The History of Vision, Colour, & Light Theories
The History of Vision, Colour, & Light Theories

Introductions, Texts, Problems

Gábor Á. Zemplén

Bern Studies in the History and Philosophy of Science
Educational materials 5
Contents

I  Introduction

1  The Aim of the Book .......................... 13
2  Choice of Topics .............................. 13
3  Structure .................................... 14
4  Preliminaries to the First Part of the Book ........ 15
5  Preliminaries to the Second Part of the Book ....... 18

A  Beginnings ................................. 21

II  Presocratic Theories and Their Critique .......... 23

1  The Presocratics: A Short Guide .................. 23
2  Heraclitus .................................... 24
3  The Parmenidian Challenge ....................... 24
4  Empedocles ................................... 27
5  Anaxagoras .................................... 28
6  The Critique of Theophrastus ....................... 29
7  Texts ........................................ 30
7.1  Text 1: Heraclitus of Ephesus .................. 30
7.2  Text 2: Parmenides of Elea ..................... 30
7.3  Text 3: Empedocles of Acragas ................. 31
7.4  Text 4: Anaxagoras of Clazomenae ............. 34
7.5  Text 5: Theophrastus on Empedocles ............ 36
7.6  Text 6: Theophrastus on Anaxagoras .......... 39
8  Further Problems ................................ 41
9  Suggested Readings ............................ 41

III  Plato ..................................... 43
| 1  | Introduction                                      | 43 |
| 2  | The *Timaeus* and Plato’s Writings on Vision, Colour, and Mirroring | 44 |
| 3  | Greek Colour Terms                                | 45 |
| 4  | The Pythagorean Influence                        | 49 |
| 5  | Texts                                            | 50 |
| 5.1| Text 1: *Timaeus* 45a-46c                         | 50 |
| 5.2| Text 2: *Timaeus* 67c-68d                         | 52 |
| 6  | Further Problems                                 | 56 |
| 7  | Suggested Readings                               | 57 |

*IV*  
Aristotle and the Aristotelian corpus  
| 1  | Introduction                                      | 59 |
| 2  | Atmospheric Optics                                | 60 |
| 3  | On the Soul and the Senses                        | 63 |
| 3.1| On the Soul                                      | 63 |
| 3.2| Alexander of Aphrodisias                          | 64 |
| 3.3| On the Senses                                    | 69 |
| 4  | On Coloured Bodies: *De Coloribus*                | 70 |
| 5  | Texts                                            | 72 |
| 5.1| Text 1: Meteorology                              | 72 |
| 5.2| Text 2: On the Soul                              | 73 |
|     | Book II Chapter 5                                | 73 |
|     | Book II Chapter 6                                | 75 |
|     | Book II Chapter 7                                | 76 |
| 5.3| Text 3: Sense and Sensibilia                      | 79 |
|     | Book I Chapter 1                                 | 80 |
|     | Book I Chapter 2                                 | 81 |
|     | Book I Chapter 3                                 | 84 |
| 5.4| Text 4: On colours                               | 87 |
|     | Chapter 1                                        | 87 |
|     | Chapter 2                                        | 89 |
|     | Chapter 3                                        | 90 |
|     | Chapter 4                                        | 92 |
|     | Chapter 5                                        | 92 |
|     | Chapter 6                                        | 96 |
| 6  | Further Problems                                 | 99 |
V Atomist Theories, Epicureanism, Lucretius

1 Introduction ........................................... 101
2 Early Greek Atomism .................................. 101
3 Epicurus ............................................... 102
4 Lucretius ............................................... 104
   4.1 The Poem De Rerum Natura ....................... 106
5 Texts .................................................. 108
   5.1 Text 1: Democritus on Vision ................... 108
   5.2 Text 2: Epicurus: Principal Doctrines .......... 110
   5.3 Text 3: Lucretius: De Rerum Natura IV .......... 112
6 Further Problems ....................................... 113
7 Suggested Readings .................................... 113

VI The Mathematical Tradition

1 Introduction ........................................... 115
2 The Beginnings of Mathematical Optics ............... 116
3 Euclid .................................................. 117
4 Hero .................................................... 117
5 Ptolemy ................................................ 119
   5.1 The Optics ......................................... 120
6 Texts .................................................. 122
   6.1 Text 1: Euclid’s Optics ......................... 122
       Proposition I ...................................... 123
       Proposition II ..................................... 123
       Proposition III ................................... 123
       Proposition IV ................................... 123
       Proposition V ..................................... 124
       Proposition VI ................................... 124
       Proposition VII .................................. 125
       Proposition XLV .................................. 125
       Proposition XLVIII ................................ 125
   6.2 Text 2: Hero’s Catoptrics ......................... 126
   6.3 Text 3: Ptolemy’s Optics, Book II, Introduction 133
   6.4 Text 4: Ptolemy’s Optics, Book II, Illusions .... 143
   6.5 Text 5: Ptolemy’s Optics, Book V, Refraction ... 160
2.2 Psychological Explanations in the Middle Ages: Alhazen
2.3 Natural Geometry or Contingent Cues
2.4 Further-larger-nearer Theories

<table>
<thead>
<tr>
<th>Texts</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Text 1: Ptolemy’s <em>Almagest</em></td>
<td>297</td>
</tr>
<tr>
<td>3.2 Text 2: Samuel Dunn, 1762</td>
<td>298</td>
</tr>
<tr>
<td>3.3 Text 3: Ptolemy’s <em>Optics</em></td>
<td>302</td>
</tr>
<tr>
<td>3.4 Text 4: Alhazen’s <em>Optics</em></td>
<td>302</td>
</tr>
<tr>
<td>3.5 Text 5: Descartes on Natural Geometry</td>
<td>308</td>
</tr>
<tr>
<td>3.6 Text 6: Berkeley on the Illusion</td>
<td>310</td>
</tr>
<tr>
<td>3.7 Text 7: Kaufman and Rock</td>
<td>318</td>
</tr>
</tbody>
</table>

