a$ and $b$ are the lengths of the two opposite sides and $c$ is the length of the dividing line, then the Babylonian assumptions imply the relation:[48]

$$a^2 + b^2 = 2c^2 \tag{3.5}$$

This example is not an exceptional one but rather the simplest kind on a type of tablets that deal with the division of irregular fields according to specific criteria. In all likelihood, this type of problem goes back to the problems of the surveyors, even though such problems are not documented by administrative tablets. It is known, however, that officials were allotted arable land as compensation for services and deliveries of agricultural products. It is quite possible that the allotments had to have a specific size so that, contrary to what is known from surveyors' administrative documents, length measures had to be determined under certain conditions for given area measures. At least, school tablets from the second half of the third millennium BCE remain, which ask for length measures to be calculated from a given area which was a difficult problem until the sexagesimal positional system was invented. In any case, the majority of tablets dealing with the division of fields belong to the mathematical tablets of the Old Babylonian period.

From the origin of writing at the beginning of the third millennium up to at least the first half of the first millennium BCE, all mathematical tablets dealing with areas were based on the *surveyors' formula*. From the viewpoint of modern mathematics, all calculations on these tablets are merely approximations of results which can be developed on the basis of Euclidean geometry.

But is it possible that this was also the perspective of the Babylonian scribes? There is, in fact, no evidence whatsoever that they considered their calculation of areas to be some kind of approximation of a different understanding of 'true' areas for which they had neither a theoretical nor even a conceptual basis. While they probably were aware of the fact that their calculated areas were not precise in any absolute sense, they surely did not see any difference between the dependence of area measures on external circumstances and the limited precision of measures such as the measures of lengths, weights, volumes, or economic

[47] YBC 4675, published by Neugebauer and Sachs 1945, 44–48.
[48] For a simple proof see Bruins 1955, 46 and Vogel 1959, 70.

exchange rates. But there were, of course, practitioners' *rules of thumb* for achieving reproducible results, such as checking the reliability of a balance in equilibrium by exchanging the sides of the load and the balance weight, or applying the *surveyors' formula* only to shapes which differ not too much from rectangles.

The scribes who composed Babylonian mathematics inherited this understanding of area measures from the surveyors and made the *surveyors' formula* a core operation of posing problems and establishing ways to solve them. This insight challenges the common mainstream interpretation of Babylonian mathematics altogether. While in many cases taking into account the non-Euclidean concept of area and disregarding the angles of their shapes may not lead to substantially different interpretations, an understanding of these characteristics of Babylonian mathematics turns out to be crucial as soon as the division of areas and cut-and-paste procedures are involved, and when the aim is to reconstruct the epistemic basis and the origins of the solutions to the sophisticated problems of this esoteric art of the scribes.

## 3.8 Conclusion

This chapter investigated the development of abilities to deal with spatial relations in ancient Mesopotamia.[49] It covers the time period from pre-history in the fourth millennium BCE to the first half of the second millennium BCE when Babylonian mathematics reached its first climax. At the beginning of the chapter the questions were posed as to the historical circumstances under which the nature of the conceptualization of space changed from spatial orientation techniques in pre-historic times to Babylonian geometry and what brought about these changes.

Summarizing the results of the chapter, two major innovations can be identified that triggered fundamental changes related to spatial cognition. Both innovations concern the representation of spatial relations by notation systems. The first innovation was closely related to the transition from rural communities in the alluvial plane of Mesopotamia to the centralized and stratified societies of the first city states. This transition was associated with the development of administrative tools such as seals, tokens, numerical impressions into clay, and finally the invention of the proto-cuneiform writing system, a non-glottographic system of operations with symbols. These operations represented administrative activities and were used to control them. They were based on the standardization of measures and their integration into systems. The second innovation that triggered fundamental changes was the invention of the sexagesimal positional system of numerical notation.

The overall development of knowledge which had its origin in the practitioners' knowledge of surveyors and its climax in the sophisticated geometry of Babylonian mathematics can be characterized as a fundamental shift in the socially shared conceptual structures used to understand spatial relationships. This development shows how content-specific mental models were transformed into an abstract mathematical model of space, which however, differs fundamentally from Euclidean geometry which later became the basis of mathematical developments in the European mathematical tradition.

[49] Peter Damerow passed away before he could finish this concluding section, which he had headed "Mental models of spatial relations in the third millennium BCE." We have not attempted to complete the section. – MS.

## Bibliography

Alt, Robert (1956). *Vorlesungen über die Erziehung auf frühen Stufen der Menschheitsentwicklung*. Berlin: Volk und Wissen.

Bauer, Josef, Robert K. Englund, and Manfred Krebernik (1998). *Mesopotamien: Späturuk-Zeit und Frühdynastische Zeit*. Annäherungen 1. Göttingen: Vandenhoeck & Ruprecht.

Boltz, William G. (2006). Pictographic Myths. In: *Sprache und Denken in China und Japan*. Ed. by Wolfgang Behr and Heiner Roetz. 30. Bochum Yearbook of East Asian Studies (BJOAF). Bochum: Faculty of East Asian Studies at Ruhr-Universität Bochum, 39–54.

Bruins, E. M. (1955). On the System of Babylonian Geometry. *Sumer* 11(1):44–49.

Damerow, Peter (2001). Kannten die Babylonier den Satz des Pythagoras? Epistemologische Anmerkungen zur Natur der Babylonischen Mathematik. In: *Changing Views on Ancient Near Eastern Mathematics*. Ed. by Jens Høyrup and Peter Damerow. Berlin: Reimer, 219–310.

Damerow, Peter and Robert K. Englund (1987). Die Zahlzeichensysteme der Archaischen Texte aus Uruk. In: *Zeichenliste der Archaischen Texte aus Uruk (ATU 2)*. Ed. by Margret W. Green and Hans J. Nissen. 2. Archaische Texte aus Uruk. Berlin: Gebr. Mann, 117–166.

Deimel, Anton (1923). *Schultexte aus Fara*. Die Inschriften von Fara. Leipzig: Hinrichs.

Driel, G. van (1982). Tablets from Jebel Aruda. In: *Zikir Sumin*. Ed. by G. Van Driel. Leiden: Brill, 12–25.

Dunham, Sally (1986). Sumerian Words for Foundation. *Revue d'Assyriologie* 80:31–63.

Edzard, Dietz O. (1969). Eine altsumerische Rechentafel. In: *Lisan mithurti: Festschrift Wolfram Freiherr von Soden*. Ed. by Wolfgang Röllig. Kevelaer: Butzon und Bercker.

Elias, Norbert (1984). *Über die Zeit*. Frankfurt a. M.: Suhrkamp.

Friberg, Jöran (1981). Methods and Traditions of Babylonian Mathematics: Plimpton 322, Pythagorean Triples, and the Babylonian Triangle Parameter Equations. *Historia Mathematica* 8:277–318.

— (2005a). On the Alleged Counting with Sexagesimal Place Value Numbers in Mathematical Cuneiform Texts from the Third Millennium BC. *Cuneiform Digital Library Journal* 2005:2.

— (2005b). *Unexpected Links between Egptian and Babylonian Mathematics*. New Jersey: World Scientific.

— (2007a). *A Remarkable Collection of Babylonian Mathematical Texts*. New York: Springer.

— (2007b). *Amazing Traces of a Babylonian Origin in Greek Mathematics*. New Jersey: World Scientific.

Fuÿe, Allotte de la (1915). Un cadastre de Djokha. *Revue d'Assyriologie* 12(1):47–54.

Gandz, Solomon (1929). The Origin of Angle-Geometry. *ISIS* 12(1929):452–481.

Gibson, McGuire and Robert D. Biggs (1991). *The Organization of Power: Aspects of Bureaucracy in the Ancient Near East*. 2nd ed. Chicago: The Oriental Institute of the University of Chicago.

Hilprecht, H. V. (1906). *Mathematical, Metrological and Chronological Tablets from the Temple Library of Nippur*. Philadelphia: Department of Archaeology, University of Pennsylvania.

Høyrup, Jens (1990). Algebra and Naive Geometry. *Altorientalische Forschungen* 17:27–69, 262–354.
— (2002). *Length, Width, Surfaces: A Portrait of Old Babylonian Mathematics and Its Kins*. New York: Springer.
Høyrup, Jens and Peter Damerow, eds. (2001). *Changings Views on Ancient Near Eastern Mathematics*. Berliner Beiträge zum vorderen Orient. Berlin: Reimer.
Hyman, Malcom D. (2006). Of Glyphs and Glottography. *Language and Communication* 26(3-4):231–249.
Jestin, R. (1937). *Tablettes sumérienne de Shuruppak conservées au Musée se Stamboul*. Paris: Boccard.
Limet, Henri (1973). *Étude de documents de la periode d'Agade appartenant à l'Université de Liège*. Bibliothèque de la Faculté de Philosophie et Lettres de l'Université de Liège. Paris: Société d'Editions « Les Belles Lettres ».
Neugebauer, Otto (1935/1937). *Mathematische Keilschrifttexte*. Berlin: Springer.
Neugebauer, Otto and Abraham Sachs (1945). *Mathematical Cuneiform Texts*. New Haven: American Oriental Society.
Nissen, Hans Jörg, Peter Damerow, and Robert K. Englund (1993). *Archaic Bookkeeping: Early Writing and Techniques of Economic Administration in the Ancient Near East*. Chicago: Chicago University Press.
Powell, Marvin A. (1976). The Antecedents of Old Babylonian Place Notation and the Early History of Babylonian Mathematics. *Historia Mathematica* 3:417–439.
Proust, Christine (2008). *Tablettes mathématiques de la collection Hilprecht*. Wiesbaden: Harrassowitz.
Robson, Eleanor (1999). *Mesopotamian Mathematics, 2100-1600 BC. Technical Constants in Bureaucracy and Education*. Oxford: Oxford University Press.
— (2001). Neither Sherlock Holmes nor Babylon. A Reassessment of Plimpton 322. *Historia Mathematica* 28:167–206.
— (2008). *Mathematics in Ancient Iraq. A Social History*. Princeton: Princeton University Press.
Sachs, Abraham (1947). Babylonian Mathematical Texts I. Reciprocals of Regular Sexagesimal Numbers. *Journal of Cuneiform Studies* 1(3):219–240.
Schmandt-Besserat, Denise (1992). *Before Writing*. Austin: University of Texas Press.
Schmidt, Olaf (1980). On Plimpton 322. Pythagorean Numbers in Babylonian Mathematics. *Centaurus* 24:4–13.
Sjöberg, Ake W. (1973). Der Vater und sein missratener Sohn. *Journal of Cuneiform Studies* 25:105–169.
— (1975). Der Examenstext A. *Zeitschrift für Assyriologie* 64:137–177.
Strommenger, Eva (1980). *Habuba Kabira, eine Stadt vor 5000 Jahren*. Mainz: Zabern.
Thureau-Dangin, Françoise (1938). *Textes mathématiques babyloniens*. Leiden: Brill.
Vogel, Kurt (1959). *Vorgriechische Mathematik, Teil II: Die Mathematik der Babylonier*. Hannover: Schroedel.
Wertheimer, Max (1925). Über das Denken der Naturvölker: Zahlen und Zahlgebilde. In: *Drei Abhandlungen zur Gestalttheorie*. Ed. by Max Wertheimer. Erlangen: Verlag der philosophischen Akademie, 106–163.

# Chapter 4
# Theoretical Reflections on Elementary Actions and Instrumental Practices: The Example of the *Mohist Canon*

*William G. Boltz and Matthias Schemmel*

## 4.1 Elementary actions, instrumental practices, and theoretical knowledge

This chapter is concerned with an analysis of the small body of texts usually referred to as the *Mohist Canon*, particularly focusing on the sections that are concerned with concepts of space, time, and matter. These texts, written around 300 BCE, constitute one of the most important sources for understanding ancient Chinese thinking about natural and technical aspects of the environment, what is often called 'Later Mohist Science'.[1]

When we call a historical cultural activity 'science', we usually justify this by identifying certain features of the activity as in some way 'scientific'. This practice may relate to such things as the recognition and systematic observation of regularities in the physical world, the explanation of such regularities by causal reasoning or by deductive argument, the use of mathematics, or the production of knowledge by systematic experimentation. In all cases we pick certain features of modern science, which we would not doubt to represent scientific thinking, and on this basis assess the extent to which the historical activities in question qualify as 'science'. In the present study, rather than advocating a certain definition of what 'science' is, we would like to shift the focus to the concept of *theoretical knowledge*, which clearly constitutes an important ingredient of present-day science, but which may also contain kinds of knowledge that under more rigorous criteria would not be called 'scientific'.

One of the ways to characterize theoretical knowledge is by recognizing that it is not directly related to practical problems. Theoretical knowledge may build upon knowledge from practical experience, but it is not pursued with the direct aim of solving practical problems. Somewhat aphoristically one may say that, while the purpose of practical knowledge is the control of action, the aim of theoretical knowledge is the control of knowledge itself. Theoretical knowledge emerges from the reflection on externally represented knowledge. Spoken language is the most obvious example of a means for the external representation of knowledge. Further examples are drawings, written language, and other symbol systems such as mathematical notation. This immediately explains why animals do not acquire theoretical knowledge; they lack any capacity for the cumulative external representation of knowledge. To be sure, many species do not simply act by stimulus and response but develop elaborate internal knowledge representations. All the same, their mental representations remain bound to the context of action, and there is no evidence of secondary reflection.[2]

In addition to theoretical knowledge there are of course other forms of knowledge. We may distinguish elementary and instrumental knowledge, both of which in some sense pre-

[1] Graham 1978.

[2] On this issue and the following distinction of forms of knowledge, see the discussion in Chapter 1.

cede theoretical knowledge. These different forms of knowledge are distinct in their sources, their inner structure, and their modes of transmission. *Elementary knowledge* is ontogenetically acquired, i.e., as an individual grows to maturity. Since the physical conditions of ontogenesis are largely culture-independent, a great part of this knowledge may be considered universal. An example of an elementary knowledge structure is what developmental psychologists refer to as the schema of an object. By this term they mean the mental construction of entities located in a definite place or moving along a definite spatial trajectory, independent of the self. Possession of the schema includes the ability to perceive objects as having a defined shape and size, regardless of from what changing point of view one sees them, and to know where to look for them when one has seen them vanish.

The relation between space and matter specifically does not first occur in the realm of theoretical reflection but appears as an inherent part of pre-theoretical thinking as a kind of elementary knowledge. In fact, conceptions of space and conceptions of material objects and the relation between the two co-evolve, and in this process space and objects become distinct from each other only gradually. As an illustration of this gradual process of separation consider the experiment in which the Swiss developmental psychologist Jean Piaget (1896–1980) interviewed children of different ages about the distance between two objects depending on whether or not a material barrier was placed between the two objects. A five-year-old child says about the two objects not separated by a barrier:[3]

> They're far apart.

But after the investigator has put a cigar box between the two objects, the child says:

> It isn't far, because there's a wall.

According to this child's conception, only the 'empty' space between the objects contributes to his perception of them as distant from each other; when the space is filled with material objects, there is no perception of distance.

From experiments of this kind, Piaget was able to conclude that only from an average age of about seven years is distance conceived as being independent from intervening material objects. The environment is then conceived as a huge receptacle in which material objects have their own distinct place. This place changes when objects are moved or are moving by themselves, but no two objects can ever be located in the same place at the same time. Piaget interprets the development of such a conception of the spatial environment as a result of the child's reflection on his or her interactions with the objects of the environment in question.

Piaget seems to assume that this development follows a universal pattern. This assumption is plausible as long as the reflection refers to experiences arising from actions within a universal environment, regardless of historical or cultural circumstances. As soon as specific tools or cultural practices in general are involved, one has to start from the assumption that knowledge structures are culturally dependent, since the handling of cultural artifacts and the performance of cultural practices imply the making of novel experiences. Such experiences lead to the acquisition of practical or, more specifically, *instrumental knowledge*, which is to a large extent expert knowledge acquired in the handling of artifacts such as measuring tools, mechanical instruments, and machines. An example for an instrumental knowledge

[3] Piaget, Inhelder, and Szeminska 1960, 75.

structure related to the practice of measurement is the additivity of lengths. Anyone who uses measuring rods or ropes knows implicitly that they can be apposed in order to measure lengths or distances greater than the measuring tool itself. This knowledge structure is independent of and may precede any general, or abstract, arithmetical knowledge the user may have.

The structures of elementary and instrumental knowledge do not necessarily find general and consistent expression at the level of linguistic representation. That only comes with theoretical reflection, which entails generality and consistency, giving rise to the appearance and use of abstract terms. *Theoretical knowledge* may thus be described as emerging from the systematic reflection on external representations of knowledge whereby the knowledge represented may be elementary, instrumental, or itself theoretical. We may distinguish different branches of theoretical knowledge according to form and representational type of the knowledge reflected upon. The systematic reflection on linguistic representations of elementary knowledge, for instance, brings about a branch of theoretical knowledge that may be described as *philosophy of space, time, and matter*, a prominent example being Aristotle's *Physics*; the systematic reflection on linguistic representations of instrumental knowledge brings about what is often referred to as *science*, such as the analytical concern with mechanical and optical phenomena; and the systematic reflection on symbolic representations of instrumental knowledge, including diagrams, brings about what may be identified as the origin of *mathematics*, most prominently Euclid's *Elements*.[4]

Since external representation of knowledge is a universal phenomenon in human cultures – there is, for instance, no human culture without a language – we may ask, is the presence of theoretical knowledge universal too? Does the presence of external representations of knowledge necessarily lead to theoretical knowledge? Based on historical and anthropological evidence the answer lies clearly in the negative. Historical evidence suggests that theoretical knowledge is something very late and very rare in human history. And even in periods and societies where we can document that it existed, it was often only tenuously and marginally maintained. The potential for reflection may, of course, always be realized by individuals. But for theoretical knowledge to become a historical force, the results of individual processes of reflection have to become collectively shared; they have to become part of an enduring tradition. The presence of external representations of knowledge is a necessary but not sufficient prerequisite for the emergence of theoretical knowledge. In addition there must be adequate societal and institutional support in the form of schools, academies, universities, libraries, etc. to sustain a cumulative tradition in which the intellectual achievements are preserved and perpetuated, be it orally or in written form.

Consider philosophical reflections on spatial concepts, which are the focus of this chapter. As evidenced by the various cultural techniques for spatial orientation and their linguistic representation, including the representation of spatial knowledge in mythologies, elementary spatial knowledge had existed in human history long before the advent of theoretical thinking. There are no sources from ancient Egypt or Mesopotamia that document a theoretical reflection on spatial language. The earliest texts documenting such theoretical reflection are found in ancient Greece, starting in the sixth century BCE with the Presocratic philosophers and culminating in the comprehensive Aristotelian natural philosophy. Other historical places that constituted a context for philosophical reflections about space were the

[4] For this third branch, see in particular Chapter 3.

neoplatonic schools of late antiquity, court science, philosophy and theology of the Arab Middle Ages, the scholasticism of the Latin Middle Ages, and early modern natural philosophy and practical mathematics. The philosophical activities at all of these historical places find their roots one way or another in Greek antiquity. This is most strikingly demonstrated by the central role of Aristotle's philosophy in Arabic, neoplatonic, and scholastic discussions of space, but can also be seen in early modern references to ancient Greek atomism.

Did then theoretical thinking about space emerge only once in history and then survive as a tradition? Or did it come about several times independently and was only accidentally influenced by earlier instances of spatial thinking? More broadly we may ask the following questions of the long-term development of spatial knowledge:

- What are the social, material, and intellectual conditions for the emergence of theoretical knowledge on space?
- To what extent is later thinking informed by the emergence of a tradition of such theoretical knowledge in antiquity?
- To what extent are similar structures in theoretical thinking on space the result of the influence of a single tradition, be it diachronically in a single culture or be it across cultures, and to what extent are such similarities an independent consequence of elementary and instrumental forms of thinking?
- What are the social, material, and intellectual conditions for the survival and perpetuation of a tradition of theoretical knowledge?

These are grand questions and addressing them obviously presupposes the comparison of different historical instances on theoretical thinking about space. In particular, historical instances that may be argued to be uninfluenced by the ancient Greek precedent are valuable objects of study in this context. Such instances are very hard to find. Traditions of theoretical reflection of ancient India and China may appear most promising in this respect. In fact, the text that will concern us here, the *Mohist Canon*, is one of very few sources from any culture that document theoretical thinking about spatial concepts independently from the Western tradition. It thus provides us with a particularly revealing and welcome independent source for approaching questions about necessity and contingency in the development of theoretical knowledge as those formulated above. Here we have made a first effort to interpret sections pertaining to spatial concepts within the framework outlined above. After a brief introduction to the text (4.2), we will discuss sections on space and matter (4.3), space and time (4.4), and instruments and arrangements (4.5). Finally, we shall discuss the epistemic status of Mohist spatial knowledge and argue that it provides an instance of theoretical knowledge parallel to that of Western philosophical considerations about space. At the same time, we shall point out differences between the two traditions and thereby take first steps towards addressing the fundamental questions raised above (4.6).

## 4.2 The *Mohist Canon*

The *Mohist Canon* is contained in four of the seventy-one chapters that make up the Mohist corpus, known generally simply as the *Mozi*. The corpus itself is a compilation of texts, perhaps of disparate origins, that dates in its transmitted form to about 300 BCE. Several centuries later it is ascribed by Han period scholars to what they identify as a 'Mohist school'. The period of the late fifth, fourth and third centuries BCE, known historically as the Warring

States Period, is distinguished for the richness of its intellectual ferment and for its growing social and political instability. Numerous texts from this period document the extensive concerns with what we would call social or moral philosophy. To the extent that these texts can be seen as constituting 'schools' of thought, they reveal how individual members of the learned classes competed for the attention of rulers across the land and for the consequent status that such attention promised. Argument, disputation and debate, aimed at influencing the ruling elite in matters of both political efficacy and social ethics, was the predominant enterprise of the day. Among these competing factions one group in particular, known traditionally as the Dialecticians (*biànzhě* 辯者), chose to argue from a perspective of logical provocation and an ostensible intellectual rigor, rather than couching their arguments in the more familiar terms of social morality, adherence to tradition, and political expediency.[5] The later Mohists are best known and best documented in the extant textual record for their systematically rigorous response to the Dialecticians. This response is what we find set out in those chapters of the *Mozi* that are known as the *Mohist Canon*.

The *Mohist Canon* deals directly with, among other things, spatial concepts and matters of mechanics and optics. The underlying motive for its compilation seems to have been a desire to set out a comprehensive model of terminological rigor and logical reasoning that could contribute to the Mohists' effective participation in the world of political, ethical and social disputation. In their effort to develop an objective, internally consistent, rigorous terminological scheme of their natural and technical environment the Mohists included not just descriptions, but strove to provide explanations as well. Their extended, probing analysis demanded the kind of thoughtful, reflective consideration that we call theoretical thinking.

Among sources from ancient China, the *Mohist Canon* is one of the most difficult to understand. The Mohist corpus overall contains more unknown graphs than most transmitted texts from the Warring States Period. This is likely due to the fact that it did not undergo as thorough a process of orthographic standardization in the course of its transmission as other texts, because it was not esteemed as a particularly literary work. The problem was compounded for the *Mohist Canon* because of its inherent difficulty. Furthermore, the text was garbled twice in the history of its transmission and has only become coherent and intelligible thanks to the work of twenthieth century scholars. Among these scholars, Liang Qichao (1873–1929) and A.C. Graham (1919–1991) stand out as having made exceptional contributions with their respective textual studies.[6]

The first two chapters of the *Mohist Canon* contain about 180 very short passages, the Canons proper, here designated 'C'. Two further chapters contain passages that were recognized by Liang Qichao, among others, in the early 1920s to be Explanations, here designated 'E', matching the Canons.[7] An Explanation is linked to its Canon by means of a head character, i.e., the first character of both canon and explanation is the same. The identity of head character in 'C' and 'E' turned out to be a crucial clue to the overall structure of the text. A Canon together with its co-ordinated Explanation we call a section. In our numbering of sections we follow Graham.[8]

---

[5] In Han-times (206 BCE – 220 CE) the Dialecticians were retrospectively designated as Nominalists (*míngjiā* 名家), i.e., as belonging to the 'School of Names'.

[6] Liang Qichao 梁啟超 1922; Graham 1978, reprint 2003.

[7] Liang Qichao 梁啟超 1922.

[8] Graham 1978. The Chinese text is as established in Boltz and Schemmel forthcoming, and though we are heavily indebted to Graham's pioneering textual work, our text may sometimes vary from that given in Graham 1978.

## 4.3 Magnitude, filling out, and interstice

Our analysis starts with section A 55.[9]

A 55

C: 厚，有所大也。

E: 厚：惟端無所大。

C: *hòu* 'having magnitude' means that there is something in relation to which it (i.e., the thing that has magnitude) is bigger.

E: *hòu* 'having magnitude': Only an **end-point** has nothing in relation to which it is bigger.

*Hòu*, which in everyday language means 'thick' (in the sense of a material, physical dimension), here implies spatial magnitude and is turned into an abstract term that can be used in other definitions or explanations. Thus, a later section reads:

A 65

C: 盈，莫不有也。

E: 盈：無盈無厚。於尺無所往而不得二。

C: *yíng* 'being filled out' is nowhere not having something.

E: *yíng* 'being filled out': Where there is no *filling out* there is no **magnitude**. On the measuring rod there is no place to which it extends such that you do not get both (i.e., filling out and magnitude).

The archaeological evidence for the *measuring rod* (*chí* 尺) shows clearly that it came to be a fixed, standard length of about 23–24 cm, typically subdivided into ten equal units. All the same, the word *chí* is used as a concrete way to refer to any short linear measure without necessarily specifying a fixed length.

In this section, the material aspect of the measuring rod appears to be crucial, since it represents the precondition for having magnitude. But 'filling out' may also directly be referred to attributes as the immediately following section shows. It introduces the meta-term *jiān bái* 堅白, 'hard and white', which was widely used in the disputational and philosophical texts of the Warring States Period.[10]. It stands for the co-occurrence of different and mutually pervasive attributes of a body, as in a stone that is both hard and white at the same time; either attribute may occur or not independently of the other. One can specify *jiān bái* 'hard-and-white' as the technical term for "the separation of distinct, but mutually pervasive properties."[11] It is defined, at first unexpectedly, among terms referring to spatial arrangements, because when understood literally, it refers to features that "fill out each other," that is, that are co-occurring or coincident.

[9]The 'end-point' (*duān* 端) occurring in the Explanation line is Graham's emendation; Graham 1978, 305. Here and in the following, terms that are defined in other sections than the one under consideration are marked in bold face.

[10]See the discussion in Graham 1978, 170–176.

[11]Graham 1978, 171.

A 66

C: 堅白，不相外也。

E: 堅白：異處不相盈。相非是相外也。

C: *jiān bái ‘hard-and-white’* is neither excluding the other.

E: *jiān bái ‘hard-and-white’*: (Attributes in general) when occurring in different places, do not **fill out** each other. When attributes are at odds with each other, this means they exclude each other.

The term *excluding* (*wài* 外) is to be understood primarily in terms of spatial exclusion but it also implies logical exclusion. The explanation states that attributes cannot be called *co-occurring* (*jiān bái* 堅白) if they are located on objects in different places, or if they are incompatible or *at odds with each other* (*xiāng fēi* 相非). In other words, the sense of *jiān bái* is delimited in two respects; it requires (a) spatial coincidence and (b) logical compatibility. It follows that for any two attributes to be in a *jiān bái* ‘hard-and-white’ relation they must be independent of each other.

As A 65 above suggests, the Mohist notion of space entails a dichotomy of ‘filled out’ versus ‘empty’. That section is part of a series reflecting the Mohist’s particular concern with the question of how each of these two features is used to define the other, most clearly illustrated in the following sequence of three sections (A 62, A 63, A 64) dealing with interstices.

A 62

C: 有間，不及中也。

E: 有間：謂夾之者也。

C: *yǒu jiān ‘having an interstice’* is (the sides) not joining at the **center**.

E: *yǒu jiān ‘having an interstice’*: refers to what flanks it (i.e., what flanks the interstice).

This section refers not simply to an ‘interstice’ (that is what we find in A 63), but to the object(s) in relation to which the interstice occurs. This may seem to be in some respects a subtle distinction, but it appears to be for the Mohist important.

A 63

C: 間，不及旁也。

E: 間：謂所夾者也。尺前於區穴而後於端，不夾於端與區穴。及及非齊之及也。

C: *jiān* ‘interstice’ is not reaching to the sides.

E: *jiān* ‘interstice’: refers to what is flanked. Measurements starting from an outline and ending at an **end-point** should not be considered as flanked by the **end-point** and the outline. Those two reachings are not equivalent reachings.

To be able to speak of an ‘interstice’ you need two flanking objects that are comparable in their capacity to be identified as boundaries of the interstice. Measuring from an outline with a measuring rod and considering the opposite end of the measuring rod as a flanking point does not define an interstice because on one side the measuring rod reaches the outline but on the other it "reaches" only to its own end-point. This is not a genuine 'reaching', hence the two reachings are not equivalent reachings.

The two sections A 62 and A 63 are complementary descriptions of the occurrence of an interstice and what defines an interstice. What remains to be described is the substance of an interstice, and for that the Mohists invoke a concrete example:

> A 64
>
> C: 櫨，間虛也。
>
> E: 櫨：虛也者兩木之間，謂其無木者也。
>
> C: *lú* 'king-post', the **interstices** are empty.
>
> E: *lú* 'king-post': What is empty is the **interstice** between two pieces of wood. It refers to the fact of having no wood.

The word *lú* 櫨 means a kind of 'rectangular piece of wood mounted on top of a pillar, as used, e.g., in the construction of a roof beam', what is technically known as a 'king-post'. It may be defined as "a structural member running vertically between the apex and base of a triangular roof truss" (see figure 4.1).[12]

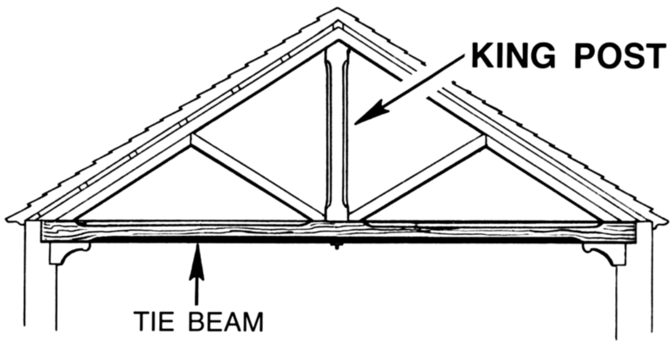


Figure 4.1: A king post.

The Mohist has recourse to this everyday object to illustrate the relation between an interstice and the material frame that forms it. This takes the understanding of 'interstice' one step beyond the descriptions of A 62 and A 63 in that it explicitly recognizes the interstice as 'empty' (*xū* 虛) relative to the material frame. The Explanation allows for the possibility that the interstice may be filled with a material other than that of the flanking objects.

[12]http://dictionary.reference.com accessed 18 October 2013. The image is taken from http://commons.wikimedia.org/wiki/File:King_post_(PSF).png, last accessed 9 January 2014. We owe the identification of the word *lú* 櫨 as 'king post' to Ian Johnston; Johnston 2010, 428–429.

## 4.4 Spatial extent and duration

We begin with the Mohists' definition of 'spatial extent'.

A 41

C: 宇，彌異所也。

E: 宇：東西蒙南北。

C: *yǔ* 'spatial extent' is spanning over different places.

E: *yǔ* 'spatial extent': east-and-west entails north-and-south.

What we translate as 'spatial extent' is in its more traditional context usually understood as 'celestial canopy', a word that generally carries cosmological overtones. Its concrete meaning is 'eaves' of a building, or more particularly, the space defined by the eaves. The sense of east-and-west "entailing" north-and-south is that the two directional spans are not separated from each other as independent manifestations of space, but are rather two different aspects or perspectives of a single comprehensive spatial extent.[13]

The verb *mí* 彌 here meaning 'to span, spread (over, out, through)' with respect to space, is used in a parallel way in the Canon line of section A 40 *jiǔ* 久 'temporal duration', i.e., 'temporal extent', the section that immediately precedes this one in the original Mohist order, given here next.

A 40

C: 久，彌異時也。

E: 久：今古合旦暮。

C: *jiǔ 'enduring'* is spanning different times.

E: *jiǔ 'enduring'*: 'present' and 'past' match 'dawn' and 'dusk'.

Just as *yǔ* 宇 'spatial extent' is expressed in A 41 as a 'span' stretching from one extreme to the other, so this section refers to the extension, or 'span', of time of a specific duration, here illustrated by the example of 'past' and 'present' as an abstract representation of the duration of time correlated with 'dawn' and 'dusk' as a concrete representation. Sections A 41 and A 40 seen in tandem suggest that the general sense of *mí* 'to span, spread (over, out, through)' is applicable both to space and to time.

The close relation that the Mohist sees between spatial extent and temporal duration also becomes clear in other sections. In particular, space and time are related in discussions of motion and rest.

A 50

C: 止，以久也。

E: 止: 無久之不止，當牛非馬。若矢過楹。有久之不止，當馬非馬。若人過梁。

C: *zhǐ 'remaining fixed'* means thereby enduring.

[13] Graham 1978, 294.

E: *zhǐ 'remaining fixed'*: The not-remaining-fixed that lacks duration corresponds to 'ox/ non-horse'; like an arrow passing a pillar. The not-remaining fixed that has duration corresponds to 'horse/non-horse'; like a person passing across a bridge.

'Remaining fixed' means 'fixed in place' and is inherently a durative phenomenon; there is no other possibility. But for the relation between 'remaining fixed' and 'not remaining fixed' there are two possibilities: (i) the 'remaining fixed' is durative and the 'not remaining fixed' is punctual or (ii) both are durative. The former is of the "ox/non-horse" type and is exemplified by an arrow passing a pillar, a momentary, punctual event. The latter is of the "horse/ non-horse" type and is exemplified by a person crossing a bridge, clearly a durative event. As the diagram in figure 4.2 shows, just as the set of 'horses' is a subset of the set of things that are 'non-oxen', but not all 'non-oxen' are 'horses', so the set of 'remaining fixed' phenomena is a subset of the set of 'durative' phenomena, but not all durative phenomena are fixed.