4. Epilogue
5. Further Problems
6. Suggested Readings

XI Molyneux’s Problem: A Congenitally Blind Recovering the Faculty of Sight
1. Introduction
2. The question
3. Answers to a Hypothetical Problem
4. The (Ir)relevance of Empirical Data
   4.1 The First Publication: Cheselden
   4.2 Testing Object Recognition: Home
   4.3 A Sphere and a Cube: Franz

5. Texts
   5.1 Text 1: William Molyneux
   5.2 Text 2: John Locke
   5.3 Text 3: George Berkeley
   5.4 Text 4: Thomas Reid
   5.5 Text 5: Edward Synge
   5.6 Text 6: William Cheselden
   5.7 Text 7: Everard Home
   5.8 Text 8: John Charles August Franz

6. Epilogue
7. Further Problems
8. Suggested Readings
List of Figures

| II.1  | The Greek peninsula and Western Asia Minor         | 25   |
| II.2  | The Italian peninsula and Sicily                   | 26   |
| III.1 | Pythagorean ratios in the spectrum                 | 50   |
| III.2 | Vision according to Plato’s *Timaeus*              | 53   |
| III.3 | Mirroring according to Plato’s *Timaeus*           | 54   |
| IV.1  | Computer-generated picture of an atmospheric halo  | 62   |
| VI.1  | North-eastern Africa and Asia Minor                | 116  |
| VI.2  | Refraction in nature from Vitruvius: *De Architectura Libri Decem* | 118  |
| VI.3  | Hero’s *Catoptrics*: surprising the observer       | 131  |
| VI.4  | Hero’s *Catoptrics*: seeing selected images        | 132  |
| VII.1 | Rainbow-formation after Theodoric of Freiberg      | 172  |
| VII.2 | The sine law of refraction                         | 182  |
| VII.3 | The Cartesian explanation of the rainbow           | 189  |
| VIII.1| The colour triangle of Johannes Zahn                | 206  |
| VIII.2| Goethe’s simple prismatic experiments              | 250  |
| IX.1  | Burning lenses in Kircher’s *Ars magna lucis et umbrae* | 264  |
| IX.2  | The eye as *camera obscura*                        | 267  |
| IX.3  | Light corpuscles moving through a prism             | 270  |
| X.1   | Alhazen’s depiction of the eye                     | 294  |
| X.2   | Descartes’s blind man                              | 308  |
| X.3   | Distance perception for Descartes                  | 309  |
| X.4   | Berkeley on the eye and vision                      | 311  |
| XI.1  | Visual recognition of objects                       | 339  |
FOREWORD

Light and sight, the subjects of this delightful history, have fascinated thinkers from the dawn of time. We do not know the origins of ruminations concerning them, but we are aware of their potency. Representations of animals on the walls of caves preceded the representation of ideas in written language. It is with the latter that this history starts, and it charts the development of observations and theories from the early Greek philosophers. This is not a standard history nor was it intended to be. On the one hand, several themes are chosen, and examined in detail. These include the nature of colour as expressed in a variety of phenomena, like colour mixing, the rainbow, and the prismatic spectrum. Spatial perception, as evidenced in the moon illusion and in vision following the restoration of sight, is also addressed. On the other hand, selected texts provide the sources from which evidence regarding theories of all these phenomena can be adduced. Throughout this historical odyssey questions are posed for the reader to answer, hints are given regarding the likely resolution of them, and the texts provide the context in which they can be understood.

The subtitle of the book, and Gábor Zemplén’s stated intent, is that this should provide an introduction to “those who are interested in theories of light, colour, and vision”. Its appeal will engage a much wider audience than the students to whom it is specifically directed. The texts selected provide ready access to materials that are often widely scattered, and pursuing the references in the bibliography will provide rich rewards. Although Aristotle received the lion’s share of reproduced text, the particular strength of the book is the coverage of pre- and post-Newtonian theories of colour. Inclusion of discussion of the moon illusion is most apposite: the phenomenon was recorded before the time this history starts and its interpretation remains enigmatic! History is not restricted to the past; it engulfs us all in myriad ways. When the history considered here commenced, light was not distinguished from sight. It was a great achievement of science to separate them. What is displayed in Gábor Zemplén’s book is that the history of sight mirrors the history of science.

Nicholas J. Wade
08.12.2003, Dundee
ACKNOWLEDGEMENTS

First and foremost I have to thank Dr. Timm Lampert for inviting me to teach at the University of Bern, Switzerland, for the winter semester of the academic year 2001/2002, and Prof. Gerd Graßhoff for supporting the idea. The department has been of much help, especially Käri Nickelsen, Hans-Christoph Liess, and Katharina Schaffer, to whom I could always turn in case of trouble. Robert Casties deserves special acknowledgment for his tireless \LaTeX support, both in Bern, and at the Max Planck Institute for History of Science, Berlin, where I finished working on the manuscript.

Helpful comments from Márta Fehér, Katerina Ierodiakonou, Timm Lampert, Csaba Pléh, Ivo Schneider, Friedrich Steinle, and Nick Wade have helped to make corrections and improvements. I thank especially Gábor Betegh for his thorough and insightful comments for the whole first part of the book, and Judit Gervain, who has done a gargantuan work by proofreading the material in different stages of the production. She, together with the rest of my family, have been entirely supportive. I am particularly grateful to my father, who remained inquisitive and rejoiced in the whole project throughout his illness.

I received financial aid from the Bavarian Free State, and the Sasakawa YLF Fund. The manuscript has been prepared with support from OTKA T 037575 and T 037504 grants (for chapters 2-4, 7-9, and 5-6, 10-11, respectively) in the History and Philosophy of Science Research Group of the Hungarian Academy of Sciences, at the Department of Philosophy, Budapest University of Technology and Economics. During the last phase of the work I was supported as Magyary Zoltán, and, finally, as Békésy György post-doctoral fellow.

The staff of the Deutsches Museum, Munich deserve special acknowledgement, as they have agreed to the free reproduction of pictures found in the Archives. Apart from some simple drawings, all reproduced pictures and some of the texts are from this source, even if not individually acknowledged.
Chapter I

INTRODUCTION

1 The Aim of the Book

This book is written with students in mind who are interested in theories of light, colour, and vision, but had no systematic training in one or more of these areas. Originally taught as a course to a wide variety of undergraduate students, the material has been rewritten as a book to suit students of various disciplines—physics, chemistry, biology, psychology, visual studies, art history, philosophy, history, cultural studies, etc.—, including postgraduate students in history of science.

Unlike in a textbook, the aim of this work has not been to provide a concise and structured body of knowledge. The emphasis is rather on helping students to gain practice in reading primary texts in the history of science, and on developing a multidisciplinary and sensitive approach to these texts. For this reason the stress is more on learning to ask questions than on learning the answers, more on seeing differences in attitude, claims, and arguments than on giving clear-cut answers.

Although to develop and defend clear positions is highly preferred in the history of science as a discipline, for a book introducing primary sources to depend stubbornly on one of the many rival positions seems undesirable.

2 Choice of Topics

The book provides ample material for a semester’s course and most of the chapters admit to a chapter-a-week approach, but certain selectivity can be beneficial. Some chapters, like the one on Aristotle, would require more than one week for most classes, and the time consumed would depend greatly on the previous training of the students.

It is notoriously difficult to select introductory texts for the history of science. The bias in selection cannot be avoided, at best it can be made explicit. This book is admittedly unable to provide generous information about the cultural context of the works investigated. This springs from the nature of the selection. In choosing these the main stress was on introducing specific problems rather than the cultural setting of these problems. As
the book contains a large selection of texts, it also necessarily presupposes a set of canonical works, from which the selection has been made. As in most cases the students will have different backgrounds, a rather wide range of topics has been included. This way both Physics, Psychology, Classics, or Art History students should feel that for certain topics their expertise can help them disentangle ambiguous passages. At the same time most students will find chapters tackling previously unencountered problem-areas.

3 Structure

In general, the chapters, after a short introduction, give information about selected aspects of the primary texts. These are followed by a number of questions, the texts themselves, and hints that help answering the questions raised. Each chapter ends with a section on further problems, that could be worked out in detail. After this list of possible essay topics a selected bibliography of books and articles that are relevant to the given chapter are listed. The first part of the book pinpoint to problems in some of the most important texts of Classical Antiquity. The second part, on the other hand, selects the texts where the development of certain problems can be further traced.

The introductions are rather short, and footnotes and references have been reduced to a minimum. They aim at giving a very coarse outline to help the appreciation of the primary texts. They might seem superfluous for some readers, as they contain general knowledge, but as the book touches on a number of topics which are widely separated in today’s divergent disciplines, most readers will find at least some of these introductions worth consulting.

In formulating the problems, several different aspects had to be considered. They should focus the reader on particular problems in the text and should not require special expertise. At the same time they should help the reader appreciate scientific problems, and, if possible, current issues of debate in the history of science. Also, they ought to facilitate attempts at interpretation, at textual understanding. The texts should, therefore, be read with these problems in mind. These will focus the attention of the reader to aspects that might otherwise be overlooked, and in some cases help to understand the texts themselves.

To further attention to the texts and not to replace it, hardly any defin-
itive answers are given to the problems posed. Instead, “hints” are provided that, once the students attempted to answer the questions, might give further insight into the problems. This extra information should help the reader arrive at an acceptable and defensible position. For individual work, these suggestions should indicate to the reader whether or not she is on the right track. In case of classwork they could serve as starting points for discussions. No hints are given for straight-forward questions.

The “Further Problems” sections provide a thread for more individual work on the themes discussed in the chapters. They are devised to help the students find topics for their end-term essays—an often recurring problem for interested non-specialists who try to fulfill course requirements in History of Science courses. Here, as in the “Suggested Readings” sections, ample literature is given that enables students to start individual research on topics that have aroused their interests.

### 4 Preliminaries to the First Part of the Book

The chapters in the first part of the book titled “Beginnings” contain writings that lie at the roots of several scientific contemporary disciplines. They touch upon areas that today belong to atmospheric, physical, or physiological optics, to vision science, colorimetry, plane geometry, philosophy or even pathology—just to list a few.

The works discussed here were written in a timespan of seven centuries, and the authors lived on three continents: Africa, Asia, and Europe. Still there is a significant continuity in the problems tackled and the methods used for finding solutions. This would make it possible to follow the development of the answers to certain problems, of methods, and arguments. (Such a horizontal history of specific problems will be the main theme of the second part of the book.) In this first part the emphasis will be laid on working with primary (i.e. English translations thereof) texts. This text-centered approach allows for the introduction of the major problem areas—many of which will resurface in the latter part of the book.

While the aim is far from providing a thorough vertical history of the Graeco-Latin theories of light, colour, and vision, the rather wide selection of texts allows the reader to appreciate the depth and scope of the problems related to these theories.