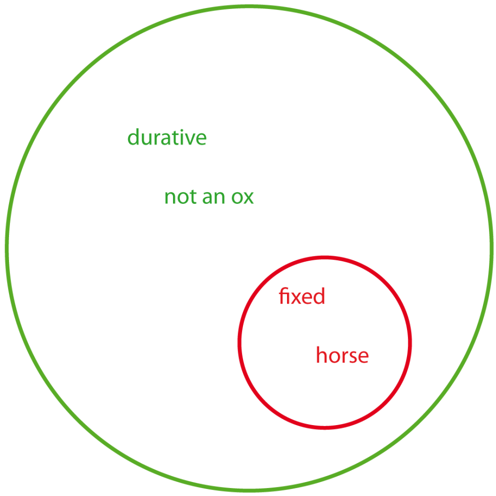


Figure 4.2: The relation between fixed and durative illustrated in terms of the set relation between 'horse' and 'not an ox'.

The text's image of "an arrow passing a pillar" is intended to represent the conjunction of 'not being fixed' and at the same time 'not being durative', since clearly a flying arrow is moving and just as clearly its passing a stationary point, here the 'pillar', is perceived as momentary and therefore not durative. Similarly, the image of a person crossing a bridge is just as obviously 'not fixed', and also clearly 'durative'. These two images, together with the original canon statement, represent all logically possible combinations of either fixed or not fixed with durative or not durative. The fourth combination, viz., fixed with not durative, entails a contradiction in terms and is not possible in actuality.

The relation between spatial extent and motion is further illustrated in B 13:

B 13

C: 宇或徙，說在長。

E: 宇:長，徙而又處宇。

C: **spatial extent**, (allows for) a shifting about somewhere. The explanation lies with 'expanding'.

E: *yǔ* 'spatial extent': as something expands and shifts about it then will occupy further **spatial extent**.

Space is here associated with a capacity for movement in some direction or another. This shows that spatial extent is not only spanning over different places, as explained in its defining entry A 41, but is a necessary aspect of motion and expansion. The immediately following section gives a characterization of the nature of the relation between the extent of space and the duration of time in an explicit, technically phrased statement:[14]

B 14

C: 宇久不堅白，說在<?>。

E: 宇:南北在旦，又在暮。宇徙久。

C: (The relation between) **spatial extent** and **temporal duration** is not of the **hard-and-white** type. The explanation lies with <?>.

E: *yǔ* 'spatial extent': South and north exist in relation to the dawn and also exist in relation to dusk. Within **spatial extent**, shifting about (entails) **temporal duration**.

The hard-and-white relation type (*jiān bái* 堅白) is defined as that relation in which one attribute may occur or not independently of the other (see above). But spatial extent exists in connection with the period of the dawn, and again separately in relation to the period of dusk. Furthermore, spatial extent is defined as that which allows for a shifting about (see B 13 above), and because shifting about entails temporal duration, spatial extent therefore has a dependent relation to temporal duration. So 'spatial extent' and 'temporal duration' are not independent attributes, but are inherently linked. Thus they are not of the hard-and-white type. Yet there is a hard-and-white type relation that holds between temporal and spatial concepts, as the following section shows.

B 15

C: 無久與宇堅白，說在因。

E: 無:堅得白必相盈也。

C: (The relation between) 'being without duration' and **spatial extent** is of the **hard-and-white** type. The explanation lies with the criterion.

E: *wú*: When the hard entails the white, each necessarily **fills out** the other.

[14] The question mark ('<?>') indicates a defective text.

The Explanation states that the hard-and-white relation type means that the two attributes are mutually pervasive, each attribute filling out the other, i.e., each is co-incident with (but independent of) the other. The fact of being mutually pervasive is the criterion referred to in the Canon. The relation between the absence of temporal duration, i.e., being temporally punctual, and spatial extent is said to be of this type. Section B 14 has just made clear that the relation between *yǔ* 宇 'spatial extent' and *jiǔ* 久 'temporal duration' is not of the hard-and-white type. We now have in a sense the complement to that, the relation between a 'point in time' (*wú jiǔ* 無久 'being without duration') and *yǔ* 宇 'spatial extent', which is said to be of the hard-and-white type. This implies that a single point in time is conceived of as filling out the whole of space, and in this respect the criterion of being mutually pervasive is met, yet neither of the two is contingent on the other; there is no dependent relation between spatial extent and a moment in time. At each moment in time there is a spatial extent being filled out by it and filling it out, somewhat anachronistically we may term them spaces of simultaneity. Different spaces of simultaneity (for instance the one existing at dawn and the one existing at dusk) are related by the shifting from one to the other, which entails duration, thereby establishing a dependent relation between temporal duration and spatial extent (B 14).

B 16

C: 在諸其所然，未然者。說在於是。

E: 在：堯善治，自今在諸古也。自古在之今，則堯不能治也。

C: Locating something in relation to where (temporally) it is properly so, or where (temporally) it has not yet become so. The explanation lies with being in relation to this (appropriate or inappropriate time).

E: *zài 'locating'*: "Yao is good at keeping order." This is, from a present perspective, locating it in the past. If one were, looking from a past perspective, to locate it in the present, then it would mean that Yao is not able to keep order.

The point seems to be that there is a non-arbitrary relation between events and time. Events are spatial occurrences and by the same token they occur over time. Therefore they are characterized as having both a 'spatial extent' (*yǔ* 宇) and 'temporal duration' (*jiǔ* 久), and this pairing is, according to B 14, not of the hard-and-white type. This means that the two features 'spatial extent' and 'temporal duration' as they pertain to events (such as Yao keeping order) are dependent in some way each on the other; events are temporally contingent and therefore are not independent of the time in which they occur; thus the example regarding Yao. When located in the proper time he is good at keeping order (an event that is historically recognized, even if legendary from a modern perspective), located in an inappropriate time, he is unable.

## 4.5 Instruments and arrangements

As the mention of the measuring rod in A 65 above indicates, we find, besides the reflection on elementary spatial knowledge, also reflection on the kind of instrumental knowledge acquired through the use of tools in the ordering of space. The following section, for instance, in which a circle is defined, reflects the use of the compass.

A 58

C: 圜，一中同長也。

E: 圜：規寫攴也。

C: *yuán* 'circle' implies (from) a single **center**, **being of the same length**.

E: *yuán* 'circle': When drawing with a compass, it is the simplest form.

'To be of the same length' and 'center' are defined in sections A 53 and A 54, respectively:[15]

A 53

C: 同長，以正相盡也。

E: 同：楗與框之同長也正。

C: *tóng cháng* 'being of the same length' means that by being laid straight (next to each other) each **exhausts** the other.

E: *tóng* 'the same': A door barrier-post and a door frame being of the same length is (an example of) being straight.

A 54

C: 中，同長也。

E: 中：自是往，相若也。

C: *zhōng* 'center' implies **being of the same length**.

E: *zhōng* 'center': extensions starting from this match one another.

The above definition of a circle (A 58) goes hand-in-hand with that of a rectangle in A 59 following.

A 59

C: 方，匡隅四雜也。

E: 方：矩見攴也。

C: *fāng* 'rectangle' implies that the frame corners number four and are closed up.

E: *fāng* 'rectangle': When drawing with a carpenter's square, it is the simplest form.

The Canon would seem to allow for any kind of quadrangle; only the Explanation by virtue of invoking the carpenter's square excludes all such that do not consist of only right angles. In normal parlance, of course, both the word *fāng* 方 and the word *kuāng* 匡 'square-frame basket' would only be used for rectangles.

In several sections on spatial arrangements of objects, the concept of a dimensionless end-point, which is introduced in section A 61, plays a constitutive role.[16]

[15]The term *jìn* 盡 'to be exhaustive' used in section A 53 is defined in section A 43 (which is not included in this selection) as meaning "that nothing is not so" (莫不然也).

[16]Note that the Chinese term *duān* 端 is used just as English 'end-point', to refer equally to the 'starting point' as well as the 'termination point' of a line or rod. A rod has two ends, a front end and a back end. Etymologically the word *duān* in fact suggests a beginning rather than an ending, as is explicitly indicated in this passage.

A 61

C: 端，體之無厚而最前者也。

E: [null.]

C: *duān* ‘end-point’ is the **element** that, having no **magnitude**, comes foremost.

E: [null.]

Not only do we have the notion of a dimensionless point, but that notion is analytically identified as a part of a network of specialized terminology, as the following passage illustrates.

A 2

C: 體，分於兼也。

E: 體：若二之一，尺之端也。

C: *tǐ ‘element’* is a part of a composite whole.

E: *tǐ ‘element’*: like one of two; an **end-point** on a measuring rod.

A *tǐ* 體 ‘element’ is not just an accidental or random part of a whole, like a piece of broken chalk, but is a ‘separable component’ of an analyzable whole. The word *tǐ* is cognate with the word *lǐ* 豊 ‘ritual vessel’ and by extension with homophonous *lǐ* 禮 ‘ritual, ceremony’. The semantic implication is that just as a *lǐ* 豊 ‘ritual vessel’ is a meaningful physical component with a precise, well-defined position and function in a *lǐ* 禮 ‘ritual or ceremonial performance’ (cf. *zhì* 豒 ‘the proper order or sequence of ritual vessels in a ceremonial performance’), so a *tǐ* 體 ‘element’ is a meaningful component in any composite whole, whether concrete or abstract, of a quotidian, non-ceremonial nature.

The Mohists recognize four different linear relations illustrated by the arrangement of two measuring rods, all dependent on the concept of a dimensionless end-point as identified in A 61 above: (i) extending to an equal length in opposite directions from a common end-point (A 60), (ii) overlapping (A 67), (iii) lying side by side to allow comparison (A 68), and (iv) being contiguous (A 69).

A 60

C: 倍，為二也。

E: 倍：二，尺與尺俱去一端，是無同也。

C: *bèi* ‘doubling’ is making two.

E: *bèi* ‘doubling’: ‘two’ means a measuring rod together with another measuring rod both extending (linearly) away from a single **end-point**, in this case (i.e., the case of doubling), they will have no shared portion.

The general notion of ‘doubling’ is illustrated very concretely in linear terms by explaining that two identical measuring rods laid end-point to end-point (in a straight line) such that there are no coincident points will give a doubled length.

A 67

C: 攖，相得也。

E: 攖：尺與尺俱不盡，端與端俱盡，尺與端或盡或不盡。堅白之攖相盡，體攖不相盡。

C: *yīng ‘overlapping’* means each entailing the other.

E: *yīng ‘overlapping’*: is when a measuring rod is put together with another measuring rod such that neither is **exhausted**, or when an **end-point** is put together with another **end-point** such that both are **exhausted**, or when a measuring rod is put together with an **end-point** such that one is **exhausted** and one is not. When attributes of the **hard-and-white type** (*jiān bái*) overlap they **exhaust** each other. When **elements** (by contrast) overlap they do not **exhaust** each other.

This section shows *tǐ* 體 ‘element’ as part of the Mohist’s specialized terminology used to establish a distinction between two different kinds of ‘overlapping’. The first example of the Explanation depicts ‘overlapping’ in the most straightforward way, one thing partially coinciding with another. The ‘overlapping’ of independent and coinciding attributes, i.e., attributes of a *jiān bái* type by contrast must by definition be exhaustive because they “fill out” each other, just as the overlapping of two end-points will be exhaustive. Similarly, the two elements (*tǐ* 體) referred to in the last phrase of the Explanation must be elements of a single object, and their overlapping corresponds to the overlapping of the two measuring rods of the first line, except now we see that an ‘element’ is understood in an abstract sense, just as *jiān bái* is the abstract counterpart to the end-point.[17]

A 68

C: 仳，有以相攖，有不相攖也。

E: 仳：兩有端而後可。

C: *bǐ* ‘side-by-side comparing’ means that there is a part where (two things) **overlap** one with the other and a part where they do not **overlap**.

E: *bǐ* ‘side-by-side comparing’: Only when the two have a (coincident) **end-point** is this possible.

It is possible, of course, to lay two measuring sticks side by side such that they partially overlap and partially do not, but unless they are positioned such that one end of one of them coincides with an end of the other, there is no meaningful comparison. The explanation of the canon here makes it clear that *bǐ* ‘side-by-side comparing’ must be of this ‘coincident end-point’ type.

A 69

C: 次，無間而不相攖也。

E: 次：端無厚而後可。

[17] Section A 2 exemplified a *tǐ* ‘element’ as an ‘end-point’, yet the overlapping of two end-points cannot be the same thing as the overlapping of two elements, since both elements must belong to a single object, and it is impossible that two end-points of a single object could ever overlap.

C: *cì* 'contiguous' is having no **interstice** but not **overlapping** one with the other.

E: *cì* 'contiguous': Only because the **end-point** has no **magnitude** is this possible.

This section shows that the notion of contiguity is possible only because end-points are without magnitude, i.e., dimensionless. Were that not the case, there would have to be either an interstice or an overlapping.

Further sections that may well be related to instrumental knowledge are A 52, A 56, and A 57. Their relation to the use of instruments remains a conjecture because there are no extant Explanations to the Canons.

A 52

C: 平，同高也。

E: [null.]

C: *píng* 'being level' means being of the same height.

E: [null.]

While it remains questionable if this passage is related to the use of leveling instruments, the following two passages are probably related to the use of gnomons.

A 56

C: 日中，正南也。

E: [null.]

C: *rì zhōng* 'the Sun at the center' is being due south.

E: [null.]

A 57

C: 直，参也。

E: [null.]

C: *zhí* 'to be straight' is to be in alignment.

E: [null.]

In the case of A 56, the 'center' refers to the mid-point on the Sun's trajectory between rising and setting, which would have been determined with a device such as a gnomon or sundial. 'To be in alignment' *cān* 参 is the standard term in Chinese astronomy for aligning two gnomons with an observed heavenly body.[18] Given the astronomical context of A 56, the reference to astronomical practice in A 57 seems plausible.[19]

[18] Graham 1978, 307.

[19] Beyond this, *cān* 参 refers to the three stars of the constellation Orion that in their linear arrangement are identified as Orion's 'belt'.

## 4.6 The epistemic status of Mohist spatial knowledge

We have claimed that the spatial knowledge documented in the *Mohist Canon* presented in the foregoing sections results from systematic reflections on the linguistic representation of elementary and instrumental knowledge and therefore constitutes a genuine case of theoretical knowledge. Let us now analyze the reflective character of this knowledge in order to corroborate this claim and to understand better how the different forms of knowledge interact and thereby shape the theoretical knowledge.

First of all, the representation of knowledge in the *Mohist Canon* clearly documents second order knowledge, i.e., knowledge resulting from reflections on the representation of knowledge. Thus, the majority of sections we encountered can be identified as definitions, statements that delineate the meaning of specific terms, which are then consistently used. The network of defined terms used in the sections discussed in this chapter is shown in Figure 4.3.

By their participation in a network of definitions, the terms become technical and, to different degrees, abstract. This is an important aspect of the transformation of meaning that takes place when concepts structuring elementary and instrumental knowledge are transferred to the realm of theoretical knowledge. While fundamental aspects of the relevant cognitive structures may be preserved in such transformations, theoretical knowledge inevitably brings about meanings alien to elementary and instrumental knowledge.

Let us, by way of example, look more closely on the relation of space and matter. As explained in the introductory section, within elementary knowledge, space and matter are inherently related ideas. Spatial concepts such as that of distance only gradually become separated from the material fillings of a space, such as an interstice between two bodies. In particular, ideas about being empty and being filled may have an impact on the perception of the extent of an interstice.

How does the relation between space and matter translate into theoretical knowledge? In the case of the *Mohist Canon*, we have a pair of concepts, *hòu* 厚 'having magnitude' (being extended) and *yíng* 盈 'filling out', that consistently differentiate the material and the spatial aspects of bodies. These are the terms defined in sections A 55 and A 65. While we have seen that the distinction between spatial and material aspects of bodies emerges in elementary knowledge, the systematic separation of the two and the reflection on their relation is clearly an aspect of theoretical thinking. Thus, the Explanation provided for the definition of 'having magnitude' refers to the *duān* 端 'end-point', a theoretical entity defined in section A 61. And the Explanation for the definition of 'being filled out' (A 65) shows that magnitude is an inherent feature of physical objects and states that spatial magnitude cannot occur without a material filling out.

In a similar manner, sections A 62 and A 63 differentiate *yǒu jiān* 有間 'having an interstice' and *jiān* 間 'interstice'. The Explanation for the definition of 'interstice' clearly demanding that the flanking things that *have* the interstice are material: the interstice, which may be gauged by means of a measuring rod, reaches from the outline of one such flanking object to that of the other. (The end-point of a measuring rod cannot be taken as the other extreme of an interstice, as the Explanation of A 63 makes clear.) Section A 64 then relates the concept of interstice to that of emptiness, stating that the interstice being empty refers to its lack of the material the flanking objects are made of.

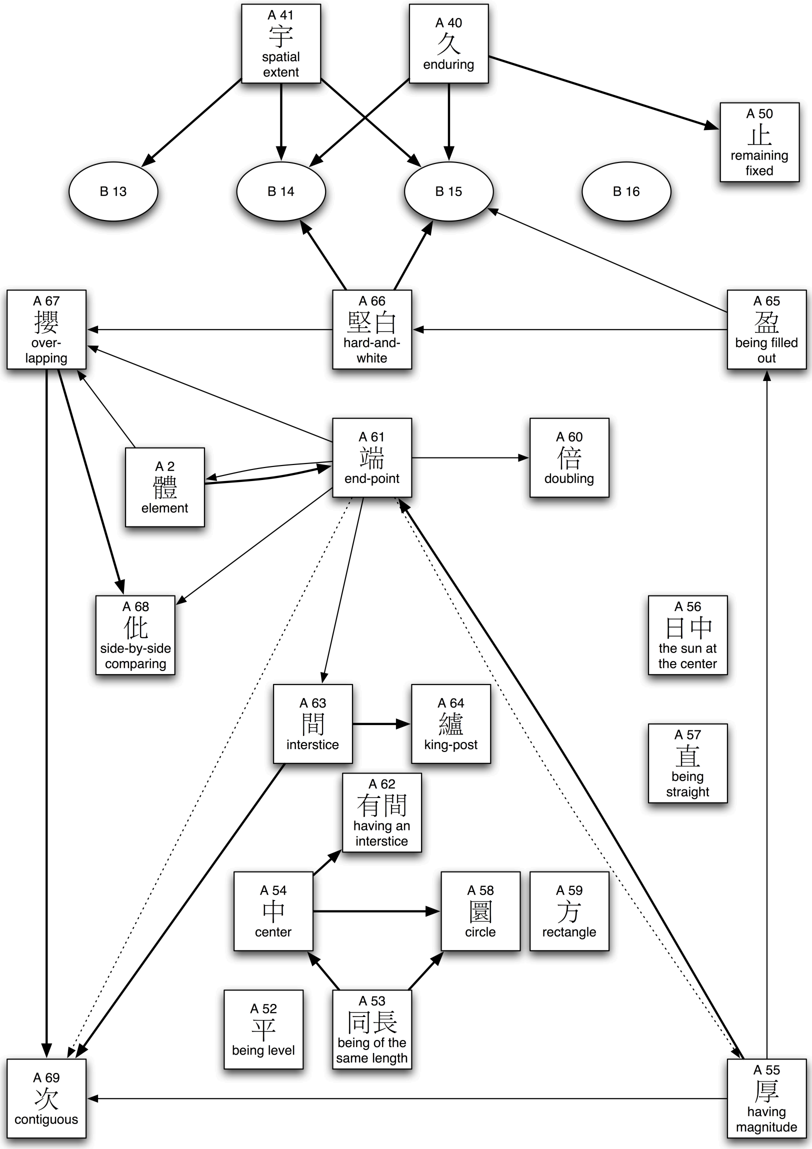


Figure 4.3: Terminological relations between sections on space, time and matter. Definitions are represented by squares, propositions by ovals. A bold arrow indicates that a defined term is used in the Canon of another section, a thin arrow that it is used in the Explanation. Dotted arrows indicate that the occurrence of the term is only conjectural.

The Mohist statement (A 65) that being filled out is a necessary precondition to having magnitude is reminiscent of Western theories of space and matter that claim that extension is a property of bodies alone, not of an alleged space independent of bodies. In a certain way, all theories that hold that space is nothing but an aspect of body maintain this view. Aristotle, for instance, extensively discusses the idea of the void as a place from which all bodies have been removed, and concludes that such void cannot exist, thereby refuting ideas about space formulated by the atomists Leucippus and Democritus.[20] A particularly radical version of this view is found in Descartes' claim that body and space are only two aspects of the same and that the walls of a vessel would be contiguous if the vessel were empty in the philosophical sense, since between its walls there would be nothing.[21]

But is the Mohist statement actually referring to such a world view, denying extension where there is no bodily filling? The every-day meaning of the term here translated as 'magnitude' (*hòu* 厚) and defined in A 55, 'to be thick', suggests that this is really about the magnitude of material objects, not about the question if the abstraction of extension still makes sense when what is abstracted from are bodies in general. In other words, it appears that the Mohist text is actually concerned with the clarification of the use of words, rather than making a claim about the existence or non-existence of space as an entity independent of bodies. If this interpretation is correct, A 65 merely states that the word 'magnitude' applies only where there is body ('filling out'). This interpretation is corroborated by the fact that a term potentially referring to spatial extension without regarding the bodies filling space is given elsewhere in the text: the 'spatial extent' (*yǔ* 宇) of section A 41. After all, this 'spatial extent' is defined as spanning over different places, not over bodies. It therefore appears amenable to a concept of space abstracted from all bodies, but this latter abstraction is also nowhere made explicit in the text.

Correspondingly, the canon A 65 on 'being filled out' seems not so much to introduce a universal material plenum, but rather to aim at complementing the immediately preceding canon dealing with the empty interstices characteristic of the structural functioning of a *lú* 櫨 'king-post'. The 'interstice' is a spatial extension described as lacking a given material, i.e., it is the part that has no wood and therefore is said to be *xū* 虛 'empty'. 'Magnitude', by contrast, is a spatial extension that is always accompanied by some material 'filling out'. The view implied by the Mohist definitions allows for the co-occurrence of an interstice and a magnitude, in that the material between the flanking objects defining the interstice can have magnitude.

It seems that there was no need for the Mohist to position himself in an argument about whether the world was a plenum or whether a perfect void existed. From all we know, such debate of physical world views was indeed absent from the disputations in Warring States China. Thus, the *Mohist Canon* shares with Aristotle's *Physics* a concern with the consistent use of terminology, and both texts particularly deal with spatial terms in this context. Accordingly, in both texts we can discern elementary structures of spatial knowledge, such as that differentiating the materiality from the extension of a body. In Aristotle there is the additional concern about the correct natural philosophy. Aristotle explicitly refutes not only what he considers errors of argumentation, but world views that he rejects, such as atomism. In the Mohist case there are no such world views either expressed or rejected.

The discursive context of the *Mohist Canon* is not so much related to systems of natural philosophy but to rules for consistent reasoning in general. This context is reflected not only in the sections on concepts of knowledge, reasoning, and moral conduct, but also in those

[20] Aristotle *Physics* IV, 8.

[21] Descartes 1984, 47–48 (Descartes, *Principles of Philosophy*, Part 2, § 18).

on spatial, mechanical, and optical terms.[22] In the case of spatial terminology this relation becomes particularly clear from the central role of the term *jiān bái* 堅白 'hard and white', which is used as a technical term in Warring States disputations. The definition of the term in A 66, in particular, reflects the close entanglement of logical and spatial arguments when the term *wài* 外 'excluding' is used in a spatial and a logical sense at the same time. Attributes are said to be of the 'hard and white' type when they fill out each other and are compatible, i.e., they spatially coincide and do not logically exclude each other. In the Aristotelian tradition, attributes pertain to bodies, or substances, while these bodies or substances then occupy a certain place.[23] Logical and spatial exclusion are discussed separately. No substance can have mutually exclusive attributes, and no two substances can be in one and the same place at the same time.[24] In the Mohist text, the argument appears to be that contradictory attributes cannot be in the same place. The section thus reflects the elementary knowledge structure of the schema of an object, i.e., no two objects can be in the same place at the same time, but it does so not by referring to some notion of an impenetrable body, but by the observation that contradictory attributes cannot exist unless in different places.

Besides a concern with the relation between spatial and material concepts, the Mohist text reflects on the relation between the concepts of spatial extent and temporal duration. The Mohist definitions of spatial extent and temporal duration (A 41 and A 40, respectively) are constructed in parallel. The use in both cases of the verb *mí* 彌 'to span, spread (over, out, through)' clearly indicates that the Mohist conceives of space and time as comparable in that both are extended. The peculiar use of the verb *zài* 在 'to locate' in a temporal context in B 14 and B 16 underlines this parallelism.

Extension is arguably the most basic structural similarity between space and time.[25] More generally, there is strong evidence that a certain parallelism between spatial and temporal concepts is a universal aspect of elementary knowledge. Spatial metaphors used for temporal designations in everyday language, for instance, are a cross-linguistic phenomenon.[26] It is a typical aspect of theoretical reflection that such structural parallelism within elementary knowledge becomes explicitly addressed on the level of technical terminology. A parallel case to the Mohist passages can again be found in Aristotelian discussions of space and time.[27]

The Mohist theoretical reflection on the relation between spatial extent and temporal duration again makes use of the concept of 'hard and white'. Thus, spatial extent and duration are said not to be of the 'hard and white' type (B 14). The reason is that they are not independent. Motion is invoked as an argument for this dependence: shifting about implies the occupation of further space (B 13) and takes time (B 14). As a matter of fact, section B 14 seems to suggest the possibility that exemplars of spatial extent can shift through time, *viz.*, the north-south extent from one instant (dawn) in time to another (dusk). Spatial extent

[22] Graham 1978; Renn and Schemmel 2006, Boltz 2006; Boltz and Schemmel 2013.
[23] Thus, according to Aristotle's *Categories*, for instance, quality and place are two different ways of predicating that which exists; see Rapp 2001, 82.
[24] This becomes clear from Aristotle *Physics* IV, for instance at 209a, 7–8 (Aristotle 1993, 282).
[25] Galton 2011.
[26] See, for instance, the recent discussion in Evans 2013. For evidence that the parallelism between space and time is not only a linguistic, but a cognitive, phenomenon, see, for instance, Boroditsky 2000 and Casasanto and Boroditsky 2008.
[27] Aristotle, for instance, describes time and space (place) as quantities related by the fact that they are both continuous, an attribute that presupposes extension; *Categories* 4b, 24–25 (Aristotle 1983, 36).

and lacking duration, by contrast, are said to be as 'hard and white' (B 15), since an instant fills out the spatial extent and vice versa.

While the particular form of the argument is specific to its cultural context, exemplified by the central role of the analytic tool of 'hard and white', there are structural commonalities to the spatio-temporal reasoning documented in the Western tradition. The idea that spatial and temporal magnitudes are related by motion, for instance, is also found in ancient Greek philosophy. As an example we may refer to Aristotle's discussion of the speed of local motion, in which the time of a motion is related to the space traversed.[28] Again, there is evidence that the connection of temporal and spatial measures via motion precedes theoretical thinking. In fact, the separation of the temporal from the spatial order in the consideration of motion is only gradually achieved in the course of ontogenesis.[29]

Despite the parallelism between space and time, there is an asymmetry in their relation as described by the Mohist. It is of spatial extent and lack of duration that the Mohist claims the relation to be of the 'hard and white' type, but not of duration and lack of spatial extent. Thus, while one instant in time fills out all spatial extent, the inverse seems not to be the case (a spatial point filling out all of time). Therefore it is instances of spatial extent that shift through time. The asymmetry may be explained by the fact that within spatial extent, motion is conceivable as well as rest. In time, by contrast, there is no rest, spatial extent and all it comprises inevitably move from one instant to the next. This attribute of time, which is not an attribute of space, has been described as *transience*.[30] In his *Physics*, Aristotle addresses this aspect of time when he "speaks of the now as progressing through time in a way comparable to that of a body progressing through a movement […]."[31]

While the concept of an instant or a 'now' has a clear enough sense in elementary thinking, in the realm of theoretical reflection it may become problematic when related to the concepts of motion and rest. In Zeno's famous paradox of the flying arrow, this problematic relation is employed when it is argued that the arrow cannot move during an instant and therefore cannot move at all. Aristotle tries to resolve this paradox by arguing that, in the 'now', there is neither motion nor rest.[32] In the Mohist case, the discussion of the instant, *wú jiŭ* 無久 'lacking duration', implies that it is compatible with *bù zhĭ* 不止 'not remaining fixed', which, for the Mohist, is equivalent to being in motion, as the example of an arrow passing a pillar suggests. It is incompatible with *zhĭ* 止 'remaining fixed', since, according to A 50, being fixed demands duration. So, while Aristotle responds to the problem by denying instantaneous motion and rest, the Mohist responds otherwise. This shows that what seems intuitively obvious at the elementary level becomes problematic at the theoretical level.

Just as the the everyday concept of an instant becomes refined in the context of theoretical thinking about motion and rest, the everyday concept of an end-point becomes refined in the context of theoretical thinking about the possible arrangement of measuring rods. Sections A 60 and A 67–69 explaining different spatial arrangements of measuring rods all rely, in one way or another, on the definition of the end-point. In A 67 we see the consideration of all possible two-item combinations of an end-point and a measuring rod, including the intuitively least obvious case of two coincident end-points. The Explanations in A 68 and

[28] *Physics* 232a, 23 – 232b, 15 (Aristotle 1993, 103–115). See further *Physics* IV, 11.
[29] Piaget 1946, Chapter 3.
[30] Galton 2011.
[31] Owen 1976, 15; the passage referred to is *Physics* 219b, 22–33.
[32] *Physics* 239b, 1–2.

A 69 both explicitly say that the configuration specified in the Canon is only possible (*ér hòu kě* 而後可) because of the particular nature of the dimensionless end-point. The definition of the end-point as something dimensionless (A 61) is clearly a result of its role in the network of concepts and can only be formulated within this network. Despite its derivation from an instrument of practical relevance – the measuring rod – the Mohist 'end-point' is therefore a typical theoretical entity. The end-point's lack of extension is conceived of as absolute, which shows that the concept does not reflect an elementary experience or a concrete perception, but a reflection on the linguistic representation of instrumental actions.

In the context of reflections on instrumental knowledge, the Mohist defines further geometrical objects such as the circle or the rectangle. Some of the Mohist geometrical definitions are strikingly reminiscent of parallel definitions in Euclid's *Elements*. Thus, Euclid defines a point as "that which has no part," and a circle as[33]

> a plane figure contained by one line such that all the straight lines falling upon it from one point [later called the center] among those lying within the figure are equal to one another[.]

The similarity of this with the Mohist definition of a circle (A 58), definitions that were certainly arrived at independently, may be explained by the similarity of the underlying practical knowledge. In both societies (Warring States China and Classical Greece), the compass, to which the Mohist Explanation of A 58 makes explicit reference, was a well-known instrument. Despite this similarity in the definitions, there is no counterpart found in the Mohist text to the Euclidean propositions. The *Mohist Canon* documents reflections on the linguistic representations of instrumental knowledge, but not on their symbolic or diagrammatic representation, such as the construction of complex figures that can be drawn with straightedge and compass. This means it is more philosophical than mathematical, and thus more Aristotelian than Euclidean, in the sense described at the outset.

The near-simultaneous but independent appearance of texts documenting theoretical thinking in Greek and Chinese antiquity raises the question how we might account for this coincidence. Are there identifiable factors that led to this development? This question becomes all the more interesting and all the more consequential when we recognize that the appearance of texts clearly representative of theoretical thinking is a markedly uncommon phenomenon in the ancient world. Whatever form a complete answer to this question might eventually take, here we can observe that both cultures, Greek and Chinese, had thinkers who characteristically constructed paradoxes as inherent parts of their arguments, the Sophists in Greece and the Dialecticians in China.[34] The dynamics of disputation resulted in both cases in a tendency to establish comprehensive doctrinal systems using consistent terminology.

Similarities in the independent reflections on spatial concepts in ancient Greece and China can, as we have seen, at least in part be explained by similarities in the elementary and instrumental knowledge reflected upon. From a Western perspective, the proximity of passages related to such diverse issues as ethics, logic, mechanics, optics, and geometry within a small text as presented by the *Mohist Canon* appears peculiar. At the same time, other fields of contemporary knowledge such as astronomy play a marginal role at best.

[33] Euclid 1956, I, 153.

[34] Beyond our concern here with cultures of disputation in China and Greece, such things as political fragmentation and the emergence of city-states, social upheaval and increased social mobility, and the flourishing of arts, crafts, and the technology of warfare all would likely be pertinent to a full account of this development.

Clearly, what knowledge is regarded relevant for a given text or textual tradition, and what knowledge is disregarded, may vary considerably among different societies and depends on the way the knowledge is shared by different societal groups.

Another difference we can observe between the Later Mohists' reflections and their Greek counterparts is that in the Chinese case there seems to be no urge to explain all of nature through certain fundamental principles, mechanisms, or elements, or to formulate encompassing natural philosophies. As concerns possible origins of this disparity between the Aristotelian and the Mohist reflections on spatial terms, the most direct cause appears to be a difference in the timing of the emergence of different types of theoretical debate. In the Greek case, the construction of cosmologies and systems of the natural world reducing all appearances to a small set of principles or elements precedes the meta-reflection about language and knowledge. The presocratics constructed competing world views of this kind long before meta-reflection arises with, or around the time of, Parmenides.[35] In the Chinese case, on the other hand, the Mohist meta-reflection precedes the establishment of comprehensive cosmologies like the Yin-Yang 陰陽 and Five-Agents (*wǔxíng* 五行) systems by several centuries. There may have been elements of these systems already present around the time of the Later Mohists, but not constituting any coherent, encompassing system. This developed only in Han times when the Mohist tradition of linguistic reflection had already lost its impetus.[36]

Finally, a notable difference that renders comparison difficult is the small size of the Chinese text corpus pertinent to theoretical reflections on space. While in Aristotle alone there are whole books devoted to the analysis and discussion of spatial concepts, in the Chinese case we mainly have the very short and very few sections that are part of the *Mohist Canon*. Furthermore, the favorable conditions for the Mohist type of reflections seem to have vanished in later times. In particular, the radical change of conditions after the foundation of the centralistic Qin empire (221 BCE) appears to have cut off this tradition. Accordingly, the Mohist deliberations never entered the mainstream of the Chinese knowledge tradition and for this reason lack the exegetic scrutiny and contextualization provided by later commentary.