Some general remarks are in order when working with these texts. First
of all, they are from very different sources. We are used to the idea that texts have a date of origin, a writer, and a language in which they were written. But with the sources in this part of the book the answers to these questions are far from obvious. We could say that the most elaborate Ancient work on light, Ptolemy’s *Optics* was written by Claudius Ptolemy in Greek possibly around 175 A.D. in the city of Alexandria, now in Egypt (see map on page 116).

Unfortunately no part of that book can be cited in the present volume, as we possess no copy of the original work. There is a single, badly mangled manuscript in Latin from the twelfth century, which is a translation of an Arabic treatise (which is probably a translation of a copy of Ptolemy’s work in Greek) by Eugene of Sicily, an important translator of Arabic texts to Latin.

Is this work Ptolemy’s *Optics*? Is it identical to the original? To the second question the answer is a definite no: the Latin version does not have the first book of the original *Optics*, and breaks off in the middle of the fifth book. Apart from these it has several other gaps, also called *lacunae*, in the text, and a number of incoherences. But the answer to the first question is nevertheless that this is the text that most resembles Ptolemy’s *Optics*; we have nothing better to rely on.

Such examples are very common, especially for our knowledge of the Presocratics—a general, but slightly misleading term for Greek philosophers who lived and worked before Plato, who died in 347 B.C.—living nearly seven centuries before Ptolemy. Although most are supposed to have written whole books, what we possess are at most fragments that surface in the writings of authors some of whom had lived nearly a millennium after the writers in question. In these cases several other issues have to be addressed.

By whom are these fragments preserved? Some, like Aristotle and Theophrastus often used the Presocratic views for polemical reasons to show their weaknesses. In certain cases we can guess whether the distortions are willful or even malignant or simply result from the difficulty of reconstructing the earlier views expressed in a significantly different terminology in spite of the most sincere interpretative efforts—in any case, the original views might remain obscure for ever.

It was also common practice to cite earlier authorities to show the origins of ideas purported by the writer. Several of Heraclitus’ fragments survive through Clement of Alexandria, who lived in the 2nd century A.D.,
a Christian father treating him as a pagan prophet of the last judgement. The writings of the same Heraclitus, in the eyes of the Stoic Cleanthes, supported his master, Zeno. Any or all of these readings can lead to deliberate or accidental misinterpretation.

Moreover, time and reliability does not always go hand in hand, the proper weighing of the often contradictory passages requires delicate balancing of many factors in cases where often both the reliability and the authenticity of the sources are dubious.

Do the fragments really represent the views of the thinker to whom they are attributed? Most of what we possess from the primary texts are short quotations dispersed in later writings, but many of these are indirect citations (‘x said that y’, or even ‘x is said to have held that y’). Even when a quote is claimed to be verbatim, i.e. word for word, it is impossible to verify this claim.

Frequently, fragments or even whole works are attributed to someone, but this judgement is questioned by later scholars: although the book On Colours, also mentioned as De Coloribus, reproduced in Chapter IV is still printed as part of Aristotle’s complete works, there is now consensus that it was not written by Aristotle.

And there are all the nitty-gritty details: sometimes only a single version of a text exists, partly destroyed by time, at other times different copies contain different versions: dissimilar letters or words, conflicting sequences of the paragraphs, incoherences, text elements that are probably the notes of the copyist or translator. At still other times, later texts are amalgamated with the original, often obscuring the initial ideas in the book. We often possess only translations of already translated texts—the case of Ptolemy’s Optics is not exceptional in this respect. It is not uncommon that translations and copying seriously damage the texts and much of the specialised terminology might drop out in this process.

Since translations are reproduced in the book, most of these difficulties remain invisible. Thus only a single interpretation is conveyed, one that is a result of careful reconstruction of the original works, at the same time itself a possible source of misinterpretation. Therefore, it is especially advised not to jump to conclusions concerning the texts without detailed study of the originals. And the note that the book is merely an introduction to these ideas is to be taken doubly seriously for this part of the work.
The chapters of the first part are roughly chronological. First some Presocratic views are introduced, a selection of the fragments of Heraclitus, Parmenides, Empedocles, and Anaxagoras. Instead of working with these directly, they are investigated together with the critique of Theophrastus: showing how a later age conserved, interpreted, and criticised them.

The next chapter discusses Plato’s views on vision, colours, and mirroring. Some inherent difficulties of Greek colour terms are highlighted.

The most bulky chapter of the first part is the analysis of the Aristotelian corpus. Not only atmospheric optical phenomena (in the Meteorology) but sense-physiology and visual perception are discussed as well. The pseudo-Aristotelian treatise On colours reveals one of the most sophisticated Ancient accounts of colours.

The atomist theories are the focus of the following chapter, mainly discussing the Lucretian poem De rerum natura.

The last chapter follows the development of the mathematical tradition from Euclid, through Hero, to Ptolemy.

5 Preliminaries to the Second Part of the Book

The second part of the book tried to satisfy three important goals. The first was to provide significant scientific texts from different periods. The sources range from texts by Alhazen and Grosseteste in the Middle Ages, through Descartes, Hooke, Newton, Berkeley, and others in the seventeenth an eighteenth centuries to medical records in the nineteenth century and important articles on perception in the second half of the twentieth century.

Secondly, various problems and theoretical considerations—some already mentioned in the first half of the book—are investigated in more detail. Areas discussed include geometrical optical problems, like the law of refraction, the explanation of apparent colours for rainbows, interference colours and refraction, debates on colour-systems, studies on size and distance perception and the moon illusion, philosophical and experimental work on the visual modality and the Molyneux problem.

The third consideration was to unite the various times and disciplines in chapters that could be studied on a week-to-week basis. The five chapters thus created try to encompass the manifold by selective attention to a relatively small number of texts. Naturally much had to be left out, but the
selection still enables the reader to have an overview of the diverging histories of several disciplines, without sacrificing the method of close reading and analysing scientific texts.

Chapter 6 on the rainbow re-investigates in more detail the Aristotelian theory, together with Robert Grosseteste’s medieval account and Descartes’s very important contribution to finding a satisfactory explanation of the bow.

The next chapter builds on Grosseteste’s and Descartes’s theories of colour and further studies the little known modificationist colour theories by discussing Robert Hooke’s colour theory in his 1666 Micrographia and parts of one of the latest and most meticulous anti-Newtonian modificationist theory by Johann W. von Goethe.

In Chapter 8 the first public exposition of Isaac Newton’s new theory of light and colours is investigated—the theory that replaced the earlier ubiquitous modificationist accounts. This middle chapter is structured somewhat differently from the others: to give a paragraph-to-paragraph close reading of Newton’s 1672 letter, the text has been broken up and not reproduced intact.

One of the longest standing unexplained phenomena is the focus of Chapter 9. The moon illusion (the increased size of the moon on the horizon), already known by the Ancients still triggers new discussions, and in the first years of the twenty-first century still no consensual account has been accepted. The texts include Ancient authors (different accounts of Ptolemy), Medieval ones (Alhazen), and later texts up to the twentieth century.

The last chapter discusses the Molyneux problem: whether a congenitally blind person, capable of distinguishing between a sphere and a cube by touch alone would, if he suddenly recovered his sight, be able to distinguish and name the two objects without touching them. The question was one of the central issues in the development of the theories of vision (and early psychology) for centuries. Positive and negative answers abound in the history of the problem, and philosophical arguments as well as experimental accounts of the still unanswered question are discussed, which might be gaining increased significance in the light of recent developments in the study of cortical plasticity and intermodality.
Part A

Beginnings
Chapter II

PRESOCRATIC THEORIES AND THEIR CRITIQUE

1 THE PRESOCRATICS: A SHORT GUIDE

We could say that the systematic study and collection of the views of earlier thinkers who had lived before Socrates started with Theophrastus, the disciple of Aristotle. These collections of “opinions” remained popular for centuries, and it is mostly due to this doxographic tradition that we can reconstruct the views of the Presocratics. These early thinkers mostly lived in Asia Minor, the southern part of the Italian peninsula, and the southern Balkans. Some travelled extensively, and the connotations we attach to the term “Presocratic philosophers” does little justice to their varied and wide-ranging interests in natural phenomena, religion, magical practices, and other areas.

The written record usually does not include exact dates of birth and death, instead the date of flourishing is recorded. The flourishing or acme attributed to the philosophers was partly based on concrete knowledge, but partly on the constructed tradition that supposed 40 years between consecutive generations of philosophers (the age difference between master and disciple).

A note on citations. The standard reference for the Presocratics is (Diels & Kranz 1922), abbreviated DK. These numbers will be listed on the margins of the cited texts. In this chapter the citations and translations are mostly taken from the widely used (Kirk, Raven & Schofield 1983), the number after the abbreviation “KRS” in the main text points to the relevant citation in the book. Also several paragraphs are cited from (Stratton 1917), which has been reprinted and is more available in (Stratton 1964). The plain numbers on the margins refer to this text (DS).

The materials cited are only English translations. As the Greek originals are usually fairly accessible, these are not included here. Reproducing all relevant texts in the original seemed unnecessary in a strictly introductory work. For a first reading and an approximate appreciation of the texts these translations suffice. No serious scholarly work, however, should be based on these alone.
Only selected fragments from Heraclitus, Parmenides, Empedocles, and Anaxagoras are included—together with the critique of Theophrastus—, for a more detailed discussion of their views any of the standard works can be consulted (see the Suggested Readings section).

2 Heraclitus

The first of the Presocratics whose fragments are included in this book is Heraclitus of Ephesus (for location see figure II.1). His acme according to Diogenes Laertius, whose *Lives of Famous Philosophers* is one of our most important sources concerning the Presocratics, was in the 69th Olympiad (DK 22 A1, KRS 190), that is around 504-501 B.C.

It is not certain whether Heraclitus left behind any systematic work on the topics he investigated. His general explanatory scheme is based on the idea that the world is constantly changing. Fire is a regulatory principle that is present in the outside world and in our souls. Wisdom is the understanding of the way the world works (and drunkenness, i.e. moisturing of the soul definitely hinders this process, DK 22 B117, KRS 231), and man should aim at comprehending the *Logos*, the underlying organising principle of things. This working is often obscured but things can reveal their deep structure even if they seem opposing on the surface: “An unapparent connexion is stronger than an apparent one” (DK 22 B54, KRS 207).

Instead of giving a detailed general summary of Heraclitus’s thought only a few fragments are included in Text 1. The interpretation of these poses special difficulties—even the Ancients considered them obscure.

3 The Parmenidian Challenge

Parmenides and the so called Eleatic philosophy had a profound influence on later Greek philosophers—not only the Presocratics, but also Plato and Aristotle had to face the Parmenidian challenge.

Born around 515 B.C. in Elea, in Sicily (see figure II.2, south of Neapoli) Parmenides is supposed to have written a single work. He is said to have travelled to Athens when sixty-five, and have met the young Socrates. The surviving fragments of the hexameter poem titled *On Nature* start with the description of a journey to a goddess at the gates of Day and Night, who teaches Parmenides of the “unshaken heart of well-rounded truth”, and of the “opinion of mortals” (DK 28 A1, KRS 288). The mystic
Figure II.1: The Greek peninsula and Western Asia Minor. This and other maps of the Ancient World are adopted from the Interactive Ancient Mediterranean Web site (http://iam.classics.unc.edu) but have been changed to include places mentioned in the text. It has been reused and redistributed under the terms of IAM’s fair use policy. Copyright 1998, Interactive Ancient Mediterranean.
journey serves as a *prooimion*—proem or introduction—; a truth-claim for the following parts of the book. Just like descensions to Hell, and other topoi, such structures were not uncommon at the time, and Plato’s *Timaeus* is also structured this way.