## Bibliography

Aristotle (1983). *The Categories. On Interpretation. Prior Analytics*. Aristotle in twenty-three volumes. Loeb classical library. Cambridge, MA: Harvard University Press.

— (1993). *The Physics*. Aristotle in twenty-three volumes. Loeb classical library. Cambridge, MA: Harvard University Press.

Boltz, William G. (2006). Mechanics in the 'Mohist Canon': Preliminary Textual Questions. In: *Studies on Ancient Chinese Scientific and Technical Texts: Proceedings of the 3rd ISACBRST*. Ed. by Hans Ulrich Vogel, Christine Moll-Murata, and Gao Xuan. Zhengzhou: Daxiang chubanshe, 32–40.

Boltz, William G. and Matthias Schemmel (2013). *The Language of 'Knowledge' and 'Space' in the later Mohist Canon*. Preprint 442, Max Planck Institute for the History of Science, Berlin.

35 Schiefsky 2012, 193–194.

36 On the early Chinese tradition of cosmos-building, see Graham 1989, 315–370.

Boltz, William G. and Matthias Schemmel (forthcoming). *Theoretical Knowledge in the Mohist Canon*. Berlin: Edition Open Access.

Boroditsky, Lera (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition* 75(1):1–28.

Casasanto, Daniel and Lera Boroditsky (2008). Time in the mind: Using space to think about time. *Cognition* 106:579–593.

Descartes, René (1984). *Principles of Philosophy*. Synthese Historical Library 24. Dordrecht: Reidel.

Euclid (1956). *The Thirteen Books of Euclid's Elements*. 2nd edition, revised with additions. Ed. by Heath. New York: Dover.

Evans, Vyvyan (2013). Temporal frames of reference. *Cognitive Linguistics* 24(3):393–435.

Galton, Antony (2011). Time flies but space does not: Limits to the spatialisation of time. *Journal of Pragmatics* 43:695–703.

Graham, Angus Charles (1978). *Later Mohist Logic, Ethics and Science*. Hong Kong: Chinese University Press.

— (1989). *Disputers of the Tao: Philosophical Argument in Ancient China*. La Salle: Open Court.

Johnston, Ian, ed. (2010). *The Mozi: A Complete Translation*. New York: Columbia University Press.

Liang Qichao 梁啟超 (1922). *Mojing jiaoshi* 墨經校釋. Shanghai: Shangwu 商務.

Owen, G. E. L. (1976). Aristotle on time. In: *Motion and time, space and matter: interrelations in the history of philosophy and science*. Ed. by Peter K. Machamer and Robert G. Turnbull. Columbus: Ohio State Univiversity Press, 3–27.

Piaget, Jean (1946). *Le développement de la notion de temps chez l'enfant*. Paris: Presses Universitaires de France.

Piaget, Jean, Bärbel Inhelder, and Alina Szeminska (1960). *The Child's Conception of Geometry*. Digital reprint 2007. Abingdon: Routledge.

Rapp, Christof (2001). *Aristoteles zur Einführung*. Hamburg: Junius.

Renn, Jürgen and Matthias Schemmel (2006). Mechanics in the Mohist Canon and Its European Counterparts. In: *Studies on Ancient Chinese Scientific and Technical Texts: Proceedings of the 3rd ISACBRST*. Ed. by Hans Ulrich Vogel, Christine Moll-Murata, and Gao Xuan. Zhengzhou: Daxiang chubanshe, 24–31.

Schiefsky, Mark (2012). The Creation of Second-Order Knowledge in Ancient Greek Science as a Process in the Globalization of Knowledge. In: *The Globalization of Knowledge in History*. Ed. by Jürgen Renn. Berlin: Edition Open Access, 191–202.

# Chapter 5
# Cosmology and Epistemology: A Comparison between Aristotle's and Ptolemy's Approaches to Geocentrism

*Pietro Daniel Omodeo and Irina Tupikova*

## 5.1 Introduction

Our present discussion on cosmological models and epistemologies is a comparison of the different argumentative strategies employed by Aristotle and Ptolemy in their approaches to geocentrism through an analysis of their discussion of the centrality of the Earth in *De caelo* II, 13–14 and *Almagest* I, 3–7. The divergence not only concerns secondary issues but rather the gnoseology underlying the theories of these two authors, and this affects also the meaning of theses on which they apparently agree. As we shall argue, this difference potentially entails momentous consequences concerning the justification and the acceptance of fundamental astronomical concepts.

The epistemological distance between the two main 'authorities' of classical cosmology already challenged authors of Late Antiquity such as Aristotle's commentators Simplicius and Philoponus. The issue was intensively debated during the Middle Ages and the Renaissance.[1] It was even crystallized as a disciplinary separation between the academic teaching of 'physical' astronomy (that is, the doctrine on the heavens from a natural-philosophical perspective) and 'mathematical' astronomy.[2] Only the former was deemed to provide the causal explanation of natural phenomena *per se* and basically rested on Aristotle's philosophy of nature (his theory of motion, of natural places and of cosmological order) and Aristotle's acceptance of Eudoxus's and Callippus's concentric-spheres model of the cosmos.[3] In twelfth-century Moorish Spain, Ibn Rushd (better known by his Latinized name Averroes) denounced the discrepancy between the homocentric heavenly mechanism propounded by Aristotle in *Metaphysics*, XII (or Λ) and *De caelo*, on the one hand, and the mathematical devices (epicycles, eccentrics and equants) employed by Ptolemy.[4] He therefore accused

[1] The standard reference for this issue is Duhem 1969, although the author's attempt to interpret the history of physics and astronomy from the perspective of twentieth-century epistemology (in particular conventionalism) is completely outdated. For some criticism of Duhem's anachronism see for instance Barker and Goldstein 1998. In another paper, Goldstein has convincingly argued that "saving the phenomena" in antiquity, and according to Ptolemy in particular, did not mean limiting astronomical consideration to computational hypotheses as merely conventional. Quite the opposite: it often meant "to seek an underlying orderly reality that can explain the disorderly appearances that are a kind of illusion," as was the case with Geminus. In the case of Ptolemy, moreover, "the phenomena are 'real' and not illusions, for they are the criteria by which the models are judged, not the other way round" (Goldstein 1997, 8).

[2] See Dijksterhuis 1986, 61–68.

[3] The classic treatment of this issue is Schiaparelli 1875. For a reassessment, see Heglmeier 1996.

[4] Theokritos Kouremenos has recently argued against Aristotle's adherence to a homocentric world system of concentric material spheres as the real physical structure of the world (Kouremenos 2010). Be that as it may, relative to the original intentions of Aristotle, the Eudoxan interpretation of his cosmology on the basis of *Metaphyisics* XII

Ptolemaic astronomy of being at odds with natural philosophy since it renounced physical tenability for computational convenience. His contemporary al-Bitruji (known by his Latinized name Alpetragius) even sought to reform mathematical astronomy in accordance with homocentrism, that is, he reduced all celestial motion to a mechanism of concentric spheres. His book on heavenly motions, translated into Latin by Michael Scot as *De motibus caelorum*, had a great impact in Christian Europe up to the Renaissance.[5] It should be noted that it was republished in 1531 in Venice shortly before analogous works of Paduan Aristotelians appeared: Giovan Battista Amico's *De motibus corporum corporum coelestium iuxta principia peripatetica sine eccentricis et epicyclis* (*On the Motion of Heavenly Bodies in Accordance with Peripatetic Principles, that is, without Eccentrics and Epicycles*, 1537 and Paris 1540) and Girolamo Fracastoro's *Homocentrica sive de stellis* (*Homocentrics, or on the Stars*, 1538).[6]

In spite of this well known criticism of Ptolemy's 'abstract mathematics', it was commonly assumed that his conceptions could be traced back to an essentially Aristotelian cosmology. As a matter of fact, Aristotle and Ptolemy were in agreement with regard to the sphericity of the Earth and its position at the center of the universe, as well as the sphericity and the circular motion of the heavens. Hence, the physical considerations of the philosopher and the mathematical arguments of the Alexandrine astronomer could reinforce each other concerning these central issues. What is more, the *Almagest* began with a mention of Aristotle's partition of speculative knowledge into the three disciplines (mathematics, physics and theology) and repeated some physical theories of Aristotle, as we shall see. In this consensual spirit, Sacrobosco, for one, assumed the essential concordance between Aristotle and Ptolemy and could therefore rely on both authorities in his (very) elementary introduction to spherical astronomy which, in spite of its intrinsic scientific limits, was one of the most successful textbooks ever. In Latin Europe, an 'Aristotelian-Ptolemaic cosmology' thus emerged, bringing together elements from both classical authorities. This unified geocentric worldview was assumed by most philosophers and theologians, for instance Robert Grosseteste.[7] In his narrative of the Copernican revolution, Kuhn therefore felt legitimized to talk about an Aristotelian-Ptolemaic 'paradigm' which Copernicus' *De revolutionibus* was to undermine. We will limit ourselves here to the issue of terrestrial centrality and, unlike Kuhn, we will focus on the premisses instead of the conclusions of *De caelo* and *Almagest* regardless of the historical fact that these sources presented close cosmological views on the Earth's position.

Before we confront the arguments for geocentricism in Ptolemy and Aristotle, we shall clarify the meaning that we attach to some particularly relevant termini. 'Cosmology' means for us a general theory of the world as a whole. It concerns the dimensions, the structure, the order and the nature of the universe.[8] We will call 'mathematical astronomy' a treatment of the heavenly phenomena based on geometry and arithmetic. This 'Greek' perspective persisted in the western astronomical tradition, as is also evidenced by Renaissance sources on

---

has largely prevailed at least from the Middle Ages onwards. For a discussion of *Metaphyisics* XII,8, see Beere 2003. See also Lloyd 2000.

[5]See F.J. Carmody's Introduction to al-Bitruji 1952, in particular Chapter Three "Al-Bitruji in Western Europe, 1217–1531" (pp. 44–38).

[6]Di Bono 1990, Di Bono 1995, and Granada and Tessicini 2005.

[7]Cf. Panti 2001.

[8]Cf. Lloyd 1991, 146: "cosmology in the strictest and fullest sense […]: by the strictest sense I mean a comprehensive view of the cosmos as an ordered whole."

astronomy. For instance, the Wittenberg mathematician Erasmus Reinhold, who played a crucial role in the dissemination of Copernicus's work through his 'Copernican' astronomical tables *Prutenicae tabulae coelestium motuum* (Tübingen, 1551), conceived of astronomy as the culmination of mathematical disciplines. According to his introduction to the second part of these "Prussian" tables, Reinhold relies on arithmetic and geometry for computation and modeling of heavenly motions. Astronomy, one reads, deals with the *ratio* and *numerus* of heavenly motions, whereby geometry and arithmetic are its two instruments or *organa*.[9]

Moreover, we shall not assume the term 'physics' in the modern sense, but rather in a restricted Aristotelian meaning of a qualitative doctrine of motion based on causal explanation. Within an Aristotelian horizon, it could be regarded as a synonym of 'natural philosophy'. In accordance with this terminology, 'physical astronomy' shall refer to a qualitative doctrine of the heavens providing causal explanations according to philosophical assumptions on motion as well as on the nature of the Earth and the heavens. Moreover, we will call a 'cosmological approach' that treatment of the world which begins with a rational investigation of the whole and makes the theory of motion, in particular the motion on Earth, dependent on this general conception. On the other hand, we will call a 'physical approach' that which begins with consideration of the observable phenomena on Earth relative to motion, gravitation and such and includes conclusions about the structure of the world as a whole. As we will argue, this distinction can conveniently encapsulate the different approaches of Aristotle and Ptolemy to the issue with which we are presently concerned: geocentricism.

## 5.2 Aristotle

Aristotle's considerations on the Earth are presented in the conclusive part of the second book of *De caelo* as ii has been handed down to us. These chapters (II, 13 and 14) appear quite self-sufficient and can be regarded as an autonomous treatise on the Earth.

It might be useful to remember that the extant works of the so-called *corpus Aristotelicum* are generally considered to be the notes of the lectures which the philosopher held at the Lyceum and were later edited by his followers. These writings often resulted from the collection of short treatises, therefore titles are often only labels attached to miscellaneous writings on closely related subjects. This is the case with *De caelo*. In spite of its title, this work does not exclusively deal with the heavens. Instead, it consists of several distinct parts: books I (or *A*) and II (or *B*) on the universe as a whole and its parts, book III (or Γ) on sublunary elements, and book IV (or Δ) on lightness and heaviness. Some scholars pointed out that Chapters 13 and 14 of the second book are apparently a juxtaposition which occurred when *De caelo* was compiled into a unified work. This can be seen by the summary at the beginning of book III, a survey on the precedent sections in which the monograph on the Earth is omitted: "We have treated earlier of the first heaven and its parts, and also of the stars which are visible in it, their composition and natural characteristics, and the fact that they are ungenerated and indestructible" (III, 1; 298 a 24–27). Alberto Jori pointed to the relative autonomy of the section on celestial bodies (II, 7–12) and that on the Earth (II, 12–13) in his introduction to *De caelo*. He explained the existence and the insertion of these two

[9] Reinhold 1551, "Logistice scrupulorum astronomicorum", f. 1v. For a discussion on this matter, see Omodeo 2014, 104–106.

monographs by the fact that they complete the treatment of the universe as a whole which is the subject of the first book and of the first part of the second.[10] Paul Moraux divided the first two books of *De caelo* into three parts: 1. Περὶ τῆς παντὸς φύσεως (on the whole nature, I and II, 1–6), 2. Περὶ τῶν καλουμένων ἄστρων (on the so-called celestial bodies, II, 6–12) and 3. Περὶ τῆς γῆς (on the Earth, II, 12–13). He claimed, however, that the treatise on the Earth is an essential part of Aristotle's books A and B, regarded as an autonomous unit.[11] For our purposes, it is only important to stress that this section has a certain self-sufficiency. In the following we are going to focus specifically on this and avoid considerations on its relation to *De caelo* as a whole.

**Aristotle's confrontation with the cosmologies of his predecessors**

In the 'monograph' on the Earth, as we might call *De caelo* II, 13–14, Aristotle considers the issue of the form and the location of the Earth. Chapter 13 is basically an overview of the theses of his predecessors, and Chapter 14 is a treatment of his own theses. However, Aristotle also presents original considerations in Chapter 13 while discussing and criticizing other authors' theories. In addition, he describes some traditional arguments for geocentrism, although he does not consider them to be cogent. We shall call these 'pseudo-arguments':

1. Pseudo-argument concerning the finiteness of the universe: Aristotle firstly observes that most of those who hold the universe to be finite place the Earth at its center, with the exception of the Pythagoreans.[12] The historical relevance of this passage lies in the discussion of the cosmology of the Pythagoreans and the theory of the motion of the Earth including reference to Plato's *Timaeus*. In the Early Modern Period, several followers of Copernicus would interpret Aristotle's treatment of the Pythagorean cosmology as evidence of the existence of ancient supporters of heliocentrism. For the present discussion, this passage is also interesting in terms of Aristotle's report that the Pythagoreans regarded the absence of stellar parallax as insufficient evidence of terrestrial centrality and immobility:[13] "Since the Earth's surface is not in any case the centre, they [the Pythagoreans] do not feel any difficulty in supposing that the phenomena are the same although we do not occupy the centre as they would be if the Earth were in the middle. For even in the current view [that is, geocentrism] there is nothing to show that we are distant from the centre by half the Earth's diameter." As we shall see, Ptolemy did not take sufficient account of these remarks.
2. Argument concerning the fall of bodies:[14] Aristotle argues for the centrality and the position of the Earth based on consideration of the fall of bodies (see Fig. 5.1, left). He assumes that a bigger body falls faster than a smaller one. If the Earth were removed

[10] See Jori 2009, 123.
[11] Moraux 1949, 159: "Wenn wir einige durch Ideenassoziationen eingeleitete Abschweifungen beiseitelassen, so können wir behaupten, daß dieser gut abgewogene Plan die strukturelle Einheit der Bücher A und B beweist. Allem Anschein nach wurden diese Bücher als ein selbständiges Ganzes konzipiert: Ein Zeichen dafür ist, daß die Abhandlung über die Erde (B 13–14) als der letzte Punkt angekündigt wird, der zu besprechen ist, um das vorgesehene Programm abzuschließen."
[12] *De caelo* II,13 293 a 19–21. In the following we shall quote from the English translation by Guthrie, Aristotle 1986.
[13] *De caelo* II,13 293 b 25–30.
[14] *De caelo* II,13 294 a 11 ff.

from its central position, he says, it would reach its point of origin very quickly, as a consequence of its huge dimensions. This argument is remarkable for two reasons. First, it seems to be based on a *petitio principii*. In fact, if the fall of heavy bodies towards the center of the Earth serves as an argument for its cosmological centrality, it is already assumed that the center of gravity and the cosmological center are one and the same. But this coincidence, i.e., the centrality of the Earth (as an element as well as an astronomical body), is precisely what has to be demonstrated. Second, it assumes that the bigger a body is, the faster it travels downwards, an assumption which is supported by empirical evidence only under certain circumstances such as, for instance, when the shape of a falling body and the friction of the medium significantly influence its fall. This argument (which was already questioned in antiquity by atomist theories of matter and motion) is interesting, however, for its historical meaning, since it was not until the Middle Ages and the Renaissance that the physical theory upon which it relied was abandoned. It was Renaissance scientist-engineers like Giovan Battista Benedetti and Galileo Galilei who succeeded in refuting this viewpoint. The Aristotelian passage proposing the argument concerning the fall of bodies is also relevant because it contains an epistemological claim concerning the logical process needed to demonstrate the centrality of the Earth:[15] "I mean that we must decide from the very beginning whether bodies have a natural motion or not, or whether, not having a natural motion, they have an enforced one. And since our decisions on these points have already been made, so far as our available means allowed, we must use them as data." Accordingly, considerations on motion, or rather on terrestrial motion, should precede considerations on the structure of the whole universe. Therefore, Aristotle does not admit discussion on why the Earth and its elements are stable, since this is a factual presupposition and not something to be demonstrated. We could say that in his treatment, terrestrial physics, in particular his theory of the natural places of the elements (and of natural and violent motion), is the presupposition of his conception of the cosmos.[16] Aristotle adds to his argument that, if the Earth moved from its place, a falling body would fall ad infinitum, since it would encounter no solid bottom to arrest its downward motion. This consideration, according to Aristotle, elicited discussions among thinkers about the foundation upon which the elements are placed:[17] "Consider too that if one removed the Earth from the path of one of its particles before it had fallen, it would travel downwards so long as there was nothing to oppose it. This question, then, has become, as one might expect, a subject of general inquiry." In Aristotle's eyes, however, such inquiry is not worth conducting. According to him, immobility had an epistemological (and ontological) priority over speculations that relied on cosmic order in general.

---

[15] *De caelo* II,13 294 b 32–295 a 2.

[16] Cf. Moraux 1961, 182, n. 10: "Il serait trop long de relever tous les cas où, dans l'étude de l'univers et du ciel, il est fait état des principes de la physique terrestre. Voici pourtant quelques exemples intéressants. Théorie des quatre éléments, des mouvements et des lieux naturels […]. Théorie de la pesanteur et lois mécaniques de la chute des corps […]. Théorie de la génération et de la corruption. Opposition du 'selon nature' et 'contre nature' ou 'par violence'. Hylémorphisme […]. Existence de détérminations telles que devant-derriére, droite-gauche, etc., chez les animaux […]."

[17] *De caelo* I,13 294 a 17–19.

3. Pseudo-argument concerning creation:[18] Aristotle remarks that those who hold that the cosmos had an origin also believe that the Earth agglomerated at its center. Aristotle not only disagrees on the assumption of a "creation" or "origin" of the world (an issue on which he does not expand here), but also rejects the argument. If one assumed with Empedocles that the various parts of the Earth were brought together by a vortex, one would ignore the fact that up and down have an ontological and epistemological priority over motion. In other words, space determinations should precede spatial displacements:[19] "Nor, again, are heavy and light defined by the vortex: rather, heavy and light things existed first, and then the motion caused them to go either to the centre or the surface. Light and heavy, then, were there before the vortex arose [...]. In an infinite space there can be no up and down, yet it is these that distinguish heavy and light." Hence, spatial determinations (up and down) come first, then the determinations of lightness and heaviness and, eventually, motion. In general terms, one can remark that the argument based on creation is not valid for Aristotle because the centrality and immobility of the Earth do not need to be demonstrated from a cosmological perspective but are already given as perceptible evidence.
4. Argument concerning lightness and heaviness:[20] The priority of the theory of natural places over cosmological considerations is also reassessed by Aristotle relative to the position of the Earth at the center. According to Anaximander and others, the reason for that is "indifference." The Earth is equidistant from all extremes, therefore it maintains its central position and is at rest. In Aristotle's eyes, this argument is ingenious but not true. In fact, he remarks that not only the centrality of the Earth and its natural tendency toward the center should be taken into account, but also the upward tendency of fire. The entire theory of elementary motions should be considered, since only the Earth falls towards the center and not the other elements:[21] "The reason is not impartial relation to the extremes, but motion towards the centre is peculiar to the Earth." As a conclusion, Aristotle repeats that only the theory of motion, in particular the consideration of the 'light-heavy' and 'up-down' determinations, contains decisive and valid arguments relating to geocentricism (see Fig. 5.1.).

**Aristotle's presentation of his own views**

Chapter II, 14 deals essentially with Aristotle's own views. It begins with considerations concerning terrestrial immobility:[22] "For ourselves, let us first state whether it [the Earth] is in motion or at rest." In fact, some thinkers believed that the Earth was a celestial body among others and other philosophers held that it was at the center but rotated about its own axis. As Aristotle already remarked, the former theory belonged to the Pythagoreans and the latter to Plato. In *De caelo* he questions the views of these predecessors but, as we shall see, he treats the problem beginning with his theory of motion rather than from a general cosmological perspective.

[18] *De caelo* II,13 295 a 13 ff.
[19] *De caelo* II,13 295 b 3–9.
[20] *De caelo* II,13 295 b 10 ff.
[21] *De caelo* II, 13 295 b 23–25.
[22] *De caelo* II, 14 296 a 24–25.

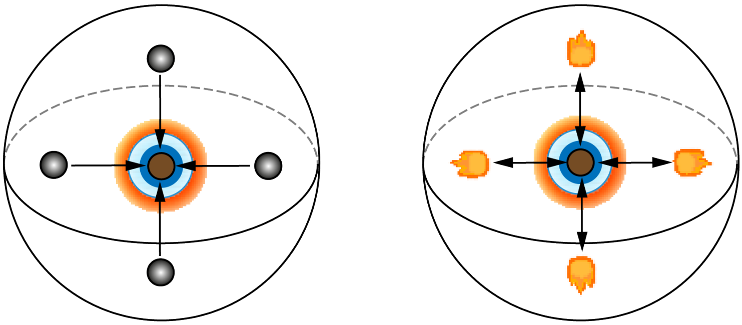

Figure 5.1: Left: the center of the sublunary world as 'gravitational' center according to the argument concerning the fall of bodies. Right: sublunary world as symmetry center according to the argument concerning lightness and heaviness.

1. Argument concerning the categorization of motion:[23] Aristotle objects to the geokinetic theories of the Pythagoreans and Plato that these are not compatible with the theory of motion, or rather with *his* theory of motion. A metaphysical premiss is also at stake: the order of the world is eternal. The reasons for this assumption should be sought elsewhere. Aristotle assumes also that a 'natural motion' is such that a whole and its parts share the same tendency. As for earth as an element, its tendency is "toward the center", as everyday experience testifies. Hence, the hypothetical motion of the Earth, like other planets, would be a 'violent' or 'enforced' motion but, since a violent motion cannot be eternal, the geokinetic theory would violate the eternal regularity of nature.
2. Argument concerning the rise and setting of stars:[24] Aristotle remarks that the terrestrial motion would affect celestial appearances, in particular the fixed stars. This argument is in striking conflict with Aristotle's previous observation that the Pythagoreans did not accept the argument concerning stellar parallax as a proper objection against their planetary conception of the Earth, since its validity depends on the dimensions of the cosmos. Aristotle's argument seems to be rather confused:[25] "Secondly, all the bodies which move with the circular movement are observed to lag behind and to move with more than one motion, with the exception of the primary sphere: the Earth therefore must have a similar double motion, whether it moves around the centre or is situated at it. But if this were so, there would have to be passing and turnings of the fixed stars. Yet these are not observed to take place: the same stars always rise and set at the same places on the Earth." It seems plausible that the double motion of planets to which Aristotle refers here concerns the daily and the periodical rotations, one along the equator and one along the ecliptic. It is, however, unclear why the ro-

[23] *De caelo* II,14 296 a 25 ff.
[24] *De caelo* II,14 296 a 34 ff.
[25] *De caelo* II, 14 296 a 35 – 296 b 6.

tation of the Earth at the center of the world should have more than one motion, if not for a priori reasons forcing the analogy between the Earth and the other planets. It is curious that Copernicus's pupil Rheticus would turn this argument against Aristotle and would argue in his *Narratio prima* that Copernicus's idea of threefold terrestrial motion (daily, annual and "of declination") conformed with Aristotle's remark that a planet must have more than one motion:[26]

> Following Plato and the Pythagoreans, the greatest mathematicians of that divine age, my teacher thought that in order to determine the causes of the phenomena circular motions must be ascribed to the spherical Earth. He saw (as Aristotle also points out) that when one motion is assigned to the Earth, it may properly have other motions, by analogy with the planets. He therefore decided to begin with the assumption that the Earth has three motions, by far the most important of all.

To sum up, the general meaning of Aristotle's argument from the rising and setting of the stars is clear, but not its details. It should be additionally noted that this argument is not based on terrestrial physics, as usual, but rather on astronomical considerations.

3. Argument from the identity of gravitational and cosmological center:[27] Aristotle remarks that the cosmological and gravitational center of the terrestrial element coincide:[28] "[…] that the Earth and the Universe have the same centre […] we see that weights moving toward the Earth do not move in parallel lines but always at the same angles to it […]." This argument obviously presupposes the sphericity of the Earth. This reasoning is therefore not based on commonsense and intuitive observations, as Aristotle presents it, but lies in theoretical assumptions (arguments for the spherical form of the Earth can be found elsewhere, for instance in *De caelo* II, 14 298 a 7–10). An observer who already knows that the Earth is spherical and notices that heavy bodies fall vertically to the ground at all latitudes will conclude that heavy bodies fall straight downwards to the center of the Earth. Still, Aristotle remarks that they fall to the center of the Earth only incidentally. He argued that their tendency is, in fact, towards the cosmological center. What counts is *place*. Earth goes to the center like fire to the periphery of the central region of the universe. Accordingly, Aristotle argued, coincidence of terrestrial and cosmological center is accidental. In other words, symmetry has an ontological and epistemological priority over gravitation. Be that as it may, the conclusion is that the Earth "must be at the center and immobile" (see Fig. 5.2).
4. Argument concerning objects thrown upwards: Aristotle adds a remark concerning objects thrown upwards. They will always come back to the ground in a straight line:[29] "To our previous reasons we may add that heavy objects, if thrown forcibly upwards in a straight line, come back to their starting-place, even if the force hurls them to an unlimited distance."
5. Argument concerning the simplicity of motion:[30] This is a reworking of considerations on natural places. A simple body, as an element, can have only one motion and

[26]Rheticus 1959, 147–148.
[27]*De caelo* II,14 296 b 6 ff.
[28]*De caelo* II, 14 296 b 15–16.
[29]*De caelo* II, 14 296 b 22–25.
[30]*De caelo* II,14 b 25 ff.

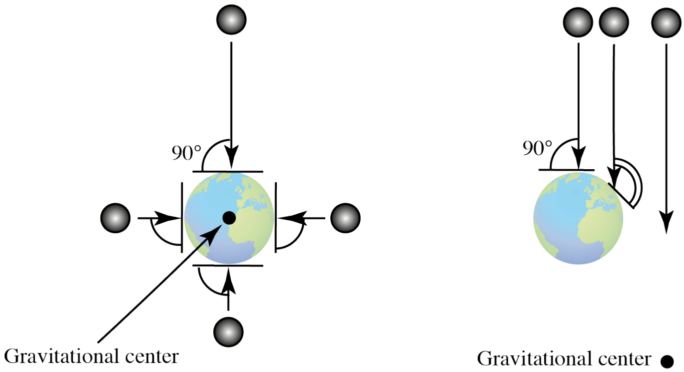


Figure 5.2: Right: Earth as a gravitational center. Falling bodies hit the Earth's surface at the same angle (90°). Left: gravitational center is outside the Earth. For a gravitational center lying at a very remote distance, the falling bodies should hit the Earth's surface in parallel lines.

cannot simultaneously move towards and away from the center, as would be the case if the Earth moves. In that case, in fact, the body's motion would have a vertical as well as a horizontal component. Additionally, the whole must be in the place which its parts tend to reach. Since no force can compel the Earth as a whole to abandon its natural place, it must be at rest at the center.

6. Confirmation from mathematical astronomy: Mathematical astronomy receives very little acknowledgment from Aristotle. Its role is merely to confirm his views based on mainly physical arguments. As he writes in the conclusion of his defense of the centrality and immobility of the Earth:[31] "This belief finds further support in the assertions of mathematicians about astronomy: that is, the observed phenomena – the shifting of the figure by which the arrangement of the stars is defined – are consistent with the hypothesis that the Earth lies at the centre. This may conclude our account of the situation and the rest or motion of the Earth." This argument would later be developed by Ptolemy for the specific case of possible displacement of the Earth to the east or to the west. What Aristotle means here is that the angular distances between stars within certain 'arrangements' such as constellations remain constant (see Fig. 5.3).

## 5.3 Ptolemy

The first book of the *Almagest* starts by mentioning Aristotle's division of theoretical philosophy into three primary categories, theology, physics and mathematics. In the following discussion, Ptolemy makes his point clear:[32]

> […] the first two divisions of theoretical philosophy should rather be called guesswork than knowledge, theology because of its completely invisible and

[31] *De caelo* II, 14 297 a 2–8.

[32] *Almagest* I H6, p. 46. Here and in the following we will quote from the English translation by Toomer, Ptolemy 1984.

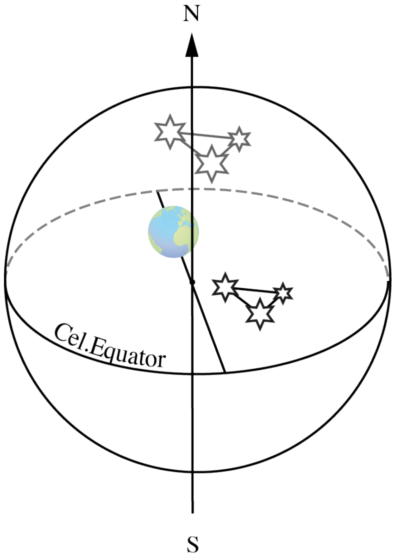


Figure 5.3: Argument concerning mathematical astronomy. The angular distances between the stars in the same constellation remain constant.

> ungraspable nature, physics because of the unstable and unclear nature of matter; hence there is no hope that philosophers will ever be agreed about them; and that only mathematics can provide sure and unshakeable knowledge to its devotees, provided one approaches it rigorously.

Ptolemy organizes his discussion of mathematical constructs modeling cosmic order along these lines. His basic principles – geocentrism, sphericity of the Earth and of the sky – are supposed to be verified by means of mathematical astronomy. As a professional astronomer he tries to "provide proofs in all of these topics by using as starting-points and foundations, as it were, for our search the obvious phenomena, and those observations made by the ancients and in our own times which are reliable."[33]

Ptolemy's thorough discussion is organized according to the following scheme (*Almagest* I 3–8):

1. that the heavens move like a sphere;
2. that the Earth, taken as a whole, is also sensibly spherical;
3. that the Earth is in the middle of the heavens;
4. that the Earth has the ratio of a point to the heavens;
5. that the Earth does not have any motion from place to place;
6. that there are two different primary motions in the heavens.

In the following we will discuss the argumentation used by Ptolemy in relation to the first five points. The last point distinguishes between the daily rotation of the celestial sphere

[33] *Almagest* I H9, p. 48.

"which carries everything from east to west" (first primary motion) and the motion of Sun, Moon and planets in the opposite direction relative to the axis, which, in turn, is inclined relative to the rotational motion of the first motion (second primary motion). The trajectory of the Sun due to this motion (relative to the sphere of the fixed stars) defines the ecliptic plane inclined relative to the equator of the celestial sphere. Ptolemy added to this list a third 'celestial motion', that is, the precession first found by Hipparchus and confirmed by Ptolemy himself. This kind of motion was not yet known in Aristotle's time.

**The heavens move like a sphere**

Let us emphasize that the statement that "the heavens move like a sphere" was considered by Ptolemy to be logically equivalent to the statement that "the stars' trajectories are circular in shape" and vice versa, only because for him the stars were thought to be fixed on the celestial sphere.[34] The arguments proposed in the *Almagest* I,3 for the sphericity of the heavens can be roughly classified as observational, 'physical' and 'mathematical'. Ptolemy suggests that 'the ancients' initially arrived at the concept of the celestial sphere from the following kind of observations:[35]

> They saw that the Sun, Moon and other stars were carried from east to west along circles which were always parallel to each other, that they began to rise up from below the Earth itself, as it were, gradually got up high, then kept on going round in similar fashion and getting lower, until, falling to Earth, so to speak, they vanished completely, then after remaining invisible for some time, again rose afresh and set; and [they saw] that the periods of these [motions], and also the places of rising and settings, were, on the whole, fixed and the same.

Ptolemy further qualifies the observational evidence for the revolution of always-visible stars and the motion of partly invisible stars. The observational arguments concerning the former, that

- their motion is circular and always takes place about one and the same center;
- that point becomes the pole of the heavenly sphere for observers;
- and those stars which are closer to the pole revolve on smaller circles;

and concerning the latter, that:

- those stars that are near the always-visible stars remain invisible for a short time;
- and those further away remain invisible for a long time in proportion to their distance,

are visualized in Fig. 5.4.

Obviously, these arguments are of 'local' geographical character: they can be put forward after just two nights of observations, without comparison to observational data from different places.[36] Stars can be observable at some localities and invisible at other places; they can belong to the category of always-visible stars at a certain geographical latitude and

[34] In the second book of his *Planetary hypotheses*, where Ptolemy extends the mathematical models of the *Almagest* to the physical realm, stars are thought to be fixed not on the spherical shell, but rather between nested spherical shells.

[35] *Almagest* I H10, p. 48.

[36] Although intuitively clear, these arguments really need some mathematical justification, namely that the intersection between a plane and a sphere is always a circle.

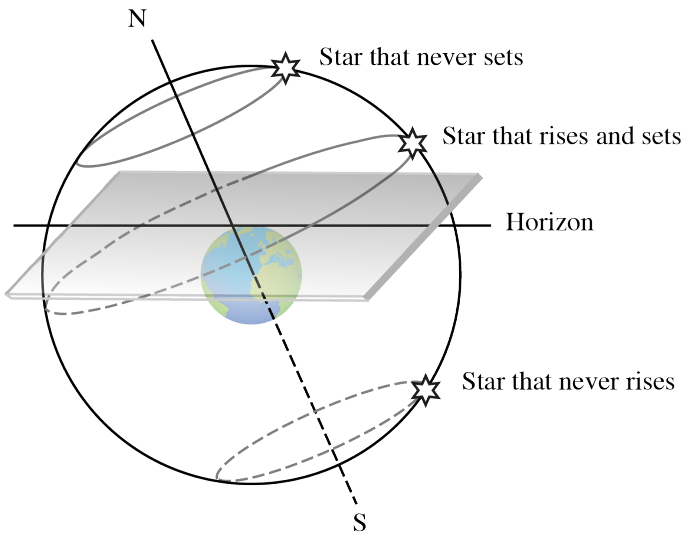


Figure 5.4: Always-visible stars and stars that rise and set at a given geographical position. Here and in the following we will depict the horizon plane drawn at an observer's position by a grey shadowed surface.

to that of stars that rise and set at other places. The position of the great circle in the sky which separates these two classes of stars is different at different latitudes. Aristotle made use of this local concept when he defined the wind directions in his *Meteorologica*.[37]

Discussing the consequences of these observational facts on astronomical knowledge, Ptolemy stresses that "absolutely all phenomena are in contradiction to the alternative notions which have been propounded."[38] It is interesting to note how deeply the paradigm of the sphericity of the cosmos has indeed prejudiced his mathematical speculations: in fact, he overlooked another mathematically equivalent explanation – in a cylindrical world (see Fig. 5.5) the observational effects would hardly be distinguished from those observed in a spherical cosmos.[39]

The other possible mathematical solution overlooked by Ptolemy is a rotational three-axis ellipsoid. For the special sort of ellipsoid with two equal axes rotating about the remaining axis, the observational effects will be the same as in the spherical universe (see Fig. 5.6).

As alternative hypotheses accounting for the visible paths of the stars, Ptolemy mentions only the untenable opinion (perhaps held by Xenophanes)[40] that stellar motions might occur in a straight line towards infinity. It is clear that such motion can be ascribed only to

[37] *Meteorologica*, II, VI, 363 b.
[38] *Almagest* I H11, p. 48.
[39] The authors would like to thank H. Mendell for a thorough discussion on the cylindrical model in relation to Anaximander.
[40] Aetius II 24.9: "The same philosopher [Xenophanes] maintains that the Sun goes forward ad infinitum, and that it only appears to revolve in a circle owing to its distance."

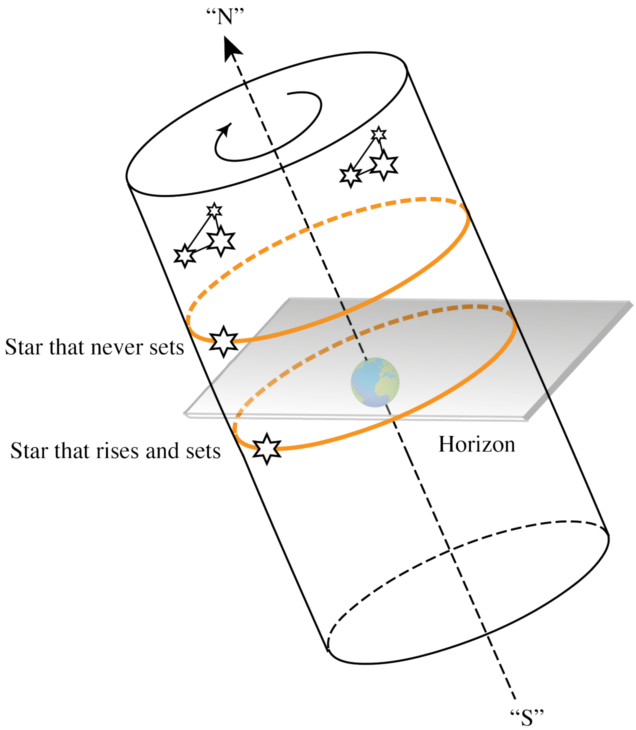


Figure 5.5: Observational effects in the 'cylindrical universe'. The stars' visible trajectories are concentric circles; the local horizon defines the different sets of always-visible stars and stars that rise and set. The mutual distances between stars remain constant.

stars that rise and set, and not to those which are ever-visible and move in circular paths. In fact, the above-mentioned arguments are sufficient to rule out this hypothesis. Nevertheless, Ptolemy proposes some other objections:[41]

- "[…] What device could one conceive which would cause each of them [stars] to appear to begin their motion from the same starting-point every day?"
- "How could the stars turn back if their motion is towards infinity?"
- "[…] If they did turn back, how could this not be obvious?"
- In this case "[…] they must gradually diminish in size until they disappear, whereas, on the contrary, they are seen to be greater at the very moment of their disappearance […]"

The first three counter-arguments have a touch of 'commonsense' reasoning or a purely rhetorical character. The last argument is totally fabricated: Ptolemy himself refers to this phenomenon a couple of lines later as being caused "by the exhalations of moisture surround-

[41] *Almagest* I H11, pp. 48–39.

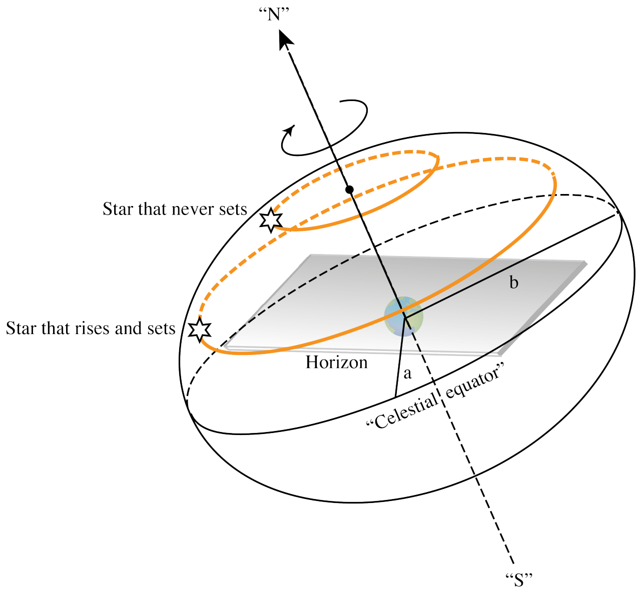


Figure 5.6: Observational effects in the 'ellipsoidal universe'. Fixed stars lie on the surface of an ellipsoid with two equal axes ($a = b$) rotating about an axis perpendicular to the plane defined by these axes. The stars' visible trajectories are concentric circles; the local horizon defines the different sets of always-visible stars and stars that rise and set. The mutual distances between stars remain constant.

ing the Earth being interposed between the place from which we observe and the heavenly bodies."[42]

Additionally, Ptolemy refers to another hypothesis which he regards as "completely absurd," namely, that "the stars are kindled as they rise out of the Earth and are extinguished again as they fall to Earth."[43] Nevertheless, he discusses this issue thoroughly.[44] Not only the necessity of cosmic order should rule this hypothesis out – because otherwise "the strict order in their size and number, their intervals, positions and periods could be restored by such a random and chance process" (in fact, the process need not necessarily be "random") – but also some other objections of special interest are proposed. Ptolemy mentions that, if this were the case, then

- "[…] One whole area of the Earth has a kindling nature, and another an extinguishing one, or rather that the same part [of the Earth] kindles for one set of observers and extinguishes for another set; and that the same stars are already kindled or extinguished for some observers, while they are not yet for others […]"

[42]This explanation is actually incorrect; in his later work (*Optics*, III 60) this phenomenon, now known as a Ponzo-illusion, is correctly explained as a pure psychological effect.
[43]Aetius II 13, 14, III 2.II: "According to Xenophanes the stars are made of clouds set on fire; they are extinguished each day and are kindled at night like coals, and these happenings constitute their settings and rising respectively."
[44]*Almagest* I 3 H12, p. 49.

- For the stars which are ever-visible in certain regions and are partly-visible at others, one should admit “that stars which are kindled and extinguished for some observers never undergo this process for other observers.”

These counter-arguments are really of ‘global’ geographical character: they can be put forward only through comparison of observational information gained at different geographical localities.

Ptolemy also presents some arguments from mathematical astronomy for the sphericity of the cosmos:[45]

- “[…] If one assumes any motion whatever, except spherical, for the heavenly bodies, it necessarily follows that their distances, measured from the Earth upwards, must vary, wherever and however one supposes the Earth itself to be situated. Hence the sizes and mutual distances of the stars must appear to vary for the same observers during the course of each revolution, since at one time they must be at a greater distance, at another at a lesser. Yet we see that no such variations occur.” (compare Fig. 5.3.)
- “[…] Since of different shapes having an equal boundary those with more angles are greater [in area or volume], the circle is greater than [all other] surfaces, and the sphere greater than [all other] solids,[46] [likewise] the heavens are greater than all other bodies.”
- “No other hypothesis can explain how sundial constructions produce correct results […]”

In fact, the first argument refers to the constancy of the stars’ mutual distances and spatial relations. Once again, Ptolemy does not mention here that the mutual distances between stars would remain intact not only in a ‘cylindrical’ world but also in a cosmos in the form of an ellipsoid (see above).

The second counter-argument is of a curiously mixed nature: a correct mathematical result intermingled with a still-naive interpretation of an extremal principle – a future tradition which survived until Leibniz.

How basic the concept of celestial sphere was for sundial constructions is widely discussed in the literature:[47] astronomical calculations with gnomons make sense only in the geocentric world and the apex of a gnomon symbolizes the Earth in the center of the spherical universe. The very visualization of the concept of the celestial sphere with gnomons and its usage in sundials can be traced back to the *analemma* construction as discussed in Vitruvius (see Fig. 5.7).

For completeness and to show the actual path of Ptolemy’s argumentation, we will list the arguments which he himself classifies as ‘physical’:

[45] *Almagest* I H13, pp. 49–40.
[46] According to Toomer (Ptolemy 1984, 41), these propositions were proved in a work by Zenodorus as early as the second century BCE.
[47] See, for example, D.R. Dicks *Early Greek Astronomy to Aristotle*, p. 166: “The data are very inaccurate for the latitude of Babylon (particularly the equinoctial and winter solstitial figures), which is not surprising since the underlying assumption seems to be that the length of the shadow increases in arithmetical progression with the height of the Sun […]. Moreover, the results are set out according to a predetermined scheme whereby the solstices and equinoxes are placed arbitrarily on the 15th day of the first, fourth, seventh, and tenth months of a schematic year of twelve months and thirty days each.”

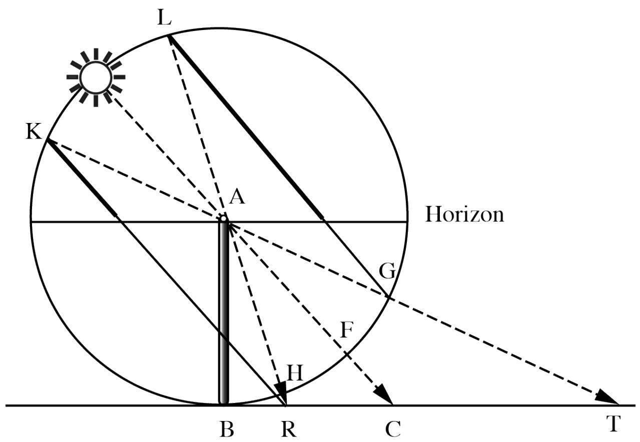


Figure 5.7: The gnomon AB is placed perpendicularly to the horizon plane. Point $R$ marks the end of the shadow at summer solstice; point $T$ marks the winter solstice. Bisecting the arc $HG$ and marking this point with $F$ one gets the point of equinox $C$ at the prolongation of the line $AF$. Obliquity of the ecliptic is depicted by the angle $RAC$.

- "[…] The motion of the heavenly bodies is the most unhampered and free of all motions; and freest motion belongs among plane figures to the circle and among solid shapes to the sphere […]"
- "[…] The aether is, of all bodies, the one with constituent parts which are finest and most like each other; now bodies with parts like each other have surfaces with parts like each other; but the only surfaces with parts like each other are the circular, among planes, and the spherical, among three-dimensional surfaces. And since the aether is not plane, but three-dimensional, it follows that it is spherical in shape."
- "[…] Nature formed all earthly and corruptible bodies out of shapes which are round but of unlike parts, but all aetherical and divine bodies out of shapes which are of like parts and spherical. For if they were flat or shaped like a discus they would not always display a circular shape to all those observing them simultaneously from different places on Earth. For this reason it is plausible that the aether surrounding them, too, being of the same nature, is spherical, and because of the likeness of its parts moves in a circular and uniform fashion."

### The Earth, taken as a whole, is sensibly spherical

The arguments aimed at demonstrating the sphericity of the Earth were widely known in Antiquity and are repeated by Ptolemy; the specification "taken as a whole" should indicate that one ignores the local irregularities of the Earth's surface. For the sake of completeness, Ptolemy also considers some other possible forms for the Earth (concave, plane, of polygonal shape, cylindrical) and shows which astronomical evidence would rule out these cases.

### 5.3.1 The Earth is in the middle of the heavens

Ptolemy treats geocentrism and enlists a series of astronomical arguments in favor of this thesis in *Alamagest* I,5. Ptolemy tries to consider all other possible cosmological arrangements with an eccentric Earth and rules them out on the basis of pure observations. The alternatives are the following:

1. that the Earth is not on the axis [of the universe] but equidistant from both poles,
2. it is on the axis but removed towards one of the poles,
3. it is neither on the axis nor equidistant from both poles.

Let us consider the first case. Two possible positions for the Earth are given in Fig. 5.8.

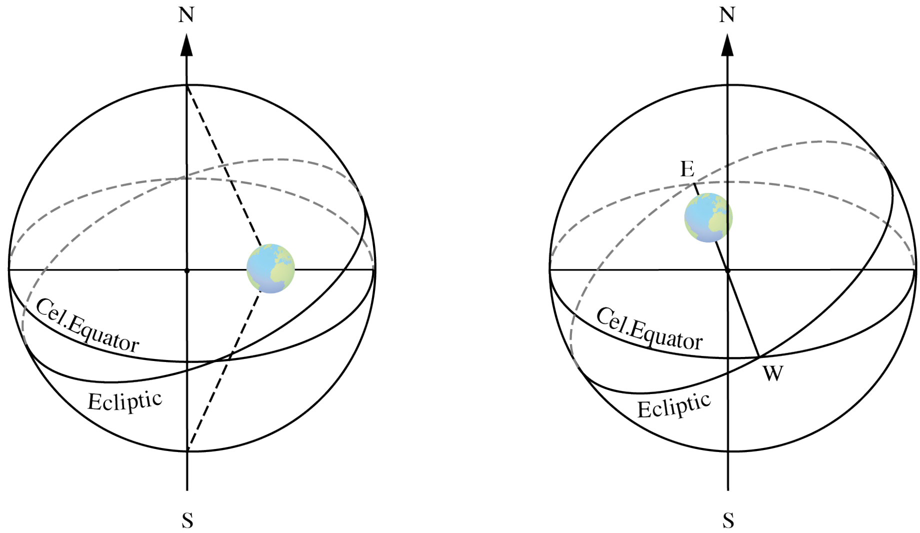


Figure 5.8: First case: the Earth is equidistant from both poles – two possible locations.

In order to understand Ptolemy's arguments, it is useful to recall that only if the Earth is in the center of the celestial sphere will the Sun rise for any observer exactly at the east point and set at the west point only twice a year, namely at equinoctials.[48] The equinox is defined as a day when the Sun's declination $\delta = 0$, that is, the Sun's trajectory lies on the celestial equator, and the length of the day is equal to the length of the night (see Fig. 5.9).

Ptolemy argues:[49]

> If the image [the Earth] removed towards the zenith or the nadir of some observer,[50] then, if he were at *sphaera recta*, he would never experience equinox,

[48]Strictly speaking, the eastern and western directions are defined locally for every observer relative to the local northern direction; for further considerations we will also use a global coordinate system with a northern direction defined through the rotational axis of the cosmos and an east-west direction coinciding with the intersection line between the ecliptic and equatorial plane.
[49]*Almagest* I H17, p. 41.
[50]Here, Ptolemy explicitly implies that the Earth's size is negligible in comparison to the distance to the stars; otherwise the Earth would not be equidistant from both poles.

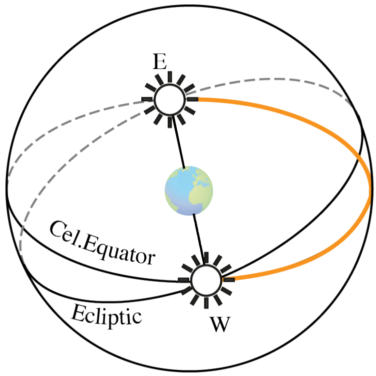


Figure 5.9: Equinox: the Sun's declination $\delta = 0$ and the visible path of the Sun coincides with the celestial equator. The Sun rises directly in the east and sets directly in the westerly direction for every observer on the Earth's surface. Here and in the following, we will depict the visible path of the Sun above the horizon plane with a thick line.

> since the horizon would always divide the heavens into two unequal parts, one above and one below the Earth; if he were at *sphaera obliqua*, either, again, equinox would never occur at all, or [if it did occur], it would not be at a position halfway between summer and winter solstices, since these intervals would necessarily be unequal, because the equator, which is the greatest of all parallel circles drawn about the poles of the [daily] motion, would no longer be bisected by the horizon; instead [the horizon would bisect] one of the circles parallel to the equator, either to the north or to the south of it. Yet absolutely everyone agrees that these intervals are equal everywhere on Earth, since [everywhere] the increment of the longest day over the equinoctial day at the summer solstice is equal to the decrement of the shortest day from the equinoctial day at the winter solstice.

Ptolemy considers separately two possible positions of observation, one at the equator (*sphaera recta*) and another at an arbitrary latitude (*sphaera obliqua*). He concludes, in fact, that in both cases one would never experience an equinox, since the horizon would always divide the heavens into two unequal parts. The argumentation is visualized in Fig. 5.10.

The completeness of Ptolemy's analysis of the astronomical consequences of this case can be seen from his remark that it can nevertheless happen that one observes the same lengths of day and night at *sphaera obliqua*. But in this case that will occur not at the true equinoctial date when the solar declination $\delta = 0$ but at some other date (see Fig. 5.11)!

The next step in Ptolemy's analysis is to consider the observational consequences of the Earth's displacement to the east or to the west. He proposes the following counter-arguments:[51]

[51] *Almagest* I H18, p. 42.

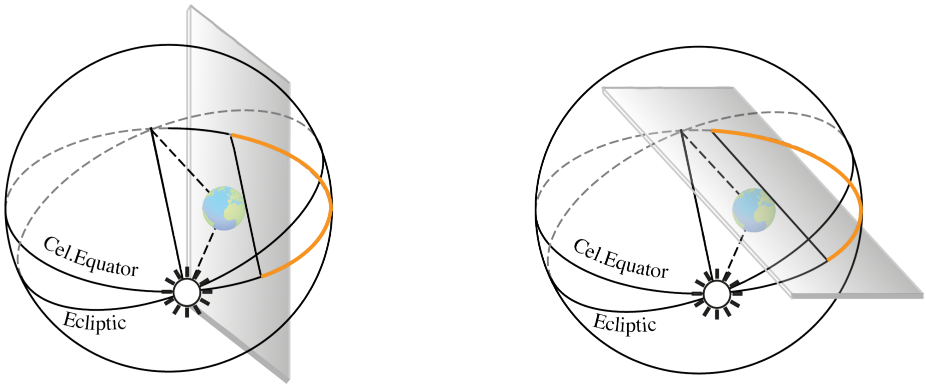


Figure 5.10: The Sun's declination $\delta = 0$. Observational situation at *sphaera recta* (left) and at *sphaera obliqua* (right).

- The sizes and distances of the stars in this case would not remain constant and unchanged at the eastern and western horizons (see Figs. 5.13 and 5.12).
- The time-interval from rising to culmination would not be equal to the interval from culmination to setting (see Fig. 5.12).

Having considered and ruled out the possible symmetrical displacement of the Earth from the rotational axis of the universe, Ptolemy begins to consider the astronomical consequences of the possible displacement of the Earth in the north-south direction *along the rotational axis*. He concludes that in this case:[52]

- The plane of the horizon would divide the heavens into unequal parts, different for different latitudes.
- The plane of the ecliptic would also be divided by the plane of the horizon into unequal parts; instead the six zodiacal signs are visible above the Earth at all times and places, while the remaining six are invisible.
- Only at *sphaera recta* could the horizon bisect the celestial sphere.
- The shadow of the gnomon at equinoxes at sunrise would no longer form a straight line with its shadow at sunset in a plane parallel to the horizon, not even sensibly.

The first and third arguments can be easily understood with the help of Fig. 5.14 and 5.15.

The last (fourth) argument in the list is of special interest:

> […] if the Earth were not situated exactly below the [celestial] equator, but were removed towards the north or south in the direction of one of the poles, the result would be that at the equinoxes the shadow of the gnomon at sunrise would no longer form a straight line with its shadow at sunset in a plane parallel to the horizon, not even sensibly.

Actually, this is old 'evidence' for geocentricism, which was also used as a proof by Pliny in his *Natural History*:[53]

[52] *Almagest* I H18–19, p. 42.
[53] Pliny, *Natural History* I, Chapter 70.

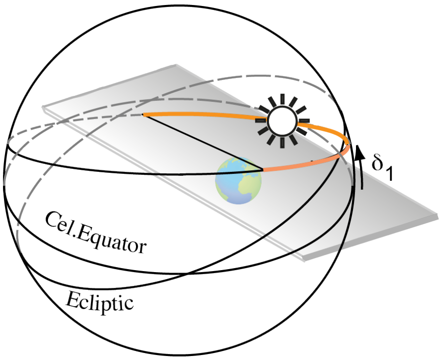


Figure 5.11: 'False equinox': the Earth is not on the rotational axis of the universe but equidistant from both poles. One can possibly observe the same length of day and night not at the true equinoctial date $\delta = 0$ but at some other date with some other Sun's declination $\delta_1$.

> It is demonstrated by dioptra, which affords the most decisive confirmation of the fact, that unless the Earth was in the middle, the days and nights could not be equal; for, at the time of the equinox, the rising and setting of the Sun, are seen on the same line, and the rising of the Sun, at the summer solstice, is on the same line with its setting at the winter solstice; but this could not happen if the Earth was not situated in the centre.

A visualization of the above-mentioned argument for the case of the equinox in Pliny is given in Fig. 5.16.

A similar line of argumentation can be found in Euclid:[54]

> Let Cancer, at point Γ in the east, be observed through a dioptra placed at point Δ, and then through the same dioptra Capricorn will be observed in the west at point $A$. Since points $A\Delta\Gamma$ are all observed through the dioptra, the line $A\Delta\Gamma$ is straight.

It should be noted that Cancer and Capricorn as zodiacal signs are not *observable* as points on the celestial sphere; on the other hand, the position of the Sun at the summer solstice is marked by its entrance into the tropic of Cancer and the longitudinal difference between the two signs is equal to 180 degrees. That means that Pliny's argument can simply be a reformulation of the 'mental observation'[55] illustrated by Euclid. It is remarkable that Ptolemy uses this statement only as a counter-argument. Ptolemy completed his argumentation for the third case (the Earth is neither on the rotational axis nor equidistant from both poles) by concluding that it is impossible because "the sorts of objection which we made to the first [two] will both arise in that case."[56]

[54]Euclid, *Phaenomena* I.

[55]To our knowledge, this kind of observation was, in fact, never made: not only the atmospheric refraction but also a final size of the Sun would make the precision of such 'proofs' mathematically invalid.

[56]*Almagest* I H19, p. 42.

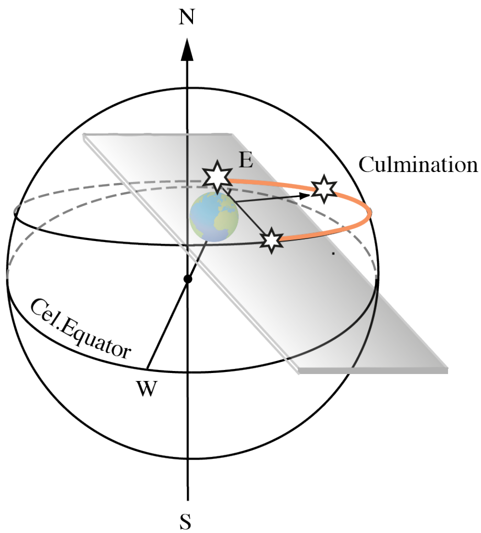


Figure 5.12: Displacement in the eastern direction. Stars appear to be bigger in the eastern direction and smaller in the western direction. The peak moment does not lie in the middle of the time-interval between the rising and setting of stars.

The final reason for the central position of the Earth comes from the observation of the Moon's eclipses:[57]

> Furthermore, eclipses of the Moon would not be restricted to situations where the Moon is diametrically opposite the Sun (whatever part of the heaven [the luminaries are in]),[58] since the Earth would often come between them when they are not diametrically opposite, but at intervals of less than a semi-circle.

Ptolemy does not discuss this argument in detail: in fact, it presupposes that both the Sun and the Moon rotate in circular motion around the center of the cosmos. This is certainly not the case for the more elaborate lunar and solar theories developed in the *Almagest*.

### The Earth has the ratio of a point to the heavens

One should emphasize that Ptolemy's statements considered above are practically all valid only if one neglects the Earth's size in comparison to the size of the universe. His continual repetition of the word *sensibly* clearly indicates that he himself was aware of the intrinsic precision of his 'proofs'. The arguments presented in this section should in fact give the necessary justification of the approximation used in the 'proofs' of the previous sections. The following arguments are proposed:[59]

---

[57] *Almagest* I H19, p. 42.
[58] That is, at opposition, at full Moon.
[59] *Almagest* I H21, p. 43.

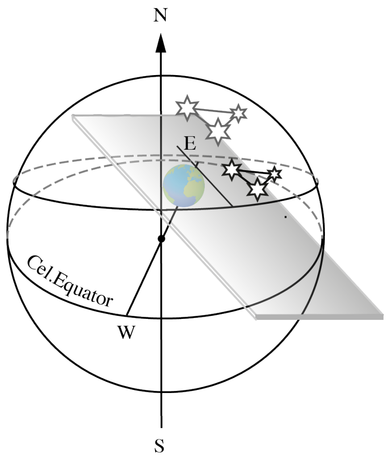


Figure 5.13: Displacement in the eastern direction. The angular distances between the stars in the same constellation appear to be bigger in the eastern direction and smaller in the western direction.

- “the sizes and the distances of the stars, at any given time, appear equal and the same from all parts of the Earth everywhere, as observations of the same [celestial] objects from different latitudes are found to have not the least discrepancy from each other”;
- “the gnomons set up in any part of the Earth […] and likewise the centers of armillary spheres, operate like the real center of the Earth; that is, the lines of sight [to heavenly bodies] and the paths of shadows caused by them agree as closely with the [mathematical] hypotheses explaining the phenomena as if they actually passed through the real center-point of the Earth”;
- “the planes drawn through the observer’s lines of sight at any point, which we call ‘horizons’, always bisect the whole heavenly sphere.”

The very nature of astronomical observations, however, limits the precision of these arguments to a perceptible level – a fact which was not lost on Ptolemy. Once more, he has to repeat that “the Earth has, to the senses, the ratio of a point to the distance of the sphere of the so-called fixed stars.”[60] What is now missing are the arguments which could rule out the displacements relative to the center of the universe which were of the Earth’s size. Such displacements would not be observable with the precision of naked-eye astronomy but could be monitored in frames of Aristotle’s physics through terrestrial observation.

### The Earth does not have any motion from place to place

As we have seen, Ptolemy thinks that geocentrism can be sufficiently demonstrated through astronomical considerations based on geometry and observation up to a *perceptible level*.

[60] We can only agree here with Toomer’s comment that the classification ‘so-called’ used for the fixed stars means that for Ptolemy the stars did in fact have a motion – that is, precession.

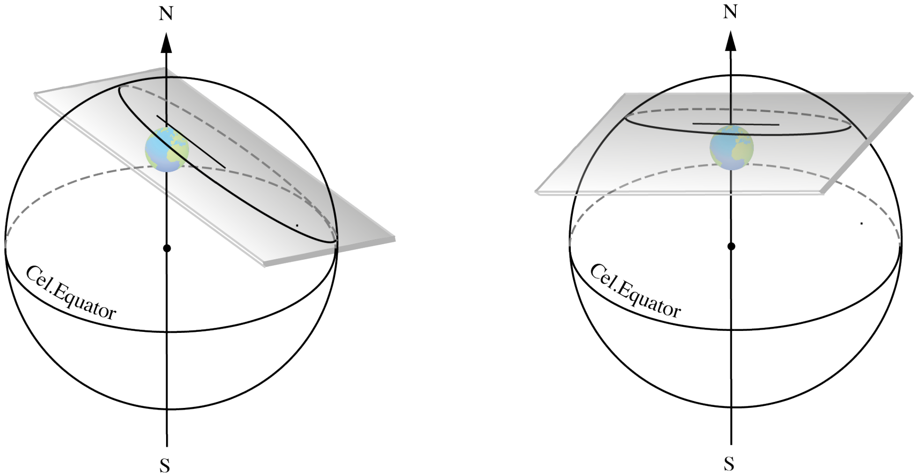


Figure 5.14: Displacement of the Earth along the rotational axis of the universe: the plane of the horizon would divide the heavens into unequal parts, varying for different latitudes.

Nevertheless, he does not use the physical arguments against the motion of the Earth in *Almagest* I, 7 to rule out the possibility of a tiny central displacement of the Earth. Unlike Aristotle, he seems to regard these arguments as irrelevant for demonstration of the centrality of the Earth. Ptolemy argues that the fall of bodies can be regarded as a corollary of geocentrism instead of an argument for it:[61]

> One can show by the same arguments [provided in support of the centrality of the Earth] that the Earth cannot have any motion in the aforementioned directions, or even move at all from its position at the center. […] Hence I think it is idle to seek for causes for the motion of objects toward the center, once it has been so clearly established from the actual phenomena that the Earth occupies the middle place in the universe, and that the heavy objects are carried toward the Earth.

Ptolemy, exactly like Aristotle (see Fig. 5.2, right) observes that the fall of heavy bodies toward the center is evident since[62] "the direction and path of the motion […] of all bodies possessing weight is always and everywhere at right angles to the rigid plane drawn tangent to the point of impact." Additionally, Ptolemy reviews a series of physical considerations which he perhaps derived from *De caelo*, although his opinions diverge from Aristotle's. Firstly, he discusses the fact that the Earth is not supported by anything in its position at the center of the universe. Unlike Aristotle,[63] he reassesses the "argument of equilibrium for indifference" ascribed to Anaximander in *De caelo*. In fact, we read, "that which is relatively smallest should be overpowered from and pressed in equally from all directions to

[61] *Almagest* I,7 p. 43.
[62] *Almagest* I, 7 p. 43.
[63] *De caelo* II, 13.

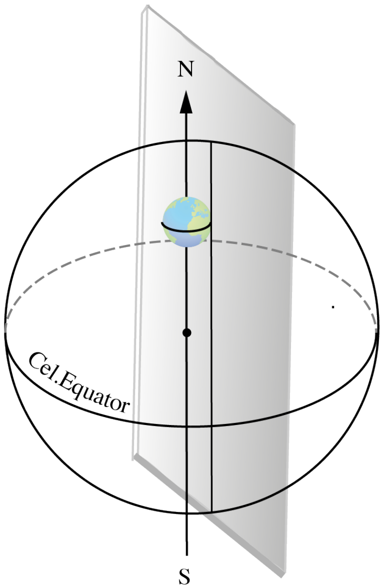


Figure 5.15: Displacement of the Earth along the rotational axis of the universe: Only at *sphaera recta* could the horizon bisect the celestial sphere.

a position of equilibrium."[64] Additionally, he affirms that there is no up-and-down motion in the universe, since directions depend on the observer. This statement is at odds with Aristotelian cosmology. In *De caelo* II,2 one reads that the heavens have an up and down, a right and a left, a back and forth. This idea is supported by an analogy between the heavens and animals, which are beings capable of moving themselves. In spite of his independence from Aristotle, Ptolemy shares his assumption that a body falls down faster the bigger it is.[65] This is also, according to him, an argument against the displacement of the Earth from its center, toward which it has a natural tendency. Moreover,[66] "living things and individual heavy objects would be left behind, riding on the air, and the Earth itself would very soon have fallen completely out of the heavens. But such things are utterly ridiculous merely to think of." Although physical arguments are not essential for demonstrating the centrality of the Earth, according to Ptolemy they are decisive for rejecting the axial rotation of the Earth, an issue which he explicitly tackles:[67]

> […] although there is perhaps nothing in the celestial phenomena which would count against that hypothesis, at least from simpler considerations, nevertheless from what would occur here on Earth and in the air, one can see that such a notion is quite ridiculous.