![Map of the Italian peninsula and Sicily](image)

**Figure II.2: The Italian peninsula and Sicily.**

The Goddess aims to teach Parmenides to distinguish between two ways, that of Truth (to perceive that existence is) and the deceptive way of Belief (to think that non-existence can also be), in which there is no truth at all.

Deduction from the basic dichotomy (that can easily be considered as trivial) leads to a number of counter-intuitive conclusions: what is cannot
have arisen from nothing (there is no nothing), it had to arise from some-
thing, but this something cannot be what is (as there is nothing else). Also
nothing else beside itself can come into existence—there is no empty space
where it could do so. According to Parmenides no rational account can be
given of origins (or cessation), including the origins of the world. There is
similarly no rarefaction or condensation, no discontinuity of the existing.
Some fragments are included in Text 2.

The result is a continuous, immovable, indivisible, finite, spherical
ple-num, and thus a conception of the world that is contrary to our everyday
senses in many ways. To reconcile the sensory impressions to this view
or find the hidden logical flaws of the argument remained a challenge for
Greek thinkers in the coming centuries.

One aim of the philosophers of the following generation had been to
escape certain consequences of the Parmenidian picture, i.e. to reintroduce
motion, or the possibility of empty space, etc.

4 EMPEDOCLES

Already in Parmenides we have seen the close connection between what
seems to be rational philosophising and shamanic journey and initiation.
The fragments of Empedocles display a similar coexistence of philosophy,
mystery, and magic.

Living in Akragas (also known as Agrigentum) in Sicily (see map on
page 26), Empedocles, the son of a wealthy citizen is said to have visited
Thourioi shortly after its foundation in 444/3 BCE\(^1\) (see figure II.2). He
was a well-known poet, statesman, natural historian, and philosopher. He
also worked in medicine, but, unlike the Pythagorean sects that flourished
in lower Italy, he is thought to have less interest in mathematics.

For a long time it was generally held that we possess fragments of two of
his poems, one \textit{On Nature} (or \textit{Peri Phuseos}), and another entitled \textit{Katharmoi},
but in light of a recently discovered spectacular new papyrus (see Suggested
Readings) more and more now believe that these are parts of a single work.

His philosophical views contain elements of the Eleatic school (show-
ing the influence of Parmenides) and of the Pythagorean sects. He (re)intro-
duced a theory of elements or “roots”—fire, air, earth, and water. The two

\(^1\) Often data like these are our main sources of identifying the dates and chronology of
the philosophers.
major forces in the world are the attracting Love (philia) and the separating Strife (neikos).

Among the numerous fragments, for the first time, there are some discussing a theory of perception and especially vision. Theophrastus in his *De sensibus* claims that “Empedocles says concerning colours that white is of fire and black of water” (DS 59), and a number of other elements of the theory are included in Text 3.

- **Problem 1** Summarise Empedocles’ views on perception. What do the fragments in Text 3 tell you about the structure of the eyes, the process of vision? List the problems you encounter when you treat this account as a unified theory of vision and colour-perception.

5 **ANAXAGORAS**

Clazomenae was an Asian Greek city west of Smyrna, belonging to the Ionian Confederation (see figure II.1, page 25). Anaxagoras was probably born here in 500 B.C. Around 480 B.C. (DK 59 A1, KRS 459) he came to Athens and pursued philosophical activities. He was prosecuted and tried—possibly for maintaining that the sun was a red-hot mass of metal—after which he withdrew to Lampsacus (north of Illium) and died around 428 B.C.

The surviving thousand words are probably fragments of a single book *On Nature*, testifying to Anaxagoras’s strong reaction to the views of the Eleatic philosophers. Without going into details of his views, it could be stated that for Anaxagoras matter is infinitely divisible, and that everything contains a portion of everything. Things are composite, contain seeds (see Text 4) which have colours and tastes. To eliminate the problem of coming-into-being he postulates that already the primeval mixture contains an infinite variety of substances.

- **Problem 2** Compare Anaxagoras’s and Empedocles’s views on perception in Text 3 and 4. Write a list of the similarities and of the differences, concentrating on those questions to which the texts attempt to give answers. Think of possible empirical support or reinforcement for each.
Most of the texts that have survived from the Presocratic philosophers had been preserved via the doxographic tradition. Although already Aristotle often cited the views of others, Theophrastus of Eresus (ca. 370/1-278), his disciple and follower as head of the Lyceum in Athens from 322 to 278 B.C., can be considered as the true founder of the doxographic tradition. Coming from the island of Lesbos (see figure II.1 on page 25) he joined Plato’s Academy at around the age of seventeen, but soon became a pupil of Aristotle at the Lyceum.

A large portion of the Presocratic views cited in this chapter have been preserved in one of Theophrastus’s works, usually referred to as On the Senses or De Sensibus (henceforth DS). It is a unified, well-structured treatise analysing earlier views on perception and the objects of sense. The opinions are separated into two major groups, whether ‘like is known by like’ or whether perception occurs by means of opposites. In the first group the Parmenides’s (DS 3-4), Plato’s (DS 5-6), Empedocles’s (DS 12-24) and Alcmeon’s views (DS 25-26) are investigated, in the second those of Anaxagoras (DS 27-37) and Clidemus (DS 39-48), while Democritus’s and Diogenes of Apollonia’s opinions are considered as unclear concerning the major dividing line. In the second part (DS 59-91) Democritus’s and Plato’s claims concerning sense objects are criticised.