[64] *Almagest* I, 7 p. 43.
[65] *Almagest* I, 7 p. 44.
[66] Ibid.
[67] *Almagest* H25, p. 45.

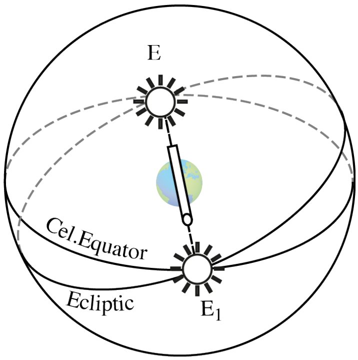


Figure 5.16: Pliny's argument: at equinoxes the sunrise and sunset points can be observed along the same line with a dioptra; therefore, the observer is located at the intersection of two great circles, i.e., one is placed in the middle of the universe.

As we have seen, Aristotle ascribes the hypothesis criticized here to Plato.[68]

Ptolemy's arguments against the axial rotation of the Earth (and terrestrial motion in general) became famous after Copernicus's refutation in the first book of his major work. They are basically derived from the excessive velocity of the terrestrial spin and the supposition that flying and thrown objects would be left behind by the terrestrial motion.

## 5.4 Conclusions and prospects

In the *Almagest*, Ptolemy's argumentative strategy in favor of geocentrism is the reverse of that employed by Aristotle in *De caelo* II, 13–14. Whereas the natural philosopher derived the centrality (and immobility) of the Earth from his theory of the elements, that is, from 'physical' observations and assumptions, the Hellenistic astronomer derived similar conclusions from geometrical-astronomical considerations. Aristotle explicitly regarded mathematical-astronomical arguments as secondary.[69] In his opinion, they merely corroborated his natural demonstration. In a certain sense, one can say that he built his cosmology on the basis of theories concerning terrestrial physics (hinged on the theory of the elements). At least in the relevant passages of the *Almagest*, Ptolemy reversed Aristotle's perspective,

[68] We do not agree with Pedersen's manner of comparing Aristotle's and Ptolemy's physics (Pedersen 1974, 43–45). On the one hand, Pedersen uncritically assumes the Aristotelian background of Ptolemy. On the other hand, he interprets *Almagest* I, 7 anachronistically and extrinsically, using expressions like "an immense pressure of the ether molecules." On top of this, Pedersen does not distinguish in his discussion between the theory of the central Earth turning about its axis and the heliocentric system. For a better insight into this issue, see Lerner 2008, 63–81, in which Michel-Pierre Lerner discusses what he calls "la critique sévère du systéme astronomique d'Aristote par Ptolémée." Ptolemy's criticism of Aristotle's natural conceptions mainly concerns the properties of the ether and heavenly bodies. Moreover, he rejects Aristotle's system of concentric spheres in favor of a model in which eccentric and epicycles are not only mathematical tools abstracted from reality, but have physical existence, as emerges from *Planetary hypotheses*.

[69] Cf. Cleary 1996, 191: "One of Aristotle's most significant steps in moving beyond Platonism was to replace mathematics with physics as the cosmological science par excellence. However, this does not mean that he rejected mathematics as a science relevant to cosmology, but rather that he subordinated it to physical inquiry."

as he considered physical arguments to be secondary: "Hence I think it is idle to seek for causes for the motion of objects towards the centre, once it has been so clearly established from the actual [astronomical] phenomena that the Earth occupies the middle place in the universe."[70] According to him, physics descends from cosmology and not the other way around. As we have seen, elementary observational phenomena, like the fall of bodies, do not require further explanation once the spherical form of the heavens and the centrality of the Earth have been demonstrated. It is precisely the inverse of Aristotle, for whom the theory of the elements comes first.

Still, to account for these divergent approaches to geocentricism, the classical distinction between mathematical and physical astronomy is not sufficient. Averroes and scholastic philosophers criticized several aspects of Ptolemaic astronomy from a natural or 'physical' perspective. The geometrical models for planetary motions seemed to be at odds with certain basic assumptions of Aristotle like the uniform circularity of celestial motions or the concentricity of heavenly spheres. Ptolemy was therefore accused of neglecting natural philosophy and his mathematical models were deemed unable to explain the real nature of the universe. Accordingly, it became customary to distinguish mathematics and physics, description of matters of fact (τὸ ὅτι or the *quia*) and causal explanation (διότι or *propter quid*).

This separation was still at work in the homocentric cosmologies of the early sixteenth century, as was the case with the Italian Aristotelians Amico and Fracastoro. In the framework of the Copernican debate, there were several attempts to distinguish mathematical from physical astronomy, in order to avoid a conflict between Copernican tables and Aristotelian physics. Mathematical astronomy was only supposed to provide useful models for celestial computation, whereas philosophy was supposed to deal with natural causes. Theologians were particularly severe in maintaining this distinction, which also entailed a hierarchic understanding of the levels of knowledge: mathematical, philosophical and, above that, revealed. Notably, this position was stubbornly supported by the Lutheran theologian Andreas Osiander, author of the conventionalist anonymous introduction to *De revolutionibus*, and later by the Catholic Inquisitor Robert Bellarmin, who played a decisive role in the trial of Bruno, the Galileo affair, and the censorship of the heliocentric theory in 1616. Both Osiander and Bellarmin limited mathematical astronomy to computation or, as Duhem put it, to "saving the phenomena" (σῴζειν τὰ φαινόμενα).

From our analysis it has become clear, however, that Ptolemy's and Aristotle's arguments for geocentrism cannot be traced back to the separation between abstract mathematical models and real physical causes (Averroism and scholastic philosophy) nor to the separation of computation and explanation (Osiander and Bellarmin). In fact, they show a more general divergence in the treatment of nature. This is an ontological and an epistemological difference at the same time. On the one hand, Aristotle tackles geocentricism from the perspective of a qualitative philosophy of nature, especially his theory of elementary motion. On the other hand, Ptolemy relies on a mathematical understanding of the cosmos as a whole. The former derives cosmology from terrestrial physics, whereas the latter proceeds in the opposite direction. It should be remarked that, in this respect, Copernicus would follow in Ptolemy's footsteps, claiming in *De revolutionibus*, Book One, that terrestrial physics should be corrected to agree with his general cosmological assumptions, in particular with

[70] *Almagest* I, 7.

the Earth's motion.[71] The divergence between Aristotle and Ptolemy is that between a qualitative and a mathematical approach to nature as well as that between a terrestrial and a heavenly perspective.

Concerning Ptolemy's epistemology – one could say, his 'mathematical epistemology' – an enlightening introduction to it is the first chapter of the first book of the *Almagest*, which contains fundamental philosophical considerations and claims. Ptolemy mentions the Aristotelian idea that there are three speculative disciplines, physics, mathematics and theology, possibly relying on *Metaphysics* V, 1. However, he alters Aristotle's perspective, since he exploits this quotation to extoll firstly the nobility of mathematics and even to hint, in the following, at the superiority of mathematics over the other two speculative disciplines. This superiority concerns at least the certainty of its demonstrations. Whilst philosophers will never reach agreement in their speculations owing to the profound uncertainty of their discipline, "mathematics can provide sure and unshakable knowledge to its devotees, provided one approaches it rigorously. For its kind of proof proceeds by indisputable methods, namely arithmetic and geometry."[72] In a very Platonic mood, Ptolemy surmises that mathematics gives access to divine things, because its objects occupy a position between the sensible and the intelligible, between the changing reality given to our perceptions and the eternal, unchanging realm of divinity.[73] With fruitful intuition, Ptolemy suggests that mathematics also helps physics "for almost every peculiar attribute of material nature becomes apparent from the peculiarities of its motion from place to place."[74] Needless to say, both the idea of a mathematical theology and that of a mathematical theory of motions are in contrast with Aristotle's metaphysics and his hylomorphic physics. Ptolemy adds some considerations on the providential design underlying nature, which owes much to Pythagoreanism and Platonism, even stoicism:[75]

> With regard to virtuous conduct in practical actions and character, this science, above all things, could make men see clearly; from the constancy, order, symmetry and calm which are associated with the divine, it makes its followers lovers of this divine beauty, accustoming them and reforming their natures, as it were, to a similar spiritual state.

A cosmological perspective like that of Ptolemy virtually entails a reversal of Aristotelian physics, once the arguments for terrestrial centrality are demonstrated to be invalid from an astronomical perspective, as Copernicus demonstrated in the first book of *De revolutionibus*. Copernicus's planetary system challenged physics from a cosmological perspective but did not challenge Ptolemaic epistemology, on which his method actually relied. Geoffrey E.R. Lloyd has pointed out that the dominant cosmological view of classic antiquity was anthropocentric. According to him, "the victory of geocentricity over heliocentricity was both a symptom and a cause of this."[76] Still, there is a profound difference between

[71] Notably, Koyré not only emphasized the Copernican dependency of physics on cosmology – of the development of a new dynamics on the heliocentric planetary astronomy – but even (and, in our opinion, unduly) generalized this dependency in order to account for the entire evolution of scientific thought from Copernicus to Newton. Cf. Koyré 1978, 131.
[72] *Almagest* p. 46.
[73] *Almagest* p. 46.
[74] *Almagest* p. 46.
[75] *Almagest* pp. 46–37.
[76] Lloyd 1991, 161.

the Aristotelian physical viewpoint and the Ptolemaic astronomical one. The former author, in fact, adhered to a geocentric model on the basis of a physics that is presented as closely linked to everyday commonsense experience. In this respect, Aristotle's natural philosophy seems to be profoundly anthropocentric. By contrast, Ptolemy's geocentrism is much less anthropocentric, if at all. It is consideration of the heavens that primarily defines the position of the Earth in the cosmos. Hence, one could ascribe to him the label 'cosmocentrist' that has been usually reserved for post-Copernican cosmologies such as that of Giordano Bruno or even, *ante litteram*, for Nicholas of Cuse's idea of an infinite universe.[77]

The "Aristotelian-Ptolemaic system" is a medieval and early modern product. In spite of their different approaches, the convergence of the general cosmological conclusions of *De caelo* and of *Almagest* led to a unified geocentric cosmology based on arguments derived from both sources, as Sacrobosco testifies. From the twelfth to the seventeenth century, university students learning the basics of spherical astronomy from Sacrobosco's *De sphaera* would receive the impression of a profound unity between the two principal sources of ancient cosmology, Aristotle and Ptolemy, in relation to the essential features of the cosmos and the reasons they brought forward. Sacrobosco traced his general cosmological views back to the authority of these two sources. In the section of *De sphaera* dealing with "quod terra sit in medio firmamenti" ("that the Earth is in the middle of the firmament"), Sacrobosco refers to the authority of "Ptolemaeus et omnes philosophi" ("Ptolemy and all the philosophers") abandoning any distinction between the mathematical astronomer and the natural philosopher.[78] As a matter of fact, he skipped, shortened or oversimplified the arguments of Aristotle and Ptolemy, and tended to present their shared opinions as part of the same conception.

Although the commentators of Aristotle, especially through Averroes, became aware of the contrast between the homocentric planetary model propounded by their 'master' and Ptolemy's epicyclic-eccentric geometrical devices, the image of an Aristotelian-Ptolemaic worldview as a unity was not abandoned and was even reinforced later as an effect of the post-Copernican debate. This fundamental agreement became almost a commonplace. According to Galileo's renowned *Dialogo sopra i due massimi sistemi del mondo*, for instance, only two major world systems existed: the Ptolemaic and the Copernican, the first one coinciding with the Aristotelian. Kuhn's account of the 'Copernican revolution' owes much to this interpretative schema. By contrast, this paper has pointed out the different, if not opposite, approaches in Aristotle's and Ptolemy's treatment of a fundamental cosmological issue in the context of which they are usually mentioned together: geocentricism. A renewed focus on epistemological tensions between the two main classics of cosmology pertaining to methodology and philosophy of knowledge helps us understand that there is no 'traditional', 'ancient' or 'Greek' cosmology. This suggests that the ancient world experienced a theoretical, philosophical and cultural diversity that can be easily overlooked from the modern perspective.[79]

[77] The classic treatment of the complex relationship between microcosm and infinity is Cassirer 1927.

[78] Thorndike 1949, 84 and 122.

[79] This pluralism has been clearly stressed, among others, by Lloyd 1991, 151: "There is not such a thing as *the* cosmological theory, of the Greeks. Indeed, one can and must go further: one of the remarkable features of Greek cosmological thought is that for almost every idea that was put forward, the antithetical view was also proposed." This is what Lloyd calls the "dialectical" character of Greek science.

In fact, not only have we often received a crystallized image of Greek knowledge, but we have also relied on works which are themselves great syntheses that overshadow and hide previous debates and multiple viewpoints. Just as Aristotle's *De caelo* superseded previous cosmologies, Ptolemy's *Almagest* superseded previous mathematical astronomy. The task of the historian of ancient cosmology should therefore be to highlight argumentative tensions, deconstruct the alleged unity of views of singular authors or epochs, and seek to obtain an insight into the cultural pluralism of debates that history and tradition have veiled.

## Bibliography


Aristotle (1986). *On the Heavens*. Aristotle in twenty-three volumes. Loeb classical library. Cambridge, MA: Harvard University Press.

Barker, Peter and Bernard R. Goldstein (1998). Realism and Instrumentalism in Sixteenth Century Astronomy: A Reappraisal. *Perspectives on Science* 6(3):232–258.

Beere, Jonathan B. (2003). Counting the Unmoved Movers: Astronomy and Explanation in Aristotle's 'Metaphysics' XII.8. *Archiv für Geschichte der Philosophie* 85:1–20.

al-Bitruji (1952). *De motibus celorum*. Ed. by Francis J. Carmody. Berkeley: University of California press.

Cassirer, Ernst (1927). *Individuum und Kosmos in der Philosophie der Renaissance*. Leipzig: Teubner.

Cleary, John J. (1996). Mathematics and Cosmology in Aristotle's Philosophical Development. In: *Aristotle's Philosophical Development: Problems and Prospects*. Ed. by William Wians. Lanham: Rowman and Littlefield, 193–228.

Di Bono, Mario (1990). *Le sfere omocentriche di Giovan Battista Amico nell'astronomia del Cinquecento*. Genova: Consiglio nazionale delle ricerche, Centro di studio sulla storia della tecnica.

— (1995). Copernicus, Amico, Fracastoro and Tusi's Device: Observations on the Use and Transmission of a Model. *Journal for the History of Astronomy* 26.

Dijksterhuis, E. J. (1986). *The Mechanization of the World Picture. Pythagoras to Newton*. Princeton: Princeton University Press.

Duhem, Pierre (1969). *To Save the Phenomena: An Essay on the Idea of Physical Theory from Plato to Galileo*. Chicago: The University of Chicago Press.

Goldstein, Bernard R. (1997). Saving the Phenomena: The Background to Ptolemy's Planetary Theory. *Journal for the History of Astronomy* 28:1–12.

Granada, Miguel Angel and Dario Tessicini (2005). Copernicus and Fracastoro: The dedicatory letters to Pope Paul III, the history of astronomy, and the quest for patronage. *Studies in History and Philosophy of Science* 36(431-476).

Heglmeier, Friedrich (1996). Die griechische Astronomie zur Zeit des Aristoteles: Ein neuer Ansatz zu den Sphärenmodellen des Eudoxos und des Kallippos. *Antike Naturwissenschaft und ihrer Rezeption* 6:51–71.

Jori, Alberto (2009). Erläuterung. In: *Über den Himmel*. Aristoteles: Werke in deutscher Übersetzung 12. Darmstadt: Wissenschaftliche Buchgesellschaft.

Kouremenos, Theokritos (2010). *Heavenly Stuff: The constitution of the celestial objects and the theory of homocentric spheres in Aristotle's cosmology*. Stuttgart: Steiner.

Koyré, Alexandre (1978). *Galileo Studies*. Atlantic Highlands: Humanities Press.

Lerner, Michel-Pierre (2008). *Le monde des sphères, Vol. 1: Genèse et triomphe d'une représentation cosmique*. Paris: Les Belles Lettres.

Lloyd, Geoffrey Ernest Richard (1991). Greek Cosmologies. In: *Methods and Problems in Greek Science*. Cambridge: Cambridge University Press, 141–163.

— (2000). 'Metaphysics' Λ 8. In: *Aristotle's 'Metaphysics' Lambda*. Ed. by Michael Frede and David Charles. Oxford: Clarendon Press, 245–273.

Moraux, Paul (1949). Einige Bemerkungen über den Aufbau von Aristoteles' Schrift 'De caelo'. *Museum Helveticum* 6:157–165.

— (1961). La méthode d'Aristote dans l'étude du ciel 'De Caelo' I 1 - II,12. In: *Aristote et les problèmes de méthode*. Louvain: Publications universitaires.

Omodeo, Pietro Daniel (2014). *Copernicus in the Cultural Debates of the Renaissance: Reception, legacy, transformation*. Leiden: Brill.

Panti, Cecilia (2001). *Moti, virtù e motori celesti nella cosmologia di Roberto Grossatesta: Studio ed edizione dei trattati 'De sphera,' 'De cometis,' 'De motu supercelestium'*. Firenze: SISMEL, Ed. del Galluzzo.

Pedersen, Olaf (1974). *A Survey of the 'Almagest'*. Odense: Odense Press.

Ptolemy (1984). *Ptolemy's Almagest*. Ed. by Gerald J. Toomer. London: Duckworth.

Reinhold, Erasmus (1551). *Prutenicae tabulae coelestium motuum*. Tubingae: per Ulricum Morhardum.

Rheticus, Georg Joachim (1959). Narratio prima. In: *Three Copernican Treatises*. Ed. by Edward Rosen. New York: Dover.

Schiaparelli, G. V. (1875). Le sfere omocentriche di Eudosso, di Calippo e di Aristotele. *Pubblicazione del Reale Osservatorio di Brera in Milano* 9.

Thorndike, Lynn (1949). *The 'Sphere' of Sacrobosco and Its Commentators*. Chicago: The University of Chicago Press.

# Chapter 6
# Space and Matter in Early Modern Science: The Impenetrability of Matter
*Peter Damerow*

## 6.1 The character of early modern science

The most conspicuous characteristic of early modern in contrast to ancient and medieval science is rightly considered the different kind of recourse to experience in the generation and justification of knowledge.[1] Results of observation and experiment were put forward against the authority of the Aristotelian doctrine. It is true that already in medieval times Aristotelian philosophy had undergone fundamental modifications – just recall, for example, how the various forms of the so-called theory of *impetus* complemented Aristotle's doctrine of motion – but such modifications were the result of internal developments within the framework of an Aristotelian consensus. They were rarely understood as radical alternatives to the Aristotelian doctrine. In contrast, what protagonists of the modern science of the sixteenth and seventeenth centuries had in common, beyond their emphasis on experience, was a downright programmatic anti-Aristotelianism. Modern science was not only supposed to open up new spheres for experience, it also aspired to a new kind of theory radically different from Aristotelianism, but with a comparable claim to universal validity.

This meant that during its first phase modern science was confronted with an insurmountable problem arising out of its internal structure. On the one hand, the experiential basis was far too small to construct a general theory of nature that was generally acceptable owing to its substantiation through empirical evidence. On the other hand, the very pretension of being able to compete with the theory of the Aristotelian tradition, which explained everything, made it imperative to come up with comprehensive theoretical systems that harbored comparable explanatory potential.

From this problematic situation, it is understandable why the insistence on observation and experimentation as the ultimate authorities to judge the truth of scientific propositions was not the only characteristic of the early phase of modern science. There was also the nearly desperate hope and expectation that, through philosophical reflection,[2] it would be possible to create an apodictic theoretical framework for the interpretation of empirical experiences, a framework justified solely on the principles of reason. This program was designated by a term going back to Aristotle: metaphysics.

[1] This chapter is based on a lecture given by Peter Damerow in 1994 at the University of Constance. The supervision of the translation from German into English and the inclusion of additional notes left by Peter Damerow were done by MS. The central concept of the chapter, *conceptual models* (*Modellvorstellungen*), is closely related to the *mental models* introduced in Chapter 1. In particular, the concept is used here to illuminate how metaphysical ideas result from the absolutization of experiences. – [MS]

[2] This reflection was actually not all that different in principle from speculative ancient philosophy.

The topic of this chapter concerns an idea typical of this modern metaphysics, an idea which runs as a thread throughout its development: the idea of distinguishing between space and matter by assuming that matter has the property of being impenetrable.

## 6.2 Ancient atomism

The early modern idea that matter has the essential property of being impenetrable has its origins in ancient philosophy, namely in a thought experiment. If one conceives of the perceptible bodies of everyday experience as being divided into ever smaller parts, the following alternative appears to be inevitable: Either the division can, theoretically, be continued infinitely; or at some point one runs up against a fundamental limit, against the *indivisible*, the *atom*. And if this smallest unit of matter is to be indivisible then matter must be impenetrable by its nature, that is, no matter can reach the place of other matter without pushing it away.[3]

The assumed existence of such impenetrable atoms of matter became the point of departure for the ancient philosophical tradition of atomism. In their speculative endeavor to reduce the coming-into-being and passing-away of reality to a permanent, rationally conceivable explanation, Greek philosophers developed the idea that the world of macroscopic phenomena and occurrences could be explained by processes in a microscopic world which is no longer accessible by our senses but only by reason. According to this explanatory model the founders of ancient atomism, the Greek philosoper Leucippus and his disciple Democritus, promoted the idea that the qualities of things we perceive are not qualities of these things themselves but rather represent the way our senses react to qualities of these things. Aetius the philosopher wrote in his *Opinions of Philosopers* about the Greek atomists:[4]

> The atomists taught that everything is entirely colorless; sensory qualities emerge from (bodies) without qualities, which are perceivable only by reason.

According to the ancient atomists, our senses merely lead us to believe the colorful world of sensory impressions. In fact there are only atoms, their shapes and their movements, which affect our senses.[5]

The postulated atoms are different from the objects of our direct experience in yet another respect. Since they are supposed to be indivisible they must have characteristics principally different from these objects, for otherwise there would be no reason why the process of division could not be continued. According to tradition, Leucippus and Democritus thus characterized atoms using a pair of absolute concepts that cannot be traced back to any other concepts. The atoms are the 'full', the 'solid', the 'being', while the space in which they

---

[3] Thus, Lucretius writes: "[…] you must yield and confess that there are things which no longer consist of any parts and are of the smallest possible nature. And since these exist, you must also confess that the first-beginnings are solid and everlasting" (*De rerum natura* I, 624–627; translation taken from Lucretius Carus 1992, 51–53).

[4] Aetios 1,15,11, see Diels 1951-1952, Vo. 2, 112; for a German translation, see Jürß, Müller, and Schmidt 1988, 175 (Fragment 209).

[5] Thus, Aristotle writes: "Democritus […] and Leucippus postulate the 'figures' and make 'alteration' and coming-to-be result from these, attributing coming-to-be and passing-away to their dissociation and association, and 'alteration' to their arrangement and position; […]." Aristotle *On Coming-to-be and Passing-Away* 315b, 7–9, translation taken from Aristotle 1992, 173.

move is the ‘void’, the ‘nothing’, the ‘infinite’. According to Simplicius, Leucippus and Democritus[6]

> thought them [i.e. the atoms] to be uncuttable, indivisible, and unaffectible on account of their being compact and having no share of void. For they said that division occurred because of the void in bodies, and that these atoms are separated from one another in the infinite void, and, differing in shape and magnitude and position and order, they move through the void and meet with and collide with one another, and that some bounce off in whatever way they will, while others become intertwined with one another because of the congruity of their shapes, magnitudes, positions, and orders, and it so happens that the generation of composite things is accomplished.

This speculative structure describing the microscopically small contains tacit assumptions about the relationship between space and matter, which were later to constitute the core of the problem to be discussed here. Apparently it is presumed as a matter of course that atoms cannot penetrate each other. Matter is granted the property that, wherever it is located, no other matter can get there without displacing it from its location. Matter, the solid, thus differs from the space in which it is located, the void, through a fundamental impenetrability. This property presents a constitutive prerequisite for the assumption of the indivisibility of atoms and thus for the atomistic explanation of creation, decay and change. The space of the atomists is an absolute void. It does not provide any resistance against matter and thus cannot exert any causal effects. Matter, by contrast, is absolutely impenetrable, such that collisions of atoms bring forth causal effects, which appear as creation and decay or as change on the macroscopic level.

Wherein lies the persuasiveness of such assumptions about the structure of the microscopically small? Doubtlessly in the fact that the structures attributed to the microscopically small appear to have such obvious validity in the macroscopic world. Thus Lucretius, for instance, justifies the basic assumptions of atomism as follows:[7]

> But now to resume my task begun of weaving the web of the discourse: the nature of the universe, therefore, as it is in itself, is made up of two things; for there are bodies, and there is void, in which these bodies are and through which they move this way and that. For sensation common to men declares that body has its separate existence; and unless our belief in sensation is first firmly established, there will be no principle of appeal in hidden matters, according to which we may establish anything by the reason. Then further, if there was no place and space which we call void, bodies could not be situated anywhere nor could they move anywhere at all in different directions […].

Atoms thus were assigned properties resembling those of the objects in the macroscopic space that surrounds us. But there is a decisive difference between the world of macroscopic things and the world of atoms. In contrast to the properties of macroscopic things, properties in the world of atoms are absolute properties. The thought experiment’s step of declaring

[6] Simplicius *On Aristotle On the Heavens*, 242, 17–27; translation taken from Simplicius 2004, 64.
[7] Lucretius *De rerum natura*, I, 418–428; translation taken from Lucretius Carus 1992, 35–37.

atoms to be absolutely indivisible entails further absolutizations. The matter of the ancient atomists is absolutely impenetrable and space is absolutely empty.[8]

## 6.3 The revival of atomism in the renaissance

Ancient atomism experienced a renaissance in the corpuscular theories of the sixteenth and seventeenth centuries. It was attractive to the representatives of modern science not only as an alternative to Aristotelianism, but above all because it corresponded to the metaphysical goal of explaining the world immanently from principles of reason. Atomism depicted the world as the result of an interplay among its smallest parts, an interplay that is complex and yet determined by laws, and about whose nature one could apparently gain insight through reasoning.

A typical example of recourse to the ancient theoretical tradition is presented by the corpuscular theory of Robert Boyle, with which he attempted to explain chemical reactions in particular.[9] As for the atomists of antiquity, for Boyle, too, the macroscopic properties of matter were secondary qualities, that is, ways of perceiving the primary qualities of size, shape and motion of the microscopically small particles of which matter actually consists. The program he pursued aimed to trace the secondary qualities back to two principles:[10]

> I should likewise, after all this, explain to you how, although matter, motion and rest, seemed to me to be the catholick principles of the universe, I thought the principles of particular bodies might be commodiously enough reduced to two, namely *matter*, and (what comprehends the two other, and their effects) the result, or aggregate, or complex of those accidents, which are the motion or rest, (for in some bodies both are not to be found) the bigness, figure, texture, and the thence resulting qualities of the small parts, which are necessary to intitle the body whereto they belong to this or that peculiar denomination; and discriminating it from others to appropriate it to a determinate kind of things, (as yellowness, fixtness, such a degree of weight, and of ductility, do make the portion of matter wherein they concur, to be reckoned among perfect metals, and obtain the name of gold) this aggregate or result of accidents you may if you please, call either *structure*, or texture (though indeed, that do not so properly comprehend the motion of the constituent parts especially in case some of them be fluid) or what other appellation shall appear most expressive.

Yet there are two essential differences between Boyle's corpuscular theory and ancient atomism. For one, Boyle was convinced that he did not only have to *postulate* the microscopic corpuscles hypothetically, he believed that they could actually be *proven* using empirical methods. Second, Boyle believed that the movements of the corpuscles were subject to the laws of mechanics formulated in the modern era. Through Boyle's 'mechanical explanations' the macroscopic properties of matter were thus supposed to be traced back to the mechanical laws, which were obtained macroscopically, but now applied to microscopic corpuscles.

[8] Cf. the quote from Lucretius in note 3.
[9] Cf. Boyle 1666.
[10] Boyle 1937, 201.

For this immanent explanation of the world based on mechanical laws, Boyle selected a metaphor common in his day: The world functions like a mechanical clock composed of innumerable small parts, which, once set in motion by its creator, continues moving in accordance with strict laws without any further intervention.

## 6.4 Consequences of mechanical models

Modern corpuscular theories like that of Boyle were thus, like ancient atomism, associated with constitutive assumptions about the relation between space and matter. Such assumptions drew their persuasiveness from being based on conceptual models that were transferred from the macroscopic sphere, in which they appeared evident, to the microscopically small.

In ancient atomism these conceptual models represented the absolutized elementary structures of experiencing objects.[11] By contrast, in the corpuscular theories of early modernity conceptual models became more varied and differentiated, because the results of modern empirical evidence were transferred to the corpuscles as well, especially results from the science most highly developed at the time: mechanics. Thus the modern mechanical world view emerged from ancient atomism through mechanical conceptual models.

The variety of such conceptual models entailed a corresponding variety of metaphysical foundations for natural science.[12] Among the controversial points was particularly the question as to whether there actually were in fact absolutely empty spaces. This question had always been negated in the Aristotelian tradition, as opposed to ancient atomism, for according to Aristotle's law of motion, the speed of a moving body would become infinitely great in a space without resistance.

A number of modern corpuscular theorists also found the atomistic assumption that atoms move in absolutely empty space problematic, albeit for different reasons. Giordano Bruno, for instance, explicitly drew on Democritus and Leucippus but did not follow the ancient atomists in the question as to the existence of a vacuum, because to him the cohesion of bodies did not appear certain if they consisted only of atoms in motion. According to his view, matter would disperse into infinity if there were not another kind of matter, which he designated the *ether*. This "glued together and encompassed"[13] the atoms, as water did for the particles of the Earth.

The ancient atomists absolutized elementary experiences that can be gained in the handling of corporeal objects and had thus arrived at the opposition of impenetrable matter and empty space. Referring to Epicurus's atomistic theory, but criticizing its dualism of full and empty, Bruno, in his *On the Infinite Universe and Worlds*, lets his protagonist Filoteo explain his alternative to the Aristotelian Elpino as follows:[14]

> We do not call aught Void as being mere nullity, but rather accept the view whereby that which is not corporeal nor doth offer sensible resistance is wont,

[11] 'Absolutization' is here to be understood as elevating cognitive structures such as properties attributed to everyday objects (e.g. the hardness of a billiard ball) to fundamental principles (e.g. the indestructibility of atoms). – [MS]
[12] Owing to the simplicity of its basic assumptions about the relationship between absolutely impenetrable matter and space as an absolute void, ancient atomism constituted little more than a general background for the modern corpuscular theories. After all, these multifarious theories were based on complex, often mutually incompatible assumptions about processes on the microscopically small level.
[13] Cited after Lasswitz 1984, 378.
[14] Singer 1968, 273.

> if it hath dimension, to be named Void, since we do not usually understand as corporeal that which hath not the property of offering resistance; whence they say that just as that is not flesh which is not vulnerable, so that which doth not offer resistance is not corporeal. In the same way we name infinite that which is an immense ethereal region in which are innumerable and infinite [numbers of] bodies such as the earth, the moon, and the sun, and these are called by us worlds, composed of Plenum and of Void: for this spirit, this air, this ether not only surroundeth these bodies but also penetrateth within them and becometh inherent in everything.

Bruno, like the ancient atomists, transferred experiences from the world of macroscopic things to the world of the microscopic, but his assumptions about the relation between matter and space were oriented on the conceptual model of an all-pervading fluid.[15]

Clearly, the question as to whether there can be a vacuum, an absolutely empty space, is not only a theoretical question but also an empirical one. In fact, even back in Scholastic natural philosophy observations of phenomena that were caused by air pressure, such as the way a clepsydra or siphon works, had been related to this question. The fact that a fluid does not flow out of a closed vessel or a narrow tube as long as no air can flow into it, was explained through a *horror vacui*, which prevented the occurrence of an absolutely empty space.

Galileo Galilei also referred to such observations to explain his corpuscular theory, the basic assumptions of which also deviated from those of ancient atomism regarding the existence of an empty space. In contrast to Bruno, Galileo already used his deliberations on the relation of matter and space in the microscopically small to elucidate many kinds of macroscopic, physical and technical issues. According to Galileo, while matter is certainly impenetrable, space is generally not empty, as this is prevented by the horror vacui, for which he mentions numerous empirical observations.

Galileo was already aware, however, that the horror vacui has only a limited magnitude. From the observation that water can only be pumped up to a certain height, which Galileo gave fairly accurately as 18 cubits,[16] and from the correct theoretical explanation that at this height the horror vacui, i.e. air pressure, is surpassed by the weight of the water column, he drew the conclusion that the horror vacui represented only a limited force, which cannot absolutely prevent the occurrence of an empty space. Galileo further assumed that for solids the horror vacui prevents atoms from separating from each other, and derived from this an explanation for the relative solidity of macroscopic bodies. Only if the fine particles of fire in lively motion push themselves between the atoms, as for example when metals melt, can they be separated without a void occurring between them.[17]

The assumption that space is not empty, but rather completely filled with matter, represented a profound change in atomism, one which inevitably led to further modifications to its basic assumptions. For instance, ancient atomism had a plausible explanation for the fact

[15]Once it has been introduced, the ether takes on further explanatory functions, which are not rooted in the metaphysical system and cannot be substantiated directly with the original conceptual model. For Bruno this is, above all, the relation between spirit and matter, which is identified with the differentiation between ether and all other matter.

[16]Galilei 1974, 24–25.

[17]Galilei 1974, 27.

that matter can change its volume,[18] which could no longer be valid, since it was based on the assumption that material bodies contained empty spaces in the form of pores that grew larger in the process of rarefaction and smaller in the process of condensation.

Galileo solved the problem as to how "condensation and rarefaction […] can be understood to take place without assuming interpenetration of bodies and [at the same time] without introducing void spaces […]"[19] through yet another drastic change to the basic assumptions of atomism, which he oriented on a model from geometry: Galileo's atoms are indeed impenetrable and indivisible but, like the equally indivisible points of a line, they are also infinitely small, and their number, like those of points on a line, infinitely large; between them are infinitely many infinitely small empty spaces.[20] On the basis of this assumption Galileo explained the possibility of condensing or rarefying matter, analogous to the possibility of projecting a small line onto a large one, that is, by uniquely assigning the points of the two lines to each other.[21]

## 6.5 The rationalist program

The examples outlined above may already be sufficient to show how the structural problem of modern science presented itself to its proponents. Common to these theories was that they were anti-Aristotelian and drew orientation from ancient atomism. The perceptible properties of objects and the changes in these properties were traced back to movements of a matter that obeys absolute mechanical laws on the microscopic level, and these laws were deduced from experiences with macroscopic objects. But owing to the alteration of the experiential basis in early modern times, this approach inevitably resulted in deviations from the traditional, canonical ideas of ancient atomism; and owing to the multiplicity of possible conceptual models it resulted in theories that could hardly be reconciled with each other. The diversity of such theories stood in blatant contradiction to the claim to apodictic certainty raised by metaphysical foundations.

It seemed obvious to blame this unfortunate situation on the fact that the claim to a rational justification was not sufficiently fulfilled, and to aspire to a more methodologically controlled approach as a way out of the situation. This was the goal of rationalism, and in particular of its outstanding representative René Descartes.

Descartes' attempt to determine the relationship between matter and space strictly, without resorting to empirical experiences, led him to a radical solution. He assumed "that the nature of matter […] consists not in its being something which is hard or heavy or coloured, or which affects the senses in any way, but simply in its being something which is extended in length, breadth and depth."[22] Descartes attempted to explain that there was no difference at all between matter and space. Through this the foundations of natural science were to be traced back entirely to mathematical considerations, which at the time still appeared to be indisputably independent from experience.

[18] Galileo offers as an extreme example "the boundless rarefaction of a small amount of gunpowder, when it is resolved into a vast bulk of fire" (Galilei 1974, 64).
[19] Galilei 1974, 64.
[20] Galilei 1974, 33.
[21] He writes: "In this way there would be no contradiction in expanding, for instance, a little globe of gold into a very great space without introducing quantifiable void spaces – provided, however, that gold is assumed to be composed of infinitely many indivisibles" (Galilei 1974, 33–34).
[22] Descartes *Principles of Philosophy*, II, 4; Descartes 1998, 224.

Descartes' identification of space and matter had a number of drastic consequences for all of the problems connected with the relation between space and matter. Thus, for instance, the question as to the existence of empty spaces became senseless, since Descartes' space was, by definition, filled with matter.[23] Nor could there be any indivisible atoms for Descartes. Like space, Descartes' matter was infinitely divisible.[24] Further, he had to conceive of material bodies as purely kinematic phenomena, as matter with a homogeneous state of motion.[25] Consequently, he thus formulated the following definition:[26]

> By 'one body' or 'one piece of matter' I mean whatever is transferred at a given time, even though this may in fact consist of many parts which have different motions relative to each other.

Finally, the impenetrability of matter became a tautology for Descartes, for if space and matter are identical, then it is a logical contradiction to assume that two different bodies could be located in the same place, that is, penetrate each other.

Descartes meticulously justified his metaphysical principles, proceeding from a theory of human cognition. Thus he emphasized that the matter-space structure he described follows strictly from the postulates of reason. Yet his rationalistic system was certainly not a deductive theory in the modern sense. In particular it is obvious that he, too, was able to make his identification of space and matter plausible only through conceptual models. So, for instance, in order to elucidate that motion was possible even in a space completely filled with matter, he compared the structure of matter with the pool of a fountain, in which fish move as freely as if they were in an empty space.[27] Just as the water moves in a circular motion around the fish, so that no empty space emerges in the place the fish just left, so he assumed "that no motion ever takes place which is not circular."[28]

In a similar manner Descartes explained why the possibility of condensing or rarefying matter did not contradict his theory, which eliminated the possibility of empty pores: he compared matter with a sponge soaked with liquid. According to this idea, a body can, depending on the conditions, absorb a greater or smaller amount of a finer kind of matter and thus increase or reduce its volume.[29]

Here, too, we thus encounter a plausible conceptual model which is absolutized for the sphere of the microscopically small. While in ancient atomism the assumption of impenetrable and indivisible atoms moving in a void was based on the absolutized conception of

---

[23] "The impossibility of a vacuum, in the philosophical sense of that in which there is no substance whatsoever, is clear from the fact that there is no difference between the extension of a space, or internal place, and the extension of a body" (Descartes *Principles of Philosophy*, II, 16; Descartes 1998, 229–230).

[24] "We also know that it is impossible that there should exist atoms, that is, pieces of matter that are by their very nature indivisible [...]. For if there were any atoms, then no matter how small we imagined them to be, they would necessarily have to be extended; and hence we could in our thought divide each of them into two or more smaller parts, and hence recognize their divisibility. For anything we can divide in our thought must, for that very reason, be known to be divisible; so if we were to judge it to be indivisible, our judgement would conflict with our knowledge" (Descartes *Principles of Philosophy*, II, 20; Descartes 1998, 231).

[25] "All the variety in matter, all the diversity of its forms, depends on motion" (Descartes *Principles of Philosophy*, II, 23; Descartes 1998, 232).

[26] Descartes *Principles of Philosophy*, II, 25; Descartes 1998, 233.

[27] Descartes *The World*, Chapter 4; Descartes 1998, 86–87. Cf. Descartes *Principles of Philosophy*, II, 17; Descartes 1998, 230.

[28] Descartes *The World*, Chapter 4; Descartes 1998, 87.

[29] Descartes *Principles of Philosophy*, II, 6 and 7; Descartes 1998, 225–226.

macroscopic objects in space, for Descartes' infinitely divisible matter that completely fills space it is the relations in a liquid and the vortices of matter which generate the movement of objects within it.

The fact that Descartes borrowed from experience in this way did not escape his contemporaries' attention, such that his rationalistic system had an ambivalent effect. On the one hand, after Descartes it was hardly possible for metaphysics to continue to avail itself of speculation as naively as had been characteristic for the pioneers of the previous generation. Yet on the other hand, his very attempt at a consistent, rationalistic justification of the foundations of natural science gave rise to doubts about the program of metaphysics, because it seemed no longer possible to rule out that any metaphysical justification would sooner or later run into conflict with the empirical orientation of modern science.

## 6.6 Newton's hidden atomism and the problem of force

Isaac Newton's attitude toward Descartes' rationalism can be regarded as representative for this awakening, fundamental doubt about the program of a metaphysical justification of natural science.

Newton's unfinished and unpublished manuscript *De Gravitatione ...*, in which he "venture[d] to dispose of [Descartes'] fictions,"[30] and in particular of the latter's identification of matter and space, already conveys not only a criticism of Descartes' specific assumptions about the relation between matter and space, but also a fundamental criticism of the very idea of a metaphysical foundation for such assumptions.

Newton wanted his own assumptions, which he formulated in a series of definitions and propositions, to be understood as being "either definitions of certain words; or axioms and postulates denied by none."[31] Newton avoided atomistic formulations of his assumptions, although the fact that they at least partly originated in atomistic theories could hardly be denied.

For instance, Newton defined a body as "that which fills place," and added the note:[32]

> I said that a body fills place, that is, it so completely fills it that it wholly excludes other things of the same kind or other bodies, as if it were an impenetrable being.

Further down he formulates:[33]

> [...] I suppose in these definitions that space is distinct from body [...].

Nevertheless, in the manuscript, Newton avoided any commitment concerning the question as to whether empty spaces exist.

In addition to such assumptions, which correspond to those of ancient atomism, Newton formulated further definitions and propositions which clearly reflect the historical distance to ancient atomism. He defined force (*vis*) as "the causal principle of motion and rest," tendency (*conatus*) as "resisted force," *impetus* as impressed force, *inertia* as "force within a body, lest its state should be easily changed by an external exciting force," pressure as the

[30]Newton 1978, 123.
[31]Newton 1978, 122.
[32]Newton 1978, 122.
[33]Newton 1978, 123.

tendency “of contiguous parts to penetrate into each others’ dimensions,” and gravity as “a force in a body impelling it to descend.”[34]

Newton’s definition of pressure, according to which pressure is only a tendency of two parts to penetrate each other, “[f]or if they could penetrate, the pressure would cease,”[35] especially brings out the contrast with an absolutized concept of impenetrability. In contrast to ancient atomism, the experience of impenetrability, that is, the experience that two bodies cannot be moved to the same place at the same time, is here brought into relation with the magnitude of resistance that a macroscopic body offers to a second body being in contact with it, a magnitude that is measurable in experiments.

In his *Principia* Newton makes clear that he regards his own determinations of matter, recorded by definitions and axioms, as the results of empirical experiences:[36]

> That all bodies are impenetrable we gather not by reason but by our senses. We find those bodies that we handle to be impenetrable, and hence we conclude that impenetrability is a property of all bodies universally. That all bodies are movable and persevere in motion or in rest by means of certain forces (which we call forces of inertia) we infer from finding these properties in the bodies that we have seen. The extension, hardness, impenetrability, mobility, and force of inertia of the whole arise from the extension, hardness, impenetrability, mobility, and force of inertia of each of the parts; and thus we conclude that any one of the least parts of all bodies is extended, hard, impenetrable, movable and endowed with a force of inertia. And this is the foundation of all natural philosophy.

For the existence of an absolute space independent of matter, Newton gave empirical proof with what is known as his ‘bucket experiment’. According to his view the water propelled upward by centrifugal force in a rotating bucket shows that one can distinguish between the rotating bucket and a bucket at rest; as such it proves the existence of absolute space in reference to which the rotation takes place. Likewise he invoked arguments for the existence of empty spaces, for instance the negligible resistance the celestial bodies experience in their movements,[37] as well as the possibility of rarefying matter to an unlimited extent.[38]

Newton attempted to avoid basic assumptions of atomism (such as the assumption of the absolute impenetrability of matter), because by their very nature they could not be confirmed directly in experimental experience. However, this attempt proved difficult to sustain. There is a great variety of evidence that Newton often tacitly accepted the validity of such assumptions. Newton was an atomist, but the basic assumptions of atomism were introduced into his work rather incidentally.

---

[34]Newton 1978, 148. In Newton 1978, *conatus* is translated as ‘endeavour’, not as ‘tendency’.

[35]Newton 1978, 148.

[36]Newton *Principia*, Book III, “Rules for the study of natural philosophy,” Rule 3; translation taken from Newton 1999, 795–796.

[37]Thus, in his *Opticks* Newton writes: “[…] to make way for the regular and lasting Motions of the Planets and the Comets, it’s necessary to empty the Heavens of all Matter, except perhaps some very thin Vapours, Steams, or Effluvia, arising from the Atmospheres of the Earth, Planets, and Comets, and from such an exceedingly rare Æthereal Medium as we described above” (Newton 1979, 368).

[38]Thus, Newton writes: “But if the quantity of matter in a given space could be diminished by any rarefaction, why should it not be capable of being diminished indefinitely?” (Newton *Principia*, Book III, Proposition 6, Corollary 3; Newton 1999, 810)

So, for instance, in his *Opticks* Newton explained chemical reactions as rearrangements of particles, which were caused by attractive forces between them. But such an explanation presupposes atomistic matter and impenetrability, such that Newton here was led to note:[39]

> All Bodies seem to be composed of hard Particles [...]. And therefore Hardness may be reckon'd the Property of all uncompounded Matter. At least, this seems to be as evident as the universal Impenetrability of Matter.

Tacit assumptions were associated also, and above all, with Newton's concept of force. Through Newton's *Principia* the concept of force became a central concept of the theory of nature, which, in the form of attractive forces like gravitation, was intimately related to novel basic assumptions about the properties of matter. Changes in the motion of material bodies are caused by forces, and the material bodies for their part affect other bodies through forces.

The concept of the impenetrability of matter could not remain unaffected by such a substantial expansion of the concept of matter. In the atomistic tradition the changes in the motions of two colliding bodies were attributed to the impenetrability of matter. According to Newton's axioms every change in motion is necessarily caused by forces. This then raises the question as to the nature of the forces that act upon the colliding bodies as a consequence of the absolute impenetrability of matter.

Newton never supplied his contemporaries with an answer to such questions. He never presented the foundations of his theory of matter in a form that would have been comparable with Descartes' *Principles of Philosophy*.

## 6.7 Euler's solution

There was certainly no shortage of attempts to reconcile Newton's findings with Descartes' rationalistic foundation of science. Leonhard Euler's foundations of mechanics constituted an important attempt of this kind. On the high technical level of eighteenth-century analytical methods, Euler created a mathematically formulated Cartesian theory of matter and its motion in space, which was based on three basic metaphysical determinations of matter held to be apodictic: its *extension*, its *inertia*, and its *impenetrability*.[40] The occurrence of forces, in contrast, Euler conceived to be a phenomenon derived from impenetrability:[41]

> The impenetrability of bodies, therefore, contains the real origin of the forces, which are continually changing their [i.e., the bodies'] state in this world; and this is the true solution of the great mystery, which has perplexed philosophers so grievously.

When two bodies collide, according to Euler, their impenetrability causes the exertion of a force that changes their motions. Euler solved the problem Newton left unresolved, that the magnitude of this force cannot be determined from impenetrability alone, by introducing

---

[39] Newton 1979, 389.

[40] Euler *Letters to a German Princess*, Letter no. 121 (21 April 1761); Euler 1823, Vol. 2, 17, in this edition it is letter no. 6 of Vol. 2.

[41] Euler *Letters to a German Princess*, Letter no. 77 (18 November 1760); Euler 1823, Vol. 1, 233.

a supplementary principle: The force that occurs is the smallest possible force that is able to prevent the penetration of matter.[42]

Connecting Cartesian principles with Newtonian dynamics inevitably led to an elaboration and change in the Newtonian concept of force. As Euler pointed out, because Newton's inertia was not a force in the sense of the Newtonian axioms, the designation as a force was misleading.[43] A further result was the differentiation between absolute forces and forces of constraint, with which matter resists forces acting upon it. In this manner Euler's mechanics became the point of departure for later attempts to formulate a generalized principle of inertia and, using this principle, develop a forceless, consistently kinematic mechanics; the mechanics of Heinrich Hertz serves as an example from the nineteenth century.

### 6.8 Kant's anti-atomistic solution

Immanuel Kant's concept of matter represents a break in so far that with it the distinction of matter from space through its impenetrability was for the first time principally challenged. For Kant this distinction was inacceptable:[44]

> Absolute impenetrability is in fact nothing more nor less than an occult quality. For one asks what the cause is for the inability of matters to penetrate one other in their motion, and one receives the answer: because they are impenetrable.

What did Kant have to set against this? From the difficulty of reconciling the assumption of the impenetrability of matter with the implications of Newtonian dynamics, Kant drew conclusions that were diametrically opposed to those of Euler, bringing him in extreme opposition to the mechanistic foundation of natural science.[45] Instead of presupposing the impenetrability of matter and asking about the forces that result from this impenetrability in the case of the collision of material bodies, as Euler did, he assumed conversely that the bodies appear to us to be impenetrable only because a force counteracts the penetration. In Kant's view, what actually distinguishes matter from space are "original forces," which constitute its essence, namely an attractive force, which we perceive as gravity, and a repulsive force, which appears to us as its impenetrability. He writes:[46]

---

[42] Euler *Letters to a German Princess*, Letter no. 78 (22 November 1760); Euler 1823, Vol. 1, 233–236; see also: Euler *Theoria motus corporum solidorum sev rigidorum*, §134; Euler 1765, 50, Euler 1948, 65; for a German translation, see Euler 1853, 59.

[43] Euler *Mechanica*, Praefatio; Euler 1912, 10, for a German translation, see Euler 1848, 5.

[44] Kant *Metaphysical Foundations of Natural Science*, Chapter 2, Explication 4, Remark 2; Kant 2004, 39.

[45] "The mechanical mode of explanation […] has, under the name of *atomism* or the *corpuscular philosophy*, always retained its authority and influence on the principles of natural science, with few changes from Democritus of old, up to Descartes, and even to our time. What is essential therein is the presupposition of the *absolute impenetrability* of the primitive matter, the *absolute homogeneity* of this material, leaving only differences in the shape, and the *absolute insurmountability* of the cohesion of matter in these fundamental particles themselves" (Kant *Metaphysical Foundations of Natural Science*, Chapter 2, "General Note to Dynamics"; Kant 2004, 72).

[46] Kant *Critique of Pure Reason*, A265, B321; translation taken from Kant 1996, 327. The impenetrability of matter is traced back to the interaction of forces not only in Kant's writings during his critical period (after 1781), but also in a relatively early tract, the *Monadologia physica* of 1756, yet with a decisive difference: Here matter exists as an entity independent of forces. It consists of the smallest, indivisible components, which Kant called *monads* in keeping with Leibniz's terminology, and which, as for Leibniz, are granted the capability to exert forces. Each monad is surrounded by a *sphaera activitatis*, a sphere of activity through which it keeps other monads away (Satz VI; Kant 1907, 351), and the repulsive force with which this occurs is perceived empirically as impenetrability (Satz

> We are acquainted with substance in space only through forces that are active in space: either in propelling other substances toward the substance (attraction), or in preventing them from penetrating into the substance (repulsion and impenetrability); we are not acquainted with other properties making up the concept of the substance that appears in space and that we call matter.

As a consequence of this reduction of the concept of matter to the concept of force, for Kant it followed that matter, like space, was infinitely divisible[47] and "originally elastic."[48] It retreats from attacking forces until the growing repulsive forces are able to resist the attacking forces.[49]

According to this theory, solid bodies are constituted by the coaction of the attractive and repulsive forces. The boundary between two bodies is that surface on which the repulsive forces of the respective matters balance each other.[50] In the case of a single body in empty space, a comparable boundary of the border results from the fact that the attractive force of matter, according to Newton's law of gravitation, decreases in inverse proportion to the square of the distance, while the repulsive force, according to Kant's assumption, decreases in inverse proportion to the cube of the distance. This means that the repulsive force predominates in the vicinity, but decreases more rapidly with distance than the attractive force. Thus, for every isolated body in space there must be a closed surface in which the repulsive and the attractive forces of the body balance each other. According to Kant, this surface is the boundary of the body; for the penetration of this surface is resisted by the repulsive force, while outside this surface the repulsion cannot be perceived, because it is smaller than the attraction.[51]

Through Kant's redefinition of the concept of matter, not only impenetrability as the characteristic property of matter was traced back to the concept of force, but so was the concept of matter itself. Matter in its conventional form was declared to be a metaphysical relic that does not stand up to the *Critique of Pure Reason*. In opposition to the concept

---

VIII; Kant 1907, 353). Here Kant apparently tried to establish a connection with Newton's theory by physically reinterpreting Leibniz's monads.

[47] "Matter is *divisible to infinity*, and, in fact, into parts such that each is matter in turn" (Kant *Metaphysical Foundations of Natural Science*, Chapter 2, proposition 4; Kant 2004, 40).

[48] Kant *Metaphysical Foundations of Natural Science*, Chapter 2, proposition 2; Note 1; Kant 2004, 36–37.

[49] "Matter can be compressed to infinity, but can never be penetrated by a matter, no matter how great the compressing force of the latter may be" (Kant *Metaphysical Foundations of Natural Science*, Chapter 2, proposition 3; Kant 2004, 37). Kant writes in summary: "The *action* of the universal attraction immediately exerted by each matter on all matters, and at all distances, is called *gravitation*; the tendency to move in the direction of greater gravitation is *weight*. The action of the general repulsive force of the parts of every given matter is called its *original elasticity*. Hence this property and weight constitute the sole universal characteristics of matter, which are comprehensible a priori, the former internally, and the latter in external relations. For the possibility of matter itself rests on these two properties" (Kant *Metaphysical Foundations of Natural Science*, Chapter 2, proposition 8, Note 2; Kant 2004, 56–57).

[50] "Physical contact is the interaction of repulsive forces at the common boundary of two matters" (Kant *Metaphysical Foundations of Natural Science*, Chapter 2, Explication 6, Remark; Kant 2004, 50).

[51] "Thus the original attraction of matter would act in inverse ratio to the squares of the distance at all distances, the original repulsion in inverse ratio to the cubes of the infinitely small distances, and, through such an action and reaction of the two fundamental forces, matter filling its space to a determinate degree would be possible. For since repulsion increases with the approach of the parts to a greater extent than attraction, the limit of approach, beyond which no greater is possible by the given attraction, is thereby determined, and so is that degree of compression which constitutes the measure of the intensive filling of space" (Kant *Metaphysical Foundations of Natural Science*, Chapter 2, Proposition 8, Remark 1; Kant 2004, 59).

of matter from atomism or corpuscular theory, in which matter represents the presupposed carrier or causer of forces, Kant's concept of matter consequently eliminated any substrate that was assumed prior to the original forces of attraction and repulsion. Matter then consists of nothing more than forces operating in space.

If we allow ourselves an anachronistic comparison, Kant's concept of matter resembles the later concept of the field; and his theoretical aspiration is reminiscent of Einstein's unachieved goal of a unified field theory, in which the source terms of the field equations disappear, i.e., the fields are no longer generated from matter independent of the fields.[52]

Yet Kant's goal was not to erect a new theory of physics. Rather, with his *Critique of Pure Reason* he wanted to counter metaphysical speculation with the methodically controlled, precise determination of the most general theoretical statements of science, statements founded in reason and not in experience. He hoped, through the self-reflection of the reasonable subject (*vernünftiges Subjekt*), to extract those criteria that would permit him to distinguish between the metaphysical illusion and the apodictic truth of general judgments (*Allgemeinurteile*) independent of experience.

Kant criticized the metaphysical argumentations of his day not because he did not share the goal of grounding the theoretical framework of empirical knowledge on the principles of reason, but rather because they did not live up to this program. Kant's question was not the question of the empiricists: 'How can metaphysics be avoided in science?' but rather: 'How is metaphysics possible as a science?'

Kant formulated this program in his *Critique of Pure Reason*, and elaborated it for the field of natural science in the *Metaphysical Foundations of Natural Science*. In particular, he concluded that the following ideas do not originate from experience:

- space and time and, thereby, the laws of kinematics;
- the general properties of matter, which is constituted by attraction and repulsion, originally elastic and infinitely divisible;
- the law of conservation of mass (in Kant called "quantity of matter");
- the first and third Newtonian axioms, i.e. the law of inertia and the principle of the equivalence of acting and reacting forces (but not the proportionality of force and acceleration); the consequence for Kant was
- a law of collision (albeit an erroneous one); and
- the non-existence of Newton's absolute space.

## 6.9 The impact of Kant's criticism

The above listing alone makes clear that Kant's metaphysical foundation of natural science, too, would not bear the test of time. Nearly all of his supposedly apodictic assumptions, obtained from a critique of reason, have subsequently turned out to be of only limited validity or even wrong.

In fact, Kant's concept of matter, too, like those of his predecessors, was based on conceptual models. What distinguishes Kant's foundation, however, is the fact that his conceptual models for the conceptualization of the microscopically small hardly incorporated any categories of everyday experience unreflectively, but were based almost exclusively on

[52] Einstein 1992, 71 ff. see also Einstein's letter to Felix Pirani, 2 February 1954 (Einstein Archives 17-447.00).

categories of the most advanced areas of contemporary natural science. Compared with the particles of matter and the ethereal fluids of corpuscular theoreticians, Kant's centers of force thus appear as an extraordinarily modern theoretical construct. Indeed, in the post-Kantian theories of matter such theoretically constructed conceptual models were further developed into powerful methodological instruments of modern natural science.

In contrast, the program of a metaphysical foundation for natural science was directed toward other objectives and doomed to failure. To pursue this agenda, the protagonists of modern science, as was here demonstrated for the case of the concept of matter, brought forth an enormous variety of metaphysical systems, whose simultaneous incompatibility and claim to universality represent the clearest indication for the failure of metaphysics.

Kant subjected the presuppositions of metaphysical systems to a systematic critique. He did not see that his own assumptions about judgments independent of experience were just as problematic as those against which his critique of reason was directed. In the *Metaphysical Foundations of Natural Science* he leveled the accusation at Newton that he[53]

> by no means dared to prove this law a priori, and therefore appealed rather to *experience*.

History has shown that Newton's caution in this point was all too justified.

**Bibliography**


Aristotle (1992). *On sophistical refutations. On coming-to-be and passing-away. On the Cosmos*. Aristotle in twenty-three volumes. Loeb classical library. Cambridge, MA: Harvard University Press.

Boyle, Robert (1666). *The Origine of Formes and Qualities: According to the corpuscular philosophy*. Oxford: Davis.

— (1937). *The Sceptical Chymist*. Ed. by Ernest Rhys. Reprint. London: Dent.

Descartes, René (1998). *The philosophical writings of Descartes, Vol. 1*. Ed. by John Cottingham, Robert Stoothoff, and Dugald Murdoch. Cambridge: Cambridge University Press.

Diels, Hermann (1951-1952). *Die Fragmente der Vorsokratiker*. 6th edition. revised by W. Kranz. Berlin: Weidmann.

Einstein, Albert (1992). *Autobiographical Notes: A centennial edition*. Ed. by Paul Arthur Schilpp. La Salle: Open Court.

Euler, Leonhard (1765). *Theoria motus corporum solidorum seu rigidorum: Ex primus nostrae cognitionis principiis stabilita et ad omnes motus, qui in huis modi corpora cadere possunt, accommodata*. Rostock: Röse.

— (1823). *Letters of Euler on Different Subjects in Natural Philosophy: Addressed to a German Princess, 2 vols.* Ed. by David Brewster. Edinburgh: Tait.

— (1848). *Leonhard Euler's 'Mechanik oder analytische Darstellung der Wissenschaft von der Bewegung': Vol. 1*. Ed. by Jakob Philipp Wolfers. Greifswald: Koch.

— (1853). *Leonhard Euler's 'Mechanik oder analytische Darstellung der Wissenschaft von der Bewegung': Vol. 3*. Ed. by Jakob Philipp Wolfers. Greifswald: Koch.

— (1912). *Leonhardi Euleri Opera omnia: Vol. 2.1*. Leipzig: Teubner.


[53] Kant *Metaphysical Foundations of Natural Science*, Chapter 3, Proposition 4, Remark 1; Kant 2004, 88.

Euler, Leonhard (1948). *Leonhardi Euleri Opera omnia: Vol. 2.3*. Leipzig: Teubner.

Galilei, Galileo (1974). *Two New Sciences: Including centers of gravity & force of percussion*. Ed. by Stillman Drake. Madison: University of Wisconsin.

Jürß, Fritz, Reimar Müller, and Ernst Günther Schmidt, eds. (1988). *Griechische Atomisten: Texte und Kommentare zum materialistischen Denken der Antike*. Leipzig: Reclam.

Kant, Immanuel (1907). *Immanuel Kants kleinere Schriften zur Naturphilosophie*. Ed. by Otto Buek. Leipzig: Dürr.

— (1996). *Critique of Pure Reason: Unified edition; with all variants from the 1781 and 1787 editions*. Indianapolis: Hackett.

— (2004). *Metaphysical Foundations of Natural Science*. Cambridge: Cambridge University Press.

Lasswitz, Kurd (1984). *Geschichte der Atomistik vom Mittelalter bis Newton, Bd. 1: Die Erneuerung der Korpuskulartheorie*. 2nd print of the Hamburg und Leipzig 1890 edition. Hildesheim: Olms.

Lucretius Carus, Titus (1992). *De rerum natura*. Ed. by William Henry Denham Rouse and Martin Ferguson Smith. Loeb Classical Library 181. Cambridge, MA: Harvard University Press.

Newton, Isaac (1978). *Unpublished Scientific Papers of Isaac Newton: A selection from the Portsmouth Collection in the University Library, Cambridge*. Ed. by Alfred Rupert Hall and Marie Boas Hall. Cambridge: Cambridge University Press.

— (1979). *Opticks or a Treatise of the Reflections, Refractions, Inflections & Colours of Light*. Based on the 4th edition London, 1730. New York: Dover.

— (1999). *The 'Principia': Mathematical Principles of Natural Philosophy*. Berkeley: University of California Press.

Simplicius (2004). *On Aristotle On the Heavens 1.5–9*. Ed. by R.J. Hankinson. London: Duckworth.

Singer, Dorothea Waley (1968). *Giordano Bruno: His Life and Thought. With Annotated Translation of His Work: On the Infinite Universe and Worlds*. New York: Greenwood Press.

# Chapter 7
# Experience and Representation in Modern Physics: The Reshaping of Space

*Alexander Blum, Jürgen Renn and Matthias Schemmel*

> The essential in what I strove for during
> my long life centers on the question:
> What can be methodically concluded for
> physics from the fact of a universal law of
> light propagation and from the equivalence
> of inertial and gravitational mass?
>
> *Albert Einstein*[1]

## 7.1 Introduction

It is common knowledge that the two theories of relativity have fundamentally changed our notions of space and time. Spacetime, a union of space and time, is now conceived as a four-dimensional manifold equipped with a Lorentzian metric, in general relativity it has furthermore become dynamic. That the two theories should have such a sweeping impact on the concepts of space and time is no matter of course. After all, they started out as theories dealing with specific physical phenomena, such as electromagnetism and gravitation, phenomena which were conceived to occur *in* space and time, so that it was not obvious that their description should bring about changes in these concepts. Quantum theory, on the other hand, did not by itself change our notions of space and time, at least not yet in the form of an established theory. All surprising aspects of the quantum world have eventually been attributed to new and strange properties of matter and radiation, not of space and time. At the outset, one might have speculated that the non-local features of quantum systems required a modification of space and time concepts,[2] but up to now, quantum theory has turned out to be quite conservative as regards spacetime structures.

---

[1] "Alles Wesentliche was ich im Laufe eines langen Lebens erstrebt habe, gruppiert sich um die Frage[:] Was kann für die Physik methodisch geschlossen werden aus der Thatsache eines universellen Gesetzes der Lichtausbreitung und aus der Gleichheit der trägen und schweren Masse?" This is Albert Einstein's response to an inquiry of the *Technische Rundschau*, Bern, 31 January 1955; Albert Einstein Archives 1–199. We are grateful to Diana Buchwald, editor in chief of the *Collected Papers of Albert Einstein*, for pointing our attention to this quotation and for providing the translation.

[2] Thus, Max Born speculated in 1919 "that the way out of all quantum difficulties has to be sought starting from fundamental considerations: one must not transfer the concepts of space and time as a four-dimensional continuum from the macroscopic world of experience to the atomistic world, the latter obviously demands another kind of manifold of numbers as an adequate image." (Max Born to Wolfgang Pauli, 23 December 1919; the German original has: "daß der Ausweg von allen Quantenschwierigkeiten von ganz prinzipiellen Punkten aus gesucht werden muß: man darf die Begriffe des Raumes und der Zeit als eines 4-dimensionalen Kontinuums nicht von der

The historical examples of the relativity and quantum revolutions raise the general questions:

- Which experiences led to the establishment of new space and time concepts in the history of modern physics?
- Why did these specific experiences have this consequence and why at this particular historical moment?

From the standpoint of a strict conventionalism, these questions should make no sense. In spite of the fact that historically only specific theories did shape space and time, conventionalism argues that it is a mere matter of convenience, whether we decide to change our concepts of space and time or rather the concepts of the things and events filling them.[3] But is it really only a matter of convenience that we do not pursue relativity theory in Newtonian space and time and quantum theory with classical matter and radiation in a novel quantum spacetime?

The conventionalist's viewpoint neglects the fact that the concepts of space, time, and matter have their origin in pre-scientific structures of thinking and, no matter how advanced they may have become, derive part of their operational meaning from this origin.[4] By contrast, the conventionalist viewpoint presupposes that they may be decoupled from their origin without losing meaning.

Conventionalist arguments also often brush over the question of the historical availability of the postulated alternative formulations. As a matter of fact it is usually not easy to come up with theoretical constructions that describe physical phenomena as either being tied up with the space-time framework or as taking place within it. While it is difficult to prove the non-existence of one of the conceivable alternative constructions, historically there is always only a limited but evolving repertoire of possibilities. And even where such alternative space-time formulations of a specific physical theory are available, the novel space-time structures involved will not in general be transferrable to phenomena outside of the domain of the theory. This would not be compatible with the expectation that all of physics takes place within one and the same arena of space and time.

Why do certain experiences have an effect on the structure of space and time rather than simply modifying the physics of things in space and time, while others do not? The above discussion shows that it is not possible to dismiss such questions on conventionalist grounds. From the perspective of an historical epistemology, we formulate three criteria that together constitute the necessary conditions for representations of physical experience to have an impact on the space-time structure of physics.

Experiences may have an impact on concepts of space and time, if

- *Constructibility*: the available means for their representation allow for a consistent description of space and time structures distinct from the pre-existing ones;
- *Operationability*: the novel space-time structures incorporate prior knowledge on space and time, be it former theoretical or pre-theoretical knowledge, from which the new theoretical constructs can derive a spatio-temporal meaning;

makroskopischen Erfahrungswelt auf die atomistische Welt übertragen, diese verlangt offenbar eine andere Art von Zahlenmannigfaltigkeit als adäquates Bild." Pauli 1979, 10).

[3]For a detailed account on conventionalism, and its geometrical version in particular, see Ben-Menahem 2006.

[4]This also appears to be the point of Einstein's insistence on the importance of rods and clocks as "independent concepts" (Einstein 2001, 213) in relativity theory; see Einstein 1921.

- *Universality*: the novel space-time structures are applicable to all domains of physical knowledge.

These criteria may be illustrated by a simple example familiar from discussions of conventionalism.[5] Imagine a universe in which local measurements have led to the conclusion that space is Euclidean. Now introduce an 'expansion field' extended through all of space. It describes for each point in spacetime a local expansion of bodies at that point. An illustration of such an expansion field would be a reduced notion of temperature affecting all bodies in the same way. Length and shape of rigid rods (as well as of all other bodies) indeed depend on the temperature of their immediate environment. Suppose that the 'temperature' field varies on large scales across space. If rods are used by the people of astro-geometers to establish the large-scale geometry of space, the result may well be that it is non-Euclidean. Is this now an artifact of the temperature field or should this be interpreted as the discovery of a new space structure? The new space concept satisfies the operationability criterion because space measurements are performed in the usual way, just on larger scales. If the mathematical tool of a generalized metric is available, it then becomes possible to conceptualize the measurements with the expandable rods in terms of a Riemannian space, so that the criterion of constructibility is also fulfilled. This shows the historical contingency introduced by the constructibility criterion – indeed the present thought experiment did emerge only in the wake of the introduction of non-Euclidean geometries. But the requirement of universality may become problematic. Real physical temperature does not affect all bodies in the same way, unlike our hypothetic expansion field. Also, there may be other legitimate forms of space measurement, e.g. using light signals. Only if several independent types of space measurements give convergent results, does it become plausible to accept these new experiences as establishing a new space concept.