Interestingly, the views of Aristotle, who was Theophrastus’ mentor and colleague for forty years, are not investigated. Similarly, after a surprisingly detailed and impassioned summary of the opinions of others, Theophrastus ruthlessly criticises these views—but does not give his own opinion about the subject. The most probable explanation for these surprising facts is that Theophrastus mainly accepted the Aristotelian explanation of perception, and further developed it in other (now lost) works.

This critique of earlier thinkers could therefore have served as a preliminary study, preceding the writing of his own thesis. Also, the structure shows that the aim was to apply criteria of the Peripatetic\(^2\) school for the evaluation of earlier (and more rudimentary) theories.

---
\(^2\) Aristotle set up his school in the Lyceum. As he liked walking around and to discuss his ideas with his colleagues, his school is often referred to as the Peripatetic School. Peripatetics are “people who walk around.”
• **Problem 3** Compare your criticism of Empedocles’ theory given to Problem 1 and those of Theophrastus in Text 5. Based on Texts 5 and 6 find the criteria according to which Theophrastus finds faults with the Presocratic theories. Describe the types of arguments he uses.

7  **Texts**

7.1  **Text 1: Heraclitus of Ephesus**

• The selected fragments included here are translated in (Kirk et al. 1983), and their numbers refer to the numbering of that volume. The numbers from (Diels & Kranz 1922) are on the margins.

DK 22 B 1  **KRS 194** Of the Logos which is as I describe it men always prove to be uncomprehending, both before they have heard it and when once they have heard it. For although all things happen according to this Logos men are like people of no experience, even when they experience such words and deeds as I explain, when I distinguish each thing according to its constitution and declare how it is; but the rest of men fail to notice what they do after they wake up just as they forget what they do when asleep.

DK 22 B 55  **KRS 197** The things of which there is seeing and hearing and perception, these do I prefer.

DK 22 B 107  **KRS 198** Evil witnesses are eyes and ears for men, if they have souls that do not understand their language.

DK 22 B 78  **KRS 205** Human disposition does not have true judgement, but divine disposition does.

DK 22 B 26  **KRS 233** A man in the night kindles a light for himself when his vision is extinguished; living he is in contact with the dead, when asleep, and with the sleeper, when awake.

7.2  **Text 2: Parmenides of Elea**

• The selected fragments labelled KRS are translated in (Kirk et al. 1983), and their numbers refer to the numbering of that volume. The numbers from (Diels & Kranz 1922) are on the margins. DS refers to (Stratton 1917), adapting the marginal pagination as well as the angle brackets.

DK 28 B 2  **KRS 291** Come now, and I will tell you (and you must carry my account
with you when you have heard it) the only ways of enquiry that are to be thought of. The one that [it] is and it is impossible for [it] not to be, is the path of Persuasion (for she attends upon Truth); the other that [it] is not and it is needful that [it] not be, that I declare to you is an altogether indiscernible track: for you could not know what is not—that cannot be done—nor to indicate it.

KRS 294 For never shall this be forcibly maintained, that things that are not are, but you must hold back your thought from this way of enquiry, nor let habit, born of much experience, force you down this way, by making use of an aimless eye or an ear and a tongue full of meaningless sound: judge by reason the strife-encompassed refutation spoken by me.

DS 3-4 Parmenides gives no definition whatsoever, saying merely that there are two elements, and that our knowledge depends upon the excess of one or the other. For according as the hot or the cold predominates does the understanding vary, there being a better and purer understanding derived from the hot; yet even such knowledge requires a certain proportion. “For ever as it finds the blend in their far-wandering members,” he says, “so does mind come to men; for that which has intelligence in men each and all is the same,—the substance of their members; since what is there in greater measure is their thought.”

For to perceive by the senses and to have intelligence are treated by him as identical; consequently both remembering and forgetting arise, by the mixture <of the elements mentioned>. But if there should occur an exact equality in the mixture, he does not make it clear whether there would or would not be thought, nor what would be the general state <resulting>. But that he also attributes perception to the opposite <element> in its own right is evident from the passage where he says that a dead man—since now the fire has left him—does not perceive light and warmth and sound, but does perceive cold and silence and the other contrasting qualities; and that absolutely all being possesses some power of knowing. Accordingly by this thesis he seems arbitrarily to preclude discussion of the difficulties attending his position.

7.3 TEXT 3: EMPEOCLES OF ACRAGAS

- The fragments included are translated in (Kirk et al. 1983). The numbers from (Diels & Kranz 1922) are on the margins. The longer citation is from (Stratton 1917), with the angle brackets of the original edition.

KRS 342 Narrow are the powers that are spread through the body, and many are the miseries that burst in, blunting thought. Men behold in their span...
but a little part of life, then swift to die are carried off and fly away like smoke, persuaded of one thing only, that which each has chanced on as they are driven every way: who, then, boasts that he has found the whole? Not so are these things to be seen or heard by men, or grasped by the understanding. You then, since you have turned aside to this place, shall learn: no further can mortal wit reach.

**KRS 343** Come now, observe with all your powers how each thing is clear, neither holding sight in greater trust compared with hearing, nor noisy hearing above the passages of the tongue, nor withhold trust from any of the other limbs [organs, parts of the body], by whatever way there is a channel to understand, but grasp each thing in the way in which it is clear.

**KRS 389** As when someone planning a journey through the wintry night prepares a light, a flame of blazing fire, kindling for whatever the weather a linen lantern, which scatters the breath of the winds when they blow, but the finer light leaps through outside and shines across the threshold with unyielding beams: so at that time did she [sc. Aphrodite] give birth to the round eye, primeval fire confined within membranes and delicate garments, and these held back the deep water that flowed around, but they let through the finer fire to the outside.