Experiences to which these criteria are applied have to be conceived of as always being 'pre-processed', i.e., processed prior to entering a new spacetime framework or other theoretical discourses. The essence of pre-processing is to assimilate any kind of input to pre-existing internal or external knowledge representations. The question of whether experiences give rise to new spacetime structures thus translates into the question of whether such pre-existing knowledge representations can be transformed or re-interpreted as such spacetime structures.

In the case of the expandable rods, the pre-existing knowledge representation to which results of measurements are assimilated is that of Euclidean space. The failure of this assimilation causes puzzlement and does not automatically lead to the establishment of a new spacetime structure. As we have argued, such establishment requires the above criteria to be fulfilled. Given that the pre-existing spacetime structure fulfills the universality criterion as a default, the most natural reaction to such a failure is to search for a problem in the specific measurement procedure. This procedure has not just spacetime aspects but also involves other physical properties such as the rigidity of the rods and their behavior under temperature changes, which might play a role for the failure. Only if it turns out that all conceivable measurement rods universally display the troubling behavior, does a change of spacetime structure become a plausible option. However, as soon as it turns out that spatial distances can also be measured by means of light, and that such measurements confirm the original Euclidean predictions, the notion of an alternative spacetime structure will be quickly dis-

[5] See Poincaré 1952, 65–68; see also Feynman, Leighton, and Sands 1989, 42/1–42/5.

carded. One might, of course, imagine a universe in which the index of refraction is related to temperature in such a manner that the measurements based on light will always produce the same results as those based on rigid rods. Only in this case the new measurements display a universality that might hint at that of a new spacetime structure, which at this point might become more plausible than the assumption of a correlation of index of refraction and temperature. There is, however, no guarantee that this new structure can actually be constructed. Indeed, for understanding historical processes, the most important criterion is that of constructibility. In principle there are two cases. Either an appropriate representation allowing the assimilation of the new experiences is already available, because it has been independently developed, or such representation has to be constructed by accommodating the existing structures to the new experiences in conflict with them. Our astro-geometers may thus either rely on the hitherto neglected mathematical knowledge and realize that the world at large scales is in fact Riemannian, or they eventually manage to develop non-Euclidean geometry as the large-scale geometry of their world.

A side remark may be in order. The above criteria may be applicable to conceptual developments more generally, even when less fundamental concepts are at stake than those of space and time. This is obvious for the criterion of constructibility. In more general cases the criterion of operationability may be reformulated in terms of the interpretability of novel constructs in terms of prior related concepts. The criterion of universality may be reformulated accordingly as a question of the domain of applicability of the new concept.

In the following section, we will briefly sketch aspects of the historical development of the relation between the concepts of space and time, up to classical physics, whose space concept is based on the exclusion of gravitation (section 7.2). We will then describe the impact of the growing corpus of experiential knowledge on optical and electromagnetic phenomena on the classical concepts of space and time (section 7.3). Next we discuss the re-introduction of gravitation into space-time-concepts, as it was brought about by general relativity (section 7.4). We will then discuss quantum theory as the case of a theory that had no comparable impact on the concepts of space and time (section 7.5). Finally, we shall come back to the questions raised at the beginning of this chapter concerning the experiences leading to new space and time concepts, and address them on the basis of the cases discussed.

## 7.2 The relation between space and time before relativity

One of the most striking novel features of relativity theory is the emergence of the concept of spacetime connecting space and time into one entity. Within relativity, time and space can only be separated depending on the state of motion of a particular observer. This particular relation between space and time can only be articulated within a specific theoretical framework. More generally speaking, however, space and time are connected in many ways within other layers of knowledge. In particular, as cognitive structures space and time have their origin in the experience of motion and change. In sensorimotor experience, spatial, temporal, material, and causal aspects are not differentiated from the beginning.[6] Yet, separate mental models of space, time, object, and cause may develop in accommodating to this experience.

[6]See, e.g., the discussion in Piaget 1959.

The specific characteristic of these models will depend on historically specific experiences and the possibilities to establish relations among them, for instance in societies that develop elaborate metrologies. In the course of history, experiential spaces have expanded considerably, including the development of instrument-based measurement operations. Only when spatial and temporal experiences and their linguistic representations become the subject of explicit reflections, general concepts of space and time can be constructed that are no longer confined to specific experiential domains, as is, e.g., the case in Aristotle's natural philosophy. At the same time, it must be possible to relate these general concepts somehow to the existing experiences, thus rendering these general concept operationable to some extent. This does not imply that these general concepts are connected to instrument-based measurement operations, as they underlie, e.g., Euclidean geometry. While the emergence of instrument-based measurement operations is a common phenomenon in complex societies such as the early civilizations of Mesopotamia, Egypt, and China, the independent emergence of explicit reflections on linguistic representations of spatial experiences have been a rather rare phenomenon.[7]

In all theoretical traditions prior to relativity theory, however, space and time largely occur as separate concepts. This is all the more astonishing as the individual concepts of space and time vary significantly among different theoretical traditions. This separation is therefore probably inherited from intuitive thinking. In particular, the elementary identification of stable entities such as objects and places implies a distinction from the aspect of change, the latter eventually giving rise to the concept of time. Temporal change, moreover, covers a much wider experiential field of which spatial change, i.e. motion, is only one particular case.

This does not preclude, however, close connections between spatial and temporal aspects within thinking, in theoretical as well as in practical and specifically in metrological contexts. The way in which elementary spatial and temporal experiences are processed depends on the available means of representations, first and foremost language. Such symbolic representations may vary among cultures but nevertheless typically display or generate a number of structural parallels between spatial and temporal categories. One instance of such parallels is the use of spatial linguistic markers in representing time which may leave its mark in theoretical reflections on space and time. Another instance is the use spatial properties in the metrology of time. This symbolic connection between space is not yet present in intuitive concepts of time which merely presuppose the correlation of different processes. Clocks make use of this in a practical way by selecting a process to which other processes can then be related, thus establishing a symbolic representation of the passage of time that can be shared. An immediate representation of points in time and of time intervals by spatial magnitudes is then established by the fact that for most clocks the process in question is a change of spatial position. Theoretical reflections may either start from such practical experiences or from the linguistic representations and the parallels of spatial and temporal experiences embodied by them.

As soon as spatial concepts and magnitudes are represented by the means of geometry, they could also be used for representing other magnitudes including (as is evident from the above) time. This possibility is for instance implicit in Euclid's concept of magnitude. Another matter is a combined geometric representation of space and time. A first extensive

[7] See Schemmel 2016; see also Chapter 1.

use of geometric techniques for representing non-spatial magnitudes is found in the work of Nicole Oresme from the fourteenth century. It occurred in the context of the scholastic elaboration of Aristotle's theory of change. Oresme developed diagrams that depicted the variation of a quality, such as heat, whiteness, or grace, over an extension, which could represent either temporal or spatial dimensions. The extension is typically represented by a line. The intensity of the quality in each point of the extension is represented by a vertical line of a given length. The extensive use, exploration, and ensuing transformation of such diagrams in the study of motion eventually led to a fundamental re-interpretation, in early modern times, of this geometrical technique in terms of a space-time-diagram for motion.[8]

While motion is naturally described in terms of space and time, it is also intimately connected to the notion of causation. However, the borderline between a descriptive (space-time) kinematics and an explanatory (causal) dynamics is subject to fundamental historical change. The main reason is that the causes of motion may themselves be connected to the structure of space and time. An extremely influential theoretical model of space and motion is Aristotelian natural philosophy, which originated in Greek antiquity and was the prevalent philosophical framework in Europe well into early modern times.

Aristotelian space is anisotropic, because it privileges directions, e.g., towards or away from the center of the universe. Space (or place, more precisely) participates in the causation of motion, since all bodies strive to reach their natural place. Motions towards natural places happen by themselves and are classified as natural motions, whereas motions caused by extrinsic forces are classified as violent motions. The eternal motions of celestial bodies are a class of their own and are supposed to proceed along circles.

Extensive exploration of this explanatory model – focused on planetary motion on the one hand and mechanical motion on the other – highlighted its problematic aspects and led to its transformation. The establishment of a new model of planetary motion given by the Copernican system constituted a major challenge, because the neat distinction between celestial and terrestrial physics in the Aristotelian model became questionable. This undermined the Aristotelian concept of spherically structured space, favoring a homogeneous and isotropic conception of space. The establishment of insights such as the parabolic shape of the projectile trajectory similarly constituted a challenge, because the neat distinction between natural and violent motion made it difficult to explain how this shape comes about. Instead it turned out to be more plausible to explain this on the basis of a uniform inertial motion and an accelerated motion of fall caused by a force.

The adaptation of the Aristotelian explanatory scheme to these challenges eventually led to the establishment of Newtonian mechanics and its explanation of celestial and terrestrial motions. Newtonian space is homogeneous and isotropic. Uniform motions (i.e., motions in a straight line with constant speed) are not in need of a causal explanation. Only accelerated motions need a causal explanation. The cause is either a force independent of the structure of space, or the inertial force caused by space in reaction to such an independent force.

In Newtonian physics, space offers an absolute frame of reference, which cannot be detected, since rest and uniform motion with respect to it are physically equivalent. This non-observability of the absolute standard of rest led to a reformulation introducing a privileged class of reference systems, called inertial systems. While in Aristotelian and Newtonian

[8] Schemmel 2014.

physics, motion was defined with respect to space, it is now the inertial structure of space that is derived from the motion of a particular type of bodies, namely 'free particles'. The space of classical mechanics is highly counter-intuitive, because space no longer offers landmarks with respect to which all motion can be described, and only allows for the identification of accelerated but not of uniform motions. The introduction of the inertial system concept was only possible after classical mechanics had been consolidated to an extent that such abstractions could be made, which, while being counter-intuitive, were uniquely suited to the formal structure of the theory.

What are the implications of Newtonian physics for the relation between space and time? Newtonian physics relates space and time by distinguishing one particular kind of motion: uniform motion in a straight line. The content of Newton's first law, the law of inertia, can be expressed by stating that free particles move uniformly along straight lines.[9] From this perspective, a unified description of space and time as spacetime becomes first reasonable, because uniform motion in a straight line can be described by straight lines in four-dimensional spacetime. No spacetime metric and not even an affine connection is required for this. A projective structure is sufficient.[10]

Uniform motion in a straight line is distinguished in Newtonian physics by not being in need of a causal explanation. Only deviations from the straight lines in four-dimensional projective space are in need of a causal explanation, namely by forces. This suggests a separation between space and time on one hand and physical processes in space and time on the other: spacetime with its inertial structure serves as an unchanging stage for physical processes that require causal explanation. Newtonian physics thus separates inertia and gravitation: inertia is part of the space-time structure and gravitation is part of the material structure of the universe. The resulting projective structure of Newtonian spacetime is universal and trivial. If gravitation is itself considered part of the spacetime framework, this framework becomes non-trivial, namely locally variable and dynamic. Obviously, to integrate gravitation with the description of inertial motion, one is in need of the principle of equivalence. A weak form of this principle is already an ingredient of Newtonian physics, although not a fundamental one. We shall later come back to the question of why Newton and his contemporaries did not formulate a strong principle of equivalence. Here we only note that the non-trivial geodesics resulting from it may have spoken for a four-dimensional formulation of physics. But still there would have been no four-metric in Newtonian spacetime. Historically, the four-metric entered mechanics through the back door of optics and electrodynamics. This will be discussed in the following section.

## 7.3 The impact of electromagnetism on the classical concepts of space and time

Since the notion of space of classical mechanics had been developed on the basis of mechanical experiences only, new perspectives on space emerged when experience from other

[9] See Pfister 2004 for a particularly lucid exposition of this formulation of Newton's first law, including definitions of 'free particle' and 'straight line'.

[10] See Pfister 2004, 56–58. Historically, the idea of space having a Euclidean measure and time flowing uniformly, which Newton explicitly adhered to, predated the formulation of the law of inertia. Note, however, that this does not imply the existence of a spacetime metric, which does indeed not exist for Newtonian spacetime. (Ehlers 1973 and later Earman 1989 designate a spacetime with Euclidean space-measure and a time-measure by the term 'Leibnizian' spacetime.)

branches of physics, in particular optics, was taken into account.[11] A wave theory of light suggested an ether, which might serve as a new standard of motion and even as a standard of rest if it has no internal motion. As such the concept of a stationary ether presents a modification of the concept of space of classical mechanics defined by the relativity of inertial motions. Indeed, historically, the further elaboration of ether-based physics led to substantial modifications of the concept of space, not by introducing a standard of rest, however, but rather by reasserting the principle of relativity for all of physics.

Initially, the absence of internal motion in the ether was strongly suggested by the phenomenon of stellar aberration, because it was possible to consistently define relative motion (of a source and an observer) as the difference between two absolute motions with respect to the ether. The rest frame defined by the ether played a more overt role in the derivation of the ray-optical law of refraction (Snell's law), because, in general, the derivation of ray optics from wave optics holds only in the rest frame of the ether. It could thus seem even plausible that the laws of refraction might be used to establish absolute motion. In particular, the direction of light that determined the magnitude of the angle of refraction was expected to be the direction of light with respect to the ether at absolute rest. However, it turned out that the experimentally observed refraction could be reproduced by naively inserting into Snell's law the direction of light actually observed in the rest frame of the medium. This fact had of course been essential for the phenomenological formulation of Snell's law. From the perspective of an ether theory, it was in need of an explanation. Such an explanation was provided by Augustin Fresnel who assumed that the medium slightly drags along the ether (in dependence on its refractive index), thereby effectively compensating any explicit effect of absolute motion relative to the ether.

The Michelson-Morley experiment was designed to probe the motion of the Earth with regard to the ether. To this end, the travel times of light rays in different directions with regard to the motion of the Earth were compared by the means of observing phase differences of light revealed by displacements of interference patterns. The non-observability of such phase differences allowed for the conclusion that the velocity of the Earth with respect to the ether must be less than one fourth of the Earth's orbital velocity, implying nearly complete dragging of the ether by the Earth, in contrast to the small dragging effect proposed by Fresnel and strongly supported by Hippolyte Fizeau's experiment measuring the index of refraction in moving media.

When optics became part of Maxwell's electrodynamic theory, realizing that light was an electromagnetic wave-phenomenon, these problems were transposed into a new context: that of an electrodynamics of moving bodies. Early attempts at formulating such a theory by Hertz and Heaviside assumed that the ether was fully dragged along inside matter, and could thus explain why the motion of the Earth did not have to be taken into account in the explanation of any terrestrial electric or magnetic phenomena. They were unable, however, to explain Fizeau's experiment or, for that matter, deliver a satisfactory theory of optics as a branch of electrodynamics. Whereas a slightly dragged-along ether seemed to be the most simple explanation for the unobservability of the motion of the Earth with regard to the ether in optics – with the exception of the Michelson-Morley experiment –, a totally dragged-along ether seemed the simple-most assumption to explain the unobservability of the motion of the Earth in the realm of electric and magnetic phenomena. The reason behind

[11] This section is partly based on a close rereading of Janssen and Stachel 2004 and personal discussion with Michel Janssen.

these different perspectives is the fact that the 'optics of moving media' was a field in which rich experiences were available, leading to such sophisticated assumptions as the Fresnel coefficient and partial dragging, whereas the electrodynamics of moving bodies was merely a formal extension of electrodynamics at rest, which was most easily achieved by simply assuming that the equations of an electrodynamics in motion were the same as those at rest, realized by the assumption of a totally dragged-along ether. Through the Michelson-Morley experiment, however, this merely formal hypothesis gained physical plausibility, clashing as it does with the optical evidence in favor of an essentially stationary ether.

With a further elaboration of electromagnetic theory it became possible to explain effects that had formerly been explained mechanically, such as ether dragging, by electromagnetic interactions. The simplest possible assumption in the sense of avoiding the introduction of convoluted hypothesis about the ether's motions was to postulate that it was at rest, an assumption key to Lorentz's elaboration of Maxwell's theory in the 1890s. On this basis it became the main challenge to explain all effects that had earlier been explained by some form of ether dragging.

Lorentz could account for the Fresnel coefficient by introducing an atomistic model of matter allowing him to account for refraction and other optical phenomena on the basis of an interaction between incoming radiation, the motion of the charged atomistic constituents of matter, and the secondary radiation produced by it.[12] This secondary radiation interfered with the primary radiation to produce the macroscopic refracted ray. In summary, the partial dragging of the ether could be replaced by local interactions of the *radiation* carried by the immobile ether with the microscopic constituents of bodies otherwise freely moving through the ether. The calculations leading up to this result were rather involved, because the Maxwell equations were assumed to hold only in the rest frame of the ether. This amounts indeed to a remarkable paradox of Lorentz's world view. While Maxwell's equations had originally been established by terrestrial experiments on an Earth that rapidly traverses the ether, the assumption of a stationary ether obliged him to assume that their validity in a terrestrial frame was actually little more than a mere coincidence.

However, Lorentz could simplify the calculations when he observed symmetries between the equations for dielectric matter at rest in the ether and for dielectric matter in motion.[13] Indeed, the two sets of equations coincided to order $v/c$ if in the equations for moving matter one introduced both a new set of electromagnetic field variables (mixing electric and magnetic field variables) and a new time variable. These new variables carried no explanatory weight but just served to simplify otherwise complicated expressions. In experiments dealing with intensities of electromagnetic radiation, however, one could treat these new variables as if they described the same physical quantities as the old ones. Indeed, there had been derivations of the Fresnel coefficient in a purely wave optical context, which had relied solely on the introduction of a new time variable equivalent to the Lorentzian auxiliary time. In these earlier derivations, however, the introduction of a new time variable had to be included as an explicit assumption rather than being derived from the invariance properties of dynamical equations. While the introduction of a new time variable might appear to be more immediate along purely optical arguments, its appearance in an electrodynamic context was more seamless as it did not require an explicit interpretation of the new variable.

[12] Lorentz 1892.
[13] Lorentz 1895.

Taking account of the null results of the Michelson-Morley experiment turned out to present a challenge for Lorentz's approach, because it represents an experiment that, in principle, could have detected the motion with respect to the stationary ether to second order in $v/c$. While Lorentz's theory suggested an effective equivalence of stationary and fully dragged ether to first order, it also ascribed real physical existence to the stationary ether, whose state of rest should be revealed in second order experiments. Whereas for first order effects the explanation could rely on electrodynamic mechanisms, now an additional hypothesis had to be introduced, whose meaning in terms of electromagnetic interactions was less evident, namely the hypothesis that bodies are contracted in the direction of their motion through the ether.[14]

The last two paragraphs may appear unnecessarily technical in the context of an account of the emergence of the relativistic concepts of space and time, but they reveal an essential mechanism at work in this process. With his auxiliary time and his contraction hypothesis, Lorentz had effectively constructed a formal framework for new time and space variables. However, Lorentz's techniques were neither derived nor presented in this manner, rather appearing as natural outgrowths of, or at least plausible assumptions within, a complex and phenomenologically rich dynamical theory, which in turn stabilized and at the same time constrained these innovations. Given the foreign character of these new spacetime variables, such a stabilization was indeed an important condition for integrating them into the larger body of physical knowledge that could not be achieved with equal ease by merely postulating them. Similarly, the constraints imposed by the underlying dynamical theory provided the new variables with a persuasive uniqueness not achievable by mere speculation.

This secured the constructibility of the new space and time variables. The price for this achievement was, however, that also the physical interpretation of the new variables was highly constrained by the framework in which they were embedded, concealing the possibility of implementing them as new concepts of space and time with their own operational meaning and a universal domain of reference.

Returning to the questions raised in the introduction, it may be asked what qualified the new space and time variables to serve as defining a new space-time framework. Lorentz had established the first of our three criteria, viz. constructibility. He had pushed the means of classical electrodynamics to the point of allowing for a consistent and complete description of space and time structures distinct from the pre-existing ones. But these new theoretical constructs were still devoid of any spatio-temporal meaning, because their relation to prior knowledge of space and time was only established indirectly through the complex framework of electrodynamics. The establishment of a more direct relation required a rethinking of space and time measurement that lay outside the scope of Lorentzian electrodynamics.

From a broader perspective on the conceptual foundations of physics such a rethinking was natural and, more importantly, possible, because the constructibility of new space and time variables had already been established. The ultimate success of such rethinking depended, however, on further conditions. It had to be checked whether such newly constructed space and time concepts could be related to prior knowledge of space and time and whether they fulfilled the criterion of universality. From such a broader perspective, both Poincaré and Einstein identified Lorentz's auxiliary time variable as giving the time in a moving system as actually measured. Einstein furthermore succinctly captured the core of

[14]Lorentz 1895.

the prior knowledge of space and time in terms of generic measurement procedures, showing how this procedure could be made compatible with the identification of Lorentzian space and time variables as giving the actually measured quantities, and argued for the universality of the resulting framework. One of the key arguments was the compatibility of the new transformation laws with all results from classical physics in the limiting case of velocities small compared to the speed of light. The universality of the framework has since been corroborated by all available physical evidence, with only gravity necessitating a further modification of this framework.

While special relativity as formulated by Einstein brought about new notions of space *and* time, it was only Minkowski's reformulation in terms of a four-dimensional formalism that unified the two into a single structure, spacetime. This single structure consists in a four-dimensional geometry with a pseudo-Euclidean metric. Its mathematical features immediately reflect the laws of special-relativistic physics, thereby reducing the elaborately constructed Lorentz transformations to mere rotations within this geometry. One far-reaching implication of this framework is a new understanding of energy-momentum conservation in terms of four-dimensional quantities. Another implication is the emergence of concepts that are distinctly spatiotemporal in character, such as the concept of proper time, which integrates spatial aspects (path dependency) and temporal aspects (transience) into an inseparable whole.

## 7.4 The re-inclusion of gravitation into the concepts of space and time.

As we have seen, in Newtonian physics, gravitation was excluded from discussions of the structure of space and time because it was considered to be a force among others *within* space and time. *Prima facie* it was to be expected that this would not change in special relativity. In Newtonian physics, the exclusion of gravitation was the premise for the postulation of homogeneous space and time endowed with an inertial structure, allowing in turn the re-introduction of gravitation, now as a force. The universality of gravitation could be accounted for by means of the additional assumption that gravitational mass is proportional to inertial mass. The universality of gravitation was thus attributed to a mere coincidence. Special relativity introduced, as we have seen, a completely new conception of spacetime. It was therefore entirely open whether the complex reconciliation of gravitation and spacetime structure could be reproduced under these conditions. Although gravitation was the most familiar of all physical forces, in this situation it was not dissimilar to a newly discovered phenomenon. Knowledge about this phenomenon was stored in the Newtonian theory of gravitation. The question was therefore which aspects of this theory were to survive in the new framework because they reflected indispensable empirical knowledge, and which had to be discarded because they just corresponded to features of a conceptual system that had to be overcome. Since these aspects were closely entangled, an answer to this question could only be found by exploring various alternatives.

These alternatives had distinct implications for the understanding of space and time. It was, for instance proposed by Poincaré and Minkowski to replace Newton's force law by a retarded action-at-a-distance law complying with the principle that no physical interaction propagates faster than light.[15] But while such a law would not require any further modifi-

[15] Poincaré 1906 and the Appendix of Minkowski 1908. English translations and discussions of both texts are found in Renn 2007a, Vol. 3.

cation of the special-relativistic understanding of space and time, it does raise fundamental questions about energy and momentum conservation. On the other hand, given the role of electromagnetism for special relativity it was plausible to build a new theory of gravitation according to this model of field theory. The first to try out such a theory was Max Abraham, but without success, since it turned out that his theory not only transgressed the bounds of the special-relativistic framework but was actually inconsistent.[16]

In order to remain within this framework, further-going modifications of classical concepts such as that of mass were required, as was realized by Gunnar Nordström and Einstein. Nordström attempted to construct a truly special-relativistic field theory of gravitation using a scalar source term, but soon encountered difficulties arising from the fact that in special relativity energy-momentum is represented by a second-rank tensor.[17] These difficulties forced him to modify his theory in such a way that it effectively left the special-relativistic spacetime framework, because coordinate differences no longer correspond directly to measured differences, their relation being determined by a dynamic factor. Therefore, the spacetime concepts of special relativity lose their operationability, a fact that may call into doubt the validity of these concepts. At the same time, the theory showed the constructibility of an alternative spacetime concept, based on the reinterpretation of the operationally defined distance measurements as providing the relevant spacetime framework. Given that this development emerged from a theory of gravity, the universal character of gravity automatically carries over to the new spacetime concepts, so that all our three criteria for the establishment of a new spacetime framework are fulfilled.

The problematic criterion in this case turns out to be the constructibility. First, the realizability of reinterpreting measured differences as the geometrically relevant distances hinges on the availability of a mathematical formalism in which coordinate differences and physical distances are conceptually divorced and can be systematically related to each other. And even given the realization that such a framework was available in the non-Euclidean geometry of Gauss and Riemann, this mathematical apparatus was rather complex given the comparatively simple deviations from pseudo-Euclidean geometry in the Nordström theory. This complexity and conceptual overhead might therefore have seemed too high a price to pay for leaving the familiar ground of special relativity.[18]

With the attempts of Abraham and Nordström, the potential of electromagnetism as a model for a relativistic field theory of gravitation was, however, not exhausted, in particular as approaches using a scalar field were of reduced structural richness compared to the electromagnetic model. It was particularly plausible to look for a generalization of the gravitational field analogous to the generalization of the electrostatic to the electrodynamic field by the introduction of additional degrees of freedom, viz. those of the magnetic field. In the framework of special relativity, this latter generalization could be seen as a consequence of the relativity principle. In fact, for Einstein himself, the existence of two equivalent descriptions involving different splits of the electromagnetic field into its electric and magnetic

[16]Abraham 1912b; Abraham 1912a. English translations of these texts and further references are found in Renn 2007a, Vol. 3; see, in particular, the discussion in Renn 2007b. See also Renn and Schemmel 2012.

[17]Nordström 1912; Nordström 1913a; Nordström 1913b. English translations of these texts are found in Renn 2007a, Vol. 3. For a discussion of Nordström's theories see, in particular, Norton 1992 (reprinted as Norton 2007).

[18]As late as 1917, Max von Laue used the conceptual unfamiliarity of general relativity to argue for Nordström's theory and its special-relativistic framework (Laue 1917).

components depending on the state of motion was an important hint used in the formulation of special relativity.

It was therefore plausible to turn this argument around and postulate the existence of a generalized gravitational field manifesting itself in different forms depending on the state of motion. In order to fill this scheme with actual physical content it was, however, required to find an adequate complement to the familiar static gravitational field. Such a complement was suggested by the somewhat artificial distinction between uniform and accelerated motion in Newtonian mechanics. By questioning this distinction one could introduce, at the same time, the idea of unifying gravitational and inertial effects and generalizing the principle of relativity to include accelerated motions. This unification was possible (strong equivalence principle) as the physical parameters determining the strength of both kinds of effects, gravitational and inertial mass, respectively, were proportional (weak equivalence principle).[19]

The distinction between inertial and non-inertial motion was prominently challenged by Ernst Mach who suggested an interaction between masses to account for the occurrence of inertial forces. Mach's critique thus provided the missing ingredient for Einstein's attempt to construct a dynamical theory of gravitation based on the idea of a generalized gravitational field. As a consequence, inertia played the role that magnetism played in the electromagnetic case, thereby providing an untapped experiential resource for the formulation of a new theory of gravitation. But since in Newtonian physics inertia was a structure of space and time, this specific implementation of the electromagnetic model implied possible consequences for spacetime. In particular, it quickly turned out that the equation of motion in the combined inertio-gravitational field had a geometrical interpretation, namely that of describing a geodesic in curved spacetime.

This means that one and the same mathematical structure, the metric field, serves to represent the inertio-gravitational field and the geometrical properties of spacetime. Accordingly, a field equation for this field, as it was part of the overall scheme associated with the electromagnetic model, assumed the meaning of an equation determining the dynamical spacetime geometry. The implications of this dynamics surpassed by far the horizon of expectations based on the experiences with prior field theory. As it turned out, it could no longer be forced into any of the available pictures about the relation between space and time on the one hand and matter on the other. Rather, these notions became entangled in new ways.

The question as to which experiences led to the establishment of new space and time concepts in general relativity can thus be answered by pointing to the process of integration of special relativity, Newtonian gravitation, and the inertial structure of classical mechanics as we have described it. But we have also raised the question of why these specific experiences had this consequence and why at this particular historical moment. It is here that we come back to our three criteria, constructibility, operationability, and universality. It is actually in this sequence that these criteria became relevant in the history of special relativity.

Remarkably, their sequence was different in the history of general relativity. Using his equivalence principle to refer concepts of space and time of special relativity to gravitational effects simulated by accelerated motion, Einstein could immediately transfer their

[19] See, e.g., Norton 1989.

operational content to the theory under construction. That this transfer would also not strip them of their universal nature was clear from the domain they were being transferred to, i.e. the equally universal phenomena of gravitational and inertial forces. In contrast, e.g., to Nordström, Einstein's modeling of gravitation with the help of accelerated motion directly brought him to explore his theory in terms of new space and time concepts. This explains the inversion of the historical sequence in which the criteria became relevant. From the outset, Einstein searched for a new spacetime framework. To a certain extent he could explore this framework with the guidance of his equivalence principle by considering various cases of accelerated motion. All such cases should qualify as legitimate manifestations of the generalized gravitational field of the theory to be constructed. The accumulation of insights following this heuristics did, however, not amount to the construction of the desired theory.

Constructibility was afforded by the introduction of a new mathematical representation from differential geometry which allowed conceptualization of the intricate relation between measured quantities and coordinate differences, suggested by the heuristic generalization of the Minkowski line element. The specific means available to the historical actors, e.g. metric geometry as opposed to affine geometry, shaped the space of constructible formulations. The criterion of constructibility can, however, not be fulfilled by solely establishing a mathematical representation of the new spacetime concepts. First of all, as the new spacetime concept had be introduced as an element of a new theory of gravitation, a complete construction requires the establishment of an adequate field equation reproducing what is empirically known about gravity. Second, the construction must not invalidate what has already been heuristically achieved in the domain of operationability. This excludes, in particular, any restriction on the admissability of accelerated motion as being interpreted as a generalized gravitational field, following, e.g., from the field equation.

The new spacetime concept brought about by general relativity is characterized by the fact that it locally corresponds to the spacetime of special relativity and that its dynamics is that of the inertio-gravitational field. This is in stark contrast to theories in which there remain non-dynamical aspects of space and time beyond the local correspondence to special relativity, such as the survival of real inertial forces in Nordström's theory. The background independence of general relativity hinges on the complete inclusion of inertial effects into the dynamic field determining spacetime. As strong as this criterion is, it does not exclude variations of general relativity, e.g., by the introduction of a cosmological constant.

## 7.5 The role of space and time in quantum theory

In quantum mechanics space and time continue to play a role and are potentially affected by its conceptual framework. The concept of position in space is related in ordinary quantum mechanics to the position operator. As a consequence, the position and the momentum of a particle can no longer simultaneously be determined, even in principle, with absolute precision, as it is the case in classical physics. Nevertheless, neither the concept of position nor its operational definition in terms of classical measurements are affected by this limitation.

It is therefore no surprise that in quantum field theory space and time can be treated in full analogy to classical field theory as parameters of the field. In essence, therefore, quantum field theory makes use of the spacetime framework of special relativity. This framework is only challenged by quantum theoretical considerations when spacetime itself is conceived as a dynamical field, as happens in general relativity. Another major difference between classical and quantum field theory is that the latter provides a successful framework also for a field theoretical description of many-particle dynamics. Hence, while in classical field theory the space-time coordinates serve a double role as dynamical particle positions and arguments of continuous field functions, only the latter role remains important in quantum field theory.

Historically the field concept has raised problems in the framework of quantum theory. These problems have motivated alternative conceptualizations that may be examined as to their potential implications of the concepts of space and time. The most radical of these attempted to get rid of the field concept altogether, as in Wheelers heuristic research program "everything as electrons." On the background of a conceptualization of space as the "positional quality of the world of material objects,"[20] this program might have led to a fundamental revision of the concepts of space and time by quantum theory. Wheeler's thought involved the identification of all electrons and positrons as a single particle whose world line went back and forth in time, thus cutting our world sheet many times. From this perspective, the indistinguishability of identical particles can be interpreted as a non-trivial connection of distant points of space.

Indeed, if only trajectories of particles are considered to be real, one may even speak of an identification of points in space. Had such a program been successful, it might have brought about new concepts of space and time. Whether the historical lack of success of this program reflects a principal obstacle or is merely a result of the limited means available, this example illustrates the crucial role of the actual *constructibility* of new concepts of space and time. In short, it turned out to be historically impossible to rebuild quantum physics as a more satisfying theory of matter and radiation by reformulating it within a new space-time framework.

One might think of representing the history of physics as a sequence of ever more general concepts of space and time. It may, however, turn out to be misguided to expect that space and time will always maintain their fundamental status through all profound changes. Instead they may lose this status to other concepts which may lack the operational foundation in prior knowledge characteristic not only for space and time but also for other canonical fundamental concepts such as matter, motion, force, and causality. This process of marginalization of canonical fundamental concepts may again be illustrated by means of the history of quantum physics.

A purely formal marginalization of space already occurs in analytical mechanics, which offers the option of focusing on an abstract space constituted by possible states of a physical system (state space, phase space, configuration space). States are typically characterized not just by spatial parameters but also by their temporal derivatives. Therefore the properties of state space are different from those of 'ordinary' physical space. Nevertheless, in classical physics, ordinary physical space remains fundamental, while state spaces may be considered as auxiliary constructions, since physical states can always be projected onto ordinary physical space as a distribution of matter along with the initial conditions, without loss of information about the dynamics.