**KRS 395** Men’s wit grows according as they encounter what is present.

**KRS 396** It is impossible to bring [the divine] near to us within reach of our eyes to grasp him with the hands—although this is the main road of persuasion entering the minds of men.

**DS 7-11** Empedocles has a common method of treating all the senses: he says that perception occurs because something fits into the passages of the particular <sense organ>. For this reason the senses cannot discern one another’s objects, he holds, because the passages of some <of the sense-organs> are too wide for the object, and those of others are too narrow. And consequently some <of these objects> hold their course through without contact, while others are quite unable to enter.

Then he attempts to tell us the character of the organ of vision. Its interior, he says, is of fire; while round about this <internal fire> are earth and air, through which the fire, by reason of its subtilty, passes like the light in lanterns. The passages <of the eye> are arranged alternately of fire and of water: by the passages of fire we perceive white objects; by those of water, things black; for in each of these cases <the objects> fit into the given <passages>. Colours are brought to our sight by an effluence.

Yet <eyes, he holds,> are not all of like construction from these opposing
elements: in some <eyes> the fire is at the centre, in others it lies more external. Because of this, certain animals see better by day, others by night: by day those whose eyes contain less of fire <have an advantage>; for with them the light within is made equal <to the water within the eye> by the <light> without. But those whose eyes have less of the opposite <element—their vision excels> by night; for with them, also, their lack is supplied <from without>. But reverse the conditions, and the opposite is true: for now even the animals that have fire in excess are dim of sight <by day>, since the fire within—increased still further by the daylight—covers and occupies the passages of water. And the same thing happens by night to those with water <in excess>, because the fire is now overtaken by the water. This goes on until for the one group the <excessive> water is cut off by the outer light; and for the other, the <excessive> fire is cut off by the air. Thus each finds its remedy in its opposite. But that <eye> is of happiest blend and is best which is composed of both <these constituents> in equal measure. This represents fairly well what he says of vision.

[...] Pleasure is excited by things that are similar <to our organs>, both in their constituent parts and in the manner of their composition; pain, by things opposed.

In a like strain he speaks also of understanding and of ignorance. The one is due to what is like; the other to what is unlike; since in his view thought is either identical with sense perception or very similar to it. For after enumerating the ways in which we recognize each element by its like, he finally adds:

“For from these have all things been fittingly conjoined, and by their means do creatures think and have delight and suffer grief.”

Accordingly, we think chiefly with the blood; for here the elements are more fully mingled than in any other of our members.

Those in whom these mingled elements are of the same or nearly the same <amount>, being neither widely separated nor too small nor of excessive size—such persons are most intelligent and keen of sense; and others are intelligent and keen of sense according as they approach to such a mixture; but those whose condition is the very reverse are the least intelligent. Again, persons in whom the elements lie loose and rare are slow and laborious; while such as have them compact and divided fine are impulsively carried away; they throw themselves into many a project, and yet accomplish little, because of the impetuous coursing of their blood. But when the composition in some single member lies in the mean, the person is accomplished in that part. For this reason some are clever orators, others artisans; for in the one case the happy mixture
is in the tongue, in the other it is in the hands. And the like holds true for all the other forms of ability.

- **Hint to Problem 1** Emanations from an object reach the perciipient organ. This is an *intromissionist* theory: the eye receives the emanations, and if these fit exactly to the small pores in the organ, perception occurs. However, the powerful simile in KRS 389 seems to suggest that fire leaves the eyes, and this would support an *extramissionist* theory.

Accepting that the intromissionist account explains perception, while the extramissionist simile only the structure of the eye (fire may leave the eye, but it is not certain that it reaches the object or that it is responsible for vision) the two are not necessarily incompatible.

It is possible that Empedocles’ description of the eye is partly based on first hand experience from dissections. At the time the crystalline lens was known (the eye contains “water”), but the retina has not been noticed. Also, the refractive properties of the eye have not been studied. As pressing the eye results in flashes of light, it was generally believed that the eye contains “fire”.

For more information see (Beare 1906), for the lantern simile in DK 31 B84 or KRS 389 (O’Brien 1970).

### 7.4 Text 4: Anaxagoras of Clazomenae

- The fragments included are translated in (Kirk et al. 1983). The longer citation is from (Stratton 1917).

**KRS 467-9** All things were together, infinite in respect of both number and smallness; for the smallness too was infinite. And while all things were together, none of them were plain because of their smallness; for air and aither held all things in subjection, both of them being infinite; for these are the greatest ingredients in the mixture of all things, both in number and in size.

But before these things were separated off, while all things were together, there was not even any colour plain; for the mixture of all things prevented it, of the moist and the dry, the hot and the cold, the bright and the dark, since there was much earth in the mixture and seeds countless in number and in no respect like one another. For none of the other things either are like to one another. And since this is so, we must suppose that all things were in the whole.

The Greeks are wrong to recognize coming into being and perishing; for
nothing comes into being or perishes, but is rather compounded or dissolved from things that are. . . .

KRS 483 And since these things are so, we must suppose that there are many things of all sorts in everything that is being aggregated, seeds of all things with all sorts of shapes and colours and tastes . . . .

KRS 472 Neither is there a smallest part of what is small, but there is always a smaller (for it is impossible that what is should cease to be). Likewise there is always something larger than what is large. And it is equal in respect of number to what is small, each thing, in relation to itself, being both large and small.

KRS 500 The sun indues the moon with brightness.
KRS 501 We call the reflexion of the sun in the clouds a rainbow.
KRS 509-10 From the weakness of our senses we cannot judge the truth. Appearances are a glimpse of the obscure.

DS 27-29 Anaxagoras holds that sense perception