In quantum mechanics, when entanglement enters, this projection is no longer possible. For instance, the wave function of a two-particle system cannot be represented as a probability distribution over ordinary space due to correlations between position measurements

[20] Albert Einstein in his foreword to Jammer 1954, xiv.

of the two particles. Therefore, in quantum mechanics ordinary space loses its ontological role since the state of the world can no longer be represented by a function over universal space and time coordinates.[21] Even when physical processes in quantum mechanics are interpreted as definite processes in space and time (path integral formulation), their interference does not take place in this spacetime, but only in configuration space. In this sense, ordinary spacetime is marginalized by the Hilbert space structure of quantum mechanics.

But Hilbert space cannot be conceived as a profoundly changed fundamental concept of physical space. It does, in particular, not replace the concept of space of classical or relativistic physics, but rather incorporates these concepts in a somewhat marginalized form. This would only change if it were possible to apply quantum physical considerations to the dynamics of spacetime itself. As long as there is no viable quantum theory of gravity and cosmology, the constructibility of the new spacetime concepts it might introduce is not given.

## 7.6 Conclusion

We have started our chapter with two questions. First, which experiences led to the establishment of new space and time concepts in the history of modern physics? And second, why did these specific experiences have such consequences and why at a particular historical moment? In the previous sections we have tried to answer the first question by reviewing the historical development of space and time concepts from the perspective of the experiences that have given rise to them. Clearly, however, an equally important component of the development were the conceptual and formal tools allowing the formulation of these concepts. What shaped the dynamics of their development? Before we come to propose answers to the second question, we would therefore like to make a few remarks on the general dynamics of conceptual frameworks.

Concepts of space change only in the context of entire theories. These are not elements of an abstract set of theories but always develop historically out of pre-existing knowledge systems. These systems comprise the available means for addressing the perceived problems and thereby define the space of possible solutions and further developments. The perceived problems possibly comprise new experiences, which have to be integrated with preprocessed experiences already incorporated in the knowledge system. The system character of knowledge has implications for the long-term development of the means of representation. These will only be elaborated and transmitted if they serve some function within a knowledge system, in particular as means for solving relevant problems. As a consequence, new means for articulating concepts of space and time will typically only emerge from the historical development of such larger systems, for instance comprehensive physical theories. Therefore, in addition to the three criteria mentioned in the introduction as conditions for the emergence of new concepts of space and time, i.e., constructibility, operationability, and universality, we actually have a fourth criterion, viability, requiring that a proposal for new space and time concepts is part of a theoretical framework that successfully addresses the relevant problems.

The criterion of constructibility is therefore not defined by the question of whether the necessary tools may have been in principle available to the historical actors, as if they were

[21] Loss of projectibility is also given in statistical mechanics, of course, but does not imply a change in ontology and only reflects our state of knowledge.

part of a universal tool box, but if the historical development of some available knowledge system could possibly have brought them about. Constructibility, in short, is defined by previous historical processes of construction and thus highly path-dependent. The historical sequence of the construction of knowledge systems involves an iterative procedure of representation and reflection. Representation is here understood in the broad sense of a set of external, i.e., material, representations of a knowledge system, such as its description in terms of language, symbolic formalisms, or artifacts. Reflection is equally broadly understood as the set of thinking processes accompanying the implementation and exploration of such a knowledge system with the help of these representations that may lead to the construction of new knowledge structures, which may then be characterized as knowledge structures of higher-order.

Accordingly, means for solving problems within knowledge systems may be distinguished by their degree of reflexivity, indicating the specific sequence of representation and reflection that gave rise to them. In the case of Euclidean geometry, for instance, the figures that can be drawn with compass and rulers may be considered as first-order representations, while its linguistic formulation within a deductive structure constitutes a second-order representation.[22] The invention of non-Euclidean geometry presupposed a degree of reflexivity that allowed to consider such second-order representations and their possible alternatives in turn as an object of reflection that may constitute the meaning of geometry independently from its first-order representations. In summary, means of construction have their own history that may be more or less closely related to the history of the subject matter under consideration.

The match between mathematical formalisms and the physical world has often been discussed as a puzzling fact, because of the difficulty to explain the adequacy of the mental constructions of mathematics for the description of physical experiences. However, when those mental constructions are understood as the result of long chains of sequences of reflections and representations, which at each stage involve specific experiences, the mathematical formalisms themselves turn out to be saturated with experience. This is, of course, not necessarily the same kind of experience that is to be captured by some physical theory. It thus may seem that, along this line of thinking, the puzzle can be reduced to the question of how to integrate different domains of experiences. This integration is, however, made even more difficult by the fact that the experiences underlying a given formalism are only implicitly represented by it, since the formalism is usually the result of a long chain of reflections and representations.

The codification of experience in terms of knowledge structures is indeed one of the reasons for the characteristic *recursive blindness*[23] of abstract thinking with regard to its own experiential sources, a recursive blindness that also accounts for the seemingly a priori character associated with the concepts of space and time. Bringing together different domains of experience by matching physical experiences with mathematical formalisms, therefore, typically raises the question which aspects of a formalism represent experiences and which have to be considered either as merely formal aspects or as representing experiences in need of reinterpretation. For instance, in the case of the electrodynamics of moving bodies, the challenge was to integrate the experiences of electromagnetism codified in Maxwell's equations with the experiences of mechanics codified in the Galilean transformations. The necessary

[22] See Damerow 1994, 268–270 and Schemmel 2016, 47–50.
[23] Renn and Hyman 2012, 493.

adaptation of the latter formalism raised the question of what aspects of it were related to an operational understanding of space and time measures and what aspects represented experiences in need of reinterpretation, such as that of simultaneity.

To sum up, it is not the case that the factor that systematically varies in the historical development of physical theories is primarily the ever larger extent of experience described by them, while the availability of adequate mathematical formalisms enters as a contingent factor or one that is governed by an entirely different logic. Rather, the development of formalisms itself involves the processing of experiences and is often closely related to the development of physical theories in the sense of our fourth criterion of viability. Therefore, from a larger perspective, there is a co-evolution of physical theories and the formalisms they employ to cover the experiences they strive to explain: while they may belong to separate intellectual or disciplinary traditions, they are typically still part of the shared knowledge available to society at large.

This co-evolution also accounts for the global dominance, despite the persistent emergence of locally viable alternative solutions, of a single stream of development in the sense of 'the winner takes all'. As in evolutionary theory, optimization is a local phenomenon always working with the available means, rather than within an abstract set of theories. In this process of optimization, any established solution (e.g. Newtonian classical mechanics) – 'established' both in an intellectual and institutional sense – is typically stabilized and extended by assimilating a maximal range of experiences, thereby gaining an advantage with respect to conceivable alternatives (e.g. Leibnizian mechanics) that are not granted a similar chance of being implemented.

It is this local dynamics that accounts for an overall development that, at the level of a history of ideas decoupled from their embodiment in the material means and concrete experiences, may seem to display a rather astonishing movement back and forth among fundamental notions of space. In particular, there is, as we have seen, the dissociation of space from gravity in Newtonian physics, by which Newtonian space becomes homogenous and isotropic, whereas in general relativity gravitation becomes a feature of spacetime. General relativity thus returns to the notion that space (or spacetime) guides motions under the influence of gravitation, a notion closer to Aristotelian physics than to Newtonian. One may therefore ask if there could not have been a more direct pathway connecting the Aristotelian notion of anisotropic space to that of Einstein. Here we have argued that the apparent swaying motion of this long-term development cannot be understood at the level of ideas of space alone. As we have also argued, such ideas only inherit their viability from the broader theories they are part of. As a matter of fact, in modern science it is the integration of an increasing corpus of experience by means of formalisms that defines progress. Therefore, on a global level, the historical development is actually much more constrained than a history of abstract ideas can account for.

This we can see more clearly, when we take into account that there is another fundamental reason for the streamlining of global developments, in addition to the winner-takes-all logic described above explaining the extrusion of alternatives. Alternatives themselves typically only emerge in the process of exploring the available means (e.g. as special relativity developed out of Lorentzian electrodynamics) and are the more viable, the richer their experiential basis, which in turn is largely provided by the established solution. The elaboration of alternatives to such a given solution (with the help of the means it provides) may either result in its abandonment (e.g. when special relativity emerged from searching alternatives to Lorentzian ether theory) or in its reconceptualization in new and different terms, for instance with the help of a new formalism (e.g. Minkowski's formalism being the result of a reconceptualization of special relativity). Such reformulations have a double function: they serve to assimilate already existing but not yet integrated experiences to a theory, thus rooting it even more deeply in experience (e.g. assimilating optics to electrodynamics by reformulating the laws of electromagnetism in terms of Maxwell's equations), and they may become the starting point for the integration of novel experiences and, possibly, the eventual overcoming of the established theory (e.g. when special relativity is reformulated in terms of Minkowski's formalism enabling the development of general relativity).

From this perspective, let us therefore once more review our account of the historical evolution of the concepts of space and time in physics with particular attention to the viability of alternative trajectories. For a long time, gravitation was a natural component of conceptions of space. What was later identified as the influence of gravitation is primarily motion taking place spontaneously and in a specific direction. This spontaneity and directedness was indeed an aspect of natural motion in Aristotelian physics, in which sublunar natural motion was directed towards the center of the Universe (which coincided with the center of the Earth) or away from it, while celestial motion was circular around the center.

Historically, alternatives to the geocentric world view were formulated in ancient Greece. The question thus arises whether the fact that a specific Earth-bound perspective was elaborated into a comprehensive system of knowledge should be deemed contingent. There may have been systematic reasons for the dominance of the geocentric world view, for instance the possibility to incorporate insights of terrestrial physics (such as the doctrine of the elements) and anthropocentric ideologies (such as Christian religion), but in view of the existence of alternatives, the dominance of geocentrism appears to be at least partially contingent. After all, the decision was based on a relatively small empirical basis, but once it was taken and an increasingly large corpus of knowledge was assimilated to it, the alternatives appeared ever less plausible.

On the other hand, the connectivity of the system developed over the centuries and the large amount of knowledge incorporated gave the addition of new insights a potentially large impact on the system as a whole. Thus, when astronomical developments eventually favored a view in which the Earth was no longer at the center of the universe, not only could the question arise whether the fall of bodies was just an Earth-bound phenomenon and not a manifestation of natural motion in a cosmological context, but the very concept of natural motions was called into question. Hence what had earlier been considered natural motions was in need of an alternative explanation. As is well known, Newton's solution was to postulate a universal force of gravitation that explained the formerly natural motion of free fall and the celestial motions. Thereby he integrated the great amount of experience pre-processed in terrestrial and celestial mechanics on the basis of new explanatory models, such as that of force as suggested by magnetism.

From terrestrial physics, and the analysis of projectile motion in particular, the idea of an inertial motion emerged, which was to become seminal for the new concept of space. Inertial motion demanded an absolute standard to judge motion, which was achieved by the concept of absolute space. Inverting the Aristotelian categorization of motions, inertial motion thus became the sublimated version of natural motion, while the motion of fall came to be interpreted as forced.

Why did the new concept of gravitation not become part of the new concept of space? Besides the question of the availability of the mathematical means of construction, a conceptualization of gravitation in terms of a four-dimensional spacetime would have rendered impossible the very formulation of Newtonian mechanics, in which gravitation served as the paradigmatic model of a force. Therefore its inclusion in a concept of spacetime was no viable option for Newton. One might well imagine that a few decades later, after Newtonian gravitation had served its historical role of integrating celestial motion and terrestrial mechanics, alternative formulations of gravitation in terms of spacetime structures would have become possible. In this context one may think of the representation of gravitation and inertia in terms of a projective or an affine structure on spacetime already within the context of Newtonian physics, as presented by Newton-Cartan theory.[24]

All such attempts fall short, however, of constructing dynamical field equations. It therefore does not seem to be coincidental that the invention of general relativity occurred soon after the formulation of Lorentzian electrodynamics, which provided the model for a field theory of gravitation and inertia. Nevertheless, the successful formulation of gravitation in terms of a dynamic spacetime within general relativity today faces a similar problem as the one it would have faced in Newton's times: it separates gravitation from the other forces. So far all attempts to geometrize the other forces have failed; as well as all attempts to integrate gravitation – despite its special status – into a quantum field theoretical framework.

## Bibliography

Abraham, Max (1912a). Der freie Fall. *Physikalische Zeitschrift* 13:310–311.

— (1912b). Zur Theorie der Gravitation. *Physikalische Zeitschrift* 13:1–4.

Ben-Menahem, Yemima (2006). *Conventionalism*. Cambridge: Cambridge University Press.

Damerow, Peter (1994). Vorüberlegungen zu einer historischen Epistemologie der Zahlbegriffsentwicklung. In: *Der Prozeß der Geistesgeschichte. Studien zur ontogenetischen und historischen Entwicklung des Geistes*. Ed. by Günter Dux and Ulrich Wenzel. Frankfurt a.M.: Suhrkamp, 248–322.

Earman, John (1989). *World Enough and Space-Time: Absolute versus relational theories of space and time*. Cambridge, MA: MIT Press.

Ehlers, Jürgen (1973). The Nature and Structure of Spacetime. In: *The Physicist's Conception of Nature*. Ed. by Jagdish Mehra. Dordrecht: Reidel.

Einstein, Albert (1921). *Geometrie und Erfahrung*. Berlin: Springer.

— (2001). *The Collected Papers of Albert Einstein, Vol. 7. English translation of selected texts*. Princeton: Princeton University Press.

Feynman, Richard Phillips, Robert B. Leighton, and Matthew L. Sands (1989). *The Feynman lectures on physics, Vol. 2*. Redwood City: Addison-Wesley.

Jammer, Max (1954). *Concepts of Space: The History of Theories of Space in Physics*. Cambridge, MA: Harvard University Press.

Janssen, Michel and John Stachel (2004). *The Optics and Electrodynamics of Moving Bodies*. Preprint 265, Max Planck Institute for the History of Science, Berlin.

[24]For an introduction to that formulation, see Misner, Thorne, and Wheeler 1973, 289–303.

Laue, Max von (1917). Die Nordströmsche Gravitationstheorie. *Jahrbuch der Radioaktivität und Elektronik* 14:263–313.

Lorentz, Hendrik Antoon (1892). La théorie électromagnétique de Maxwell et son application aux corps mouvants. *Archives néerlandaises des sciences exactes et naturelles* 25: 363–552.

— (1895). *Versuch einer Theorie der electrischen und optischen Erscheinungen in bewegten Körpern*. Leiden: Brill.

Minkowski, Hermann (1908). Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern. *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen*:53–111.

Misner, Charles W., Kip S. Thorne, and John A. Wheeler (1973). *Gravitation*. New York: Freeman.

Nordström, Gunnar (1912). Relativitätsprinzip und Gravitation. *Physikalische Zeitschrift* 13:1126–1129.

— (1913a). Träge und schwere Masse in der Relativitätsmechanik. *Annalen der Physik* 40:856–878.

— (1913b). Zur Theorie der Gravitation vom Standpunkt des Relativitätsprinzips. *Annalen der Physik* 42:533–554.

Norton, John D. (1989). What Was Einstein's Principle of Equivalence? In: *Einstein and the History of General Relativity*. Ed. by Don Howard and John Stachel. Boston: Birkhäuser, 5–47.

— (1992). Einstein, Nordström and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation. *Archive for History of Exact Sciences* 45:17–94.

— (2007). Einstein, Nordström and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation. In: *Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy*. Ed. by Jürgen Renn and Matthias Schemmel. Dordrecht: Springer, 413–487.

Pauli, Wolfgang (1979). *Scientific Correspondence with Bohr, Einstein, Heisenberg a.o., Vol. 1: 1919-1929*. Ed. by Armin Hermann, Karl von Meyenn, and Victor Frederick Weisskopf. New York: Springer.

Pfister, Herbert (2004). Newton's First Law Revisited. *Foundations of Physics Letters* 17(1): 49–64.

Piaget, Jean (1959). *The Construction of Reality in the Child*. 5th print. The Basic Classics in Psychology. New York: Basic Books.

Poincaré, Henri (1906). Sur la dynamique de l'électron. *Rendiconti del Circolo Matematico di Palermo* 21:129–175.

— (1952). *Science and Hypothesis*. New York: Dover.

Renn, Jürgen, ed. (2007a). *The Genesis of General Relativity*. 4 Vols. Boston Studies in the Philosophy of Science 250. Dordrecht: Springer.

— (2007b). The Summit Almost Scaled: Max Abraham as a Pioneer of a Relativistic Theory of Gravitation. In: *Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy*. Ed. by Jürgen Renn and Matthias Schemmel. Dordrecht: Springer, 305–330.

Renn, Jürgen and Malcolm Hyman (2012). Survey: The Globalization of Modern Science. In: *The Globalization of Knowledge in History*. Ed. by Jürgen Renn. Berlin: Edition Open Access, 491–526.

Renn, Jürgen and Matthias Schemmel (2012). Theories of Gravitation in the Twilight of Classical Physics. In: *Einstein and the Changing Worldviews of Physics*. Ed. by Christoph Lehner, Jürgen Renn, and Matthias Schemmel. Einstein Studies 12. Boston: Birkhäuser.

Schemmel, Matthias (2014). Medieval Representations of Change and Their Early Modern Application. *Foundations of Science* 19:11–34.

— (2016). *Historical Epsitemology of Space: From Primate Cognition to Spacetime Physics*. Cham: Springer.

# Index


**A**

Abraham, Max, 26, 202
absolute
directions, 12, 41, 76
forces, 186
impenetrability, 177–179, 184–186
lack of extension, 142
motion, 198
space, 4, 21, 184, 188, 196, 209
standard of rest, 196, 198, 209
accommodation, 9
action at a distance, 201
additivity of lengths, 123
Akkadian, 112
al-Bitruji, 146
Alexandria, Museion of, 18
Alpetragius (al-Bitruji), 146
Amico, Giovan Battista, 146, 170
Anaximander, 150, 156, 167
angle, 10, 16, 17, 94, 101, 107, 115–117
anthropocentrism, 171, 172, 209
anthropogenesis, 3
Arab Middle Ages, 18, 124
Aristotelian cosmology and physics, 18, 22, 23, 123, 124, 139–141, 143, 146–153, 166, 168–173, 179, 195, 196, 208, 209
Aristotelian-Ptolemaic cosmology, 146, 172
Aristotelianism, 175, 178
Aristotle, 139, 145, 146, 153, 156, 169, 175, 176
his distinction of three speculative disciplines, 171
his theory of change, 196
'Categories', 140
'De caelo', 145–153, 167–169, 172, 173, 177
'Metaphysics', 145
'Physics', 123, 139–141
assimilation, 9, 193, 194
astrophysics, 21
Athapaskan, 46, 52, 54, 81
atomism, 19, 21, 23, 124, 139, 149, 176–186, 188, 191, 192, 199
Averroes (Ibn Rushd), 145, 170, 172
Aztec surveyors, 17

**B**

Bellarmin, Robert, 170
Benedetti, Giovan Battista, 149
Boas, Franz, 37, 66, 85
Bororo (Brasilian central plateau), 14
Bruno, Giordano, 170, 172, 179–180

**C**

Carnap, Rudolf, 1
Caroline Islands navigators, 13
causal explanation, 121, 145, 147, 170, 196, 197
China, 15, 58, 124, 125, 139, 142, 195
Clarke, Samuel, 22
cognitive mapping, 6–11, 35, 37–40
compass
and straight edge, 15, 18, 19, 132, 133, 142, 207
magnetic, 12, 22
Micronesian star compass, 13
conatus, 183, 184
conceptual model, *see* mental model
conservation
of energy and momentum, 26, 201, 202

of quantity of matter, 188
of size and shape, 16
container model of space, 23, 24
continuity of object trajectories, 7
conventionalism, 145, 192, 193
coordinates, 22
geographical, 22
linguistic, 39, 40, 82
their role in field theory, 205
their role in quantum mechanics, 206
Copernican
revolution, 146, 172
system, 172, 196
tables, 147, 170
Copernicus, Nicolaus, 152
'De revolutionibus', 146, 169–171
cosmological constant, 204
cosmology
Aristotelian, *see* Aristotelian cosmology and physics
Aristotelian-Ptolemaic, *see* Aristotelian-Ptolemaic cosmology
as a theory of the world as a whole, 146
Chinese, 143
classical, 145, 172
Greek, 19, 143, 172
its relation to physics, 169–171
modern, 21, 24
of the Pythagoreans, 148, 150–152
Plato's, 148
Ptolemaic, *see* Ptolemaic cosmology
quantum, 206

**D**

dead reckoning, 22, 39, *see also* path integration
Democritus, 139, 176, 177, 179, 186
Dene Chipewyan (Cold Lake, Alberta, Canada), 12, 36, 43, 45–54, 65, 66, 79–85
dependency of the effort on the path taken, 6, 8, 10
Descartes, René, 139, 181–183, 185, 186
'Principles of Philosophy', 1, 139, 181, 182, 185
'The World', 182
detour, 6
diagrams, 27, 123, 142, 196
dichotomy of movable and unmovable objects, 7, 8
dichotomy of object and space, 7, 8, 22, 127
disciplinary structure of knowledge, 18, 23–27, 145–147, 171, 208
distance, 10, 13, 14, 22, 37–39, 42, 64, 95, 99, 122, 123, 137, 193, 202
between stars (angular), 153–155, 157–159, 163, 166
Euclidean, 19, 197
in Dene Chipewyan, 83
in Eipo, 72–74, 76–78
in Hopi, 81
division
arithmetic, 16, 110, 112, 114
of labor, 15, 51
of land, 14, 116
physical, 176, 177
drawing, 4, 11, 14–16, 27, 108, 109, 115, 121, 133
Duhem, Pierre, 145, 170

**E**

Egypt, 15, 17, 18, 115, 123, 195
Einstein field equation, 21, 203
Einstein, Albert, 23, 26, 188, 191, 192, 200–205
Eipo (central highlands of West New Guinea), 12–14, 36, 42, 43, 50, 55–78, 84, 85
electrodynamics, 21, 22, 25, 191, 197–203, 207–210
quantum, *see* quantum electrodynamics
electromagnetic
field, 199, 202

radiation, 21, 22, 198, 199
worldview, 22
empty, *see* void
equation of motion in the inertio-gravitational field, 203
equivalence principle, 26, 191, 197, 203, 204
ether, 160, 169, 179, 180, 189, 198–200, 209
Euclid
'Elements', 123, 142
'Phaenomena', 164
Euclidean
distance, *see* distance, Euclidean
geometry, *see* geometry, Euclidean
space, 1, 8, 16, 193
Euler, Leonhard, 185–186
evolutionary epistemology, 2
Ewe, south-eastern Ghana, 42

**F**

field
concept, 22, 188, 205
electromagnetic, *see* electromagnetic field
equation, 204, 210
inertio-gravitational, *see* inertio-gravitational field
field theory
of gravitation, 202–203, 210
quantum, *see* quantum field theory
figure-ground asymmetry, 39, 41, 43–45, 52, 84
Fizeau, Hippolyte (experiment measuring the index of refraction in moving media), 198
Fracastoro, Girolamo, 146, 170
frames of reference, 37, 39–41, 76, 196
free fall, 209
Fresnel coefficient, 199
Fresnel, Augustin, 198

**G**

Galilean transformations, 207
Galilei, Galileo, 149, 180–181
'Dialogo sopra i due massimi sistemi del mondo', 172
Galileo affair, 170
Gauss, Carl Friedrich, 202
general relativity, 5, 21–26, 191, 192, 202–204, 208–210
genetic epistemology, 2
geographical knowledge, 21, 22, 38, 39, 63, 67, 84
geometry
Babylonian, 15–17, 94, 112–117
differential, 204
dynamic, 26, 203
Euclidean, 19, 20, 94, 107, 115–117, 195, 197, 207
non-Euclidean, 19, 20, 193, 194, 202, 207
pseudo-Euclidean, 201, 202
Gipper, Helmut, 85
gnomon, 136, 159, 160, 163, 166
gravitation, 147, 152, 191, 197, 201–204, 208–210
law of, 187
theory of, 22–24, 26, 201, 204, *see also* field theory of gravitation
gravitational
center, 151–153
field, *see* inertio-gravitational field
force, 25, 185, 187, 201, 204, 209, 210
waves, 21, 24
Greece, 18, 123, 142, 209
Grosseteste, Robert, 146

**H**

habitus, 5
Heaviside, Oliver, 198
Hertz, Heinrich, 186, 198
Hilbert space, 206
Hilbert, David, 26
Hipparchus of Nicaea, 155
homogeneity of space, 4, 16, 24, 25, 196, 201, 208

Hopi (Arizona, USA), 42, 43, 46, 66, 74, 81, 83, 84
house building, Eipo, 14, 61, 63, 67–70, 73, 74
Hume, David, 23

**I**

Ibn Rushd, 145, 170, 172
impetus, 175, 183
India, 124
inertia, 183–186, 196–198, 201, 203, 204, 209, 210
  law of, 188, 197
inertial system, 196, 197
inertio-gravitational field, 24, 26, 203, 204
infinity of space, 150, 156, 157, 172, 177, 179, 180
institutions, 2, 17, 18, 21–23, 59, 61, 113, 123, 208
isotropy of space, 4, 24, 25, 196, 208

**J**

'Jiu zhang suan shu' ('Nine Chapters on Arithmetical Techniques'), 15

**K**

Kant, Immanuel, 1, 2, 22, 23, 186–189
  'Critique of Pure Reason', 1, 22, 186–188
  'Metaphysical Foundations of Natural Science', 1, 22, 23, 186–189
  'Monadologia physica', 186
Koyré, Alexandre, 171
Kuhn, Thomas, 146, 172

**L**

landmarks, 6, 7, 10, 13, 22, 27, 36–40, 44, 45, 56, 64, 70, 73–75, 77, 79–85, 93, 197
language, 3, 5, 11–13, 17, 18, 35–86, 93, 94, 97, 121, 123, 143, 195, 207
  formalized, 27
  technical, 26
Latin Middle Ages, 18, 124, 146
Laue, Max von, 202
Leibniz, Gottfried Wilhelm, 22, 159, 186, 187
Leibniz-Clarke correspondence, 22
Leibnizian
  mechanics, 208
  spacetime, 197
Leucippus, 139, 176, 177, 179
Lévi-Strauss, Claude, 14
light, 21
  and electromagnetism, 198
  as a means of measuring space, 193, 194
  bending, 24
  propagation, universal law of, 191
  rays, 20, 198
  speed of, 201
  wave theory of, 198
logic, 125, 127, 130, 140, 142
logical positivism, 1
Lorentz transformations, 201
Lorentz, Hendrik Antoon, 199, *see also* Maxwell-Lorentz theory
Lucretius, 176–178

**M**

Mach, Ernst, 203
magnetic
  compass, *see* compass, magnetic
  field, *see* electromagnetic field
Malotki, Ekkehart, 40, 42, 43, 74, 85
maps (geographic), 8, 22, 47, 109
Maxwell-Lorentz theory, 199–201, 208–210
Maxwell's equations, 199, 207, 209
measurement, 16, 17, 94, 123, 127, 193–195, 200–202
  in Eipo culture, 14, 69, 74
  in quantum mechanics, 24, 204, 205
  of fields, 15, 99–101, 109, 115

measuring rods, 16, 18, 20, 99, 123, 126, 128, 132–135, 137, 141, 142, 193, 194
  and clocks, 192
mechanics, 125, 142, 178, 179
  analytical, 205
  celestial, 21, 209, 210
  classical, 21, 23–26, 196–198, 203, 207, 208, 210
  Euler's, 185, 186
  Hertz's, 186
  Leibnizian, 208
  quantum, *see* quantum mechanics
  statistical, 206
mental model, 6, 9, 16, 19, 22, 27, 40, 43, 115, 117, 175, 176, 179–182, 188, 189, 194, 195
  of large-scale space, 10, 12–14, 37, 38, 40, 72, *see also* cognitive mapping
  of object, 9, 10, 179
Mesoamerica, 15
Mesopotamia, 15, 17, 18, 94–117, 123, 195
metaphysics, 23, 151, 171, 175–176, 178–183, 185, 187–189
metric tensor, 191, 193, 197, 201, 203, 204
Michelson-Morley experiment, 198–200
microscope, 21
Minkowski
  formalism, 204, 209
  spacetime, 26, 201
Minkowski, Hermann, 201
Minsky, Marvin, 9, 10
'Mohist Canon', 18, 19, 121, 124–143
monads, 186, 187
mythology, 12–14, 18, 21, 37, 51, 61, 62, 68, 70, 72, 85, 123

**N**

nativism, 5
Navajo (south-western USA), 53, 54, 66, 79
navigation
  deep-sea, 22
  Micronesian, 13, 15
  radio, 22
Neolithic, 14, 63, 74
Neoplatonism, 18, 124
Newton, Isaac, 183, 185, 189, 197, 210
  his atomism, 183–185
  his axioms, 185, 186, 188, *see also* inertia, law of
  his bucket experiment, 184
  his concept of space, 4, 5, 21–24, 188
  'De gravitatione …', 183
  'Opticks', 185
  'Philosophiae naturalis principia mathematica', 184, 185
Newton-Cartan theory, 210
Newtonian
  dynamics, 186
  gravitation, 23, 187, 201, 203, 209, 210
  physics, 26, 196, 197, 201, 203, 208, 210, *see also* mechanics, classical
  spacetime, 192, 196, 197
Ngatatjara (Australian desert), 13
Nordström, Gunnar, 26, 202, 204
nothing, 23, 139, 177
nuclear forces, 26

**O**

ontogenesis, 2, 3, 5, 11, 20, 36, 37, 84, 122, 141
Oresme, Nicole, 196
Osiander, Andreas, 170

**P**

Paleolithic, 11
Parmenides, 143
path integration, 10
path-connectedness of space, 8
Penrose, Roger, 24
perceptional space, 8, 11
perihelion motion, 24

permanent object, schema or model of, 6, 9, 10, 27
perspective, 7–11, 16, 41, 42
Philoponus, John, 145
phylogenesis, 2, 3, 11, 36
Piaget, Jean, 2, 6–8, 11, 37, 122, 141, 194
place, 7, 10, 12–14, 35, 51, 65, 76, 84, 85, 195
 Aristotle's concept of natural place, 145, 149, 150, 152, 153, 196
 definiteness and exclusivity of, 7, 122, 140, 176, 177, 182, 184
 geographical, 42, 47, 155, 156, 160, 163
 in Dene Chipewyan, 79, 81, 82
 in Eipo, 13, 67–79
 in the 'Mohist Canon', 126, 127, 129–131, 139
 names, *see* toponyms
 philosophical concept of, 139, 140, 182, 183
 sacred, 60, 62, 75
planetary motion, 21, 24, 151, 152, 155, 170, 184, 196
Plato, 1, 150–152, 169
 'Republic', 1
 'Timaeus', 148
Platonism, 169, 171
Poincaré, Henri, 193, 200, 201
position-quality model of space, 23, 24, 205
primates, nonhuman, 5–8
projectile motion, 196, 209
proportionality of force and acceleration, 188
Ptolemaic cosmology, 146, 153–173
Ptolemy, Claudius, 145, 146, 148, 153, 171, 172
 'Almagest', 145, 146, 153–159, 161–165, 167–173
 'Planetary hypotheses', 155, 169
Pythagorean
 numbers, 111
 theorem, 115
Pythagoreanism, 1, 171, *see also* cosmology of the Pythagoreans

**Q**

quantum
 electrodynamics, 25
 field theory, 24, 25, 204, 205, 210
 gravity, 24, 26
 mechanics, 5, 24, 25, 204–206

**R**

reflective abstraction, 17, 19, 20
reflexivity, degree of, 207
Reinhold, Erasmus, 147
relativism, linguistic, 43, 85
relativity
 general, *see* general relativity
 of motion, 198
 principle, 198, 202, 203
 special, *see* special relativity
representational space, 8, 11
Rheticus, Georg Joachim
 'Narratio prima', 152
Riemann, Bernhard, 202
Riemannian space, 193, 194
rigid body, 20, 193, 194
Roman surveyors (agrimensores), 17, 106

**S**

Sapir, Edward, 46, 85
Sapir-Whorf hypothesis, 43, 85
sensorimotor behavior and intelligence, 8–14, 16, 19, 37, 194
sexagesimal positional system of numerical notation, 16, 17, 105, 109–117
shared knowledge, 3, 10–14, 18, 20, 27, 36, 44, 84, 117, 123, 143, 195, 208
short cut, 6
Simplicius, 145, 177

Slavey (Great Slave Lake, Canada), 46, 54, 66, 79
Snell's law, 198
social cognition, 11
South-America, 15
special relativity, 21, 24–26, 191, 192, 194, 201–204, 208, 209
spherical cosmos, 21, 146, 155, 156, 159, 160, 170, 172, 196
sphericity of the Earth, 21, 22, 146, 152, 154, 160
star compass, Micronesian, 13
stellar aberration, 198
string theory, 24
Sumerian, 97, 108, 112
surveying, 4, 15–17, 94, 99–102, 106–110, 113, 115–117
surveyors' formula, 16, 17, 106–107, 116, 117
symbol
  systems, 11, 17, 26, 27, 36, 94, 117, 121
  use, 9, 11
symbolic
  function, 11
  notation, 16, 19, 95, 96
  representation, 9, 16, 17, 22, 27, 40, 43, 64, 94, 97, 110, 123, 142, 195, 207

**T**

Temne (northern Sierra Leone), 14
three-dimensionality of space, 7, 8, 10, 16, 95, 98, 160, 191
topological
  relations, 8, 10, 47, 67, 75, 85, 109
  space, 1, 8
toponyms, 12, 13, 37, 39, 40, 47, 61, 67, 84

**U**

universality of gravitation, 187, 201, 202, 204, 209
universals of cognition, 5, 10, 26, 40, 43, 84, 94, 110, 122, 123, 140

**V**

vacuum, *see* void
vertical direction, 7, 22, 76, 152, 153
void, 7, 19, 22, 23, 122, 127, 128, 137, 139, 177–184, 187

**W**

Whorf, Benjamin Lee, 85
wind directions, 6, 10, 13, 49, 81, 156
writing, 4, 11, 15, 16, 18, 19, 23, 94, 95, 97, 106, 110, 113, 116, 117, 121, 123

**Z**

Zeno of Elea, 141
Zenodorus, 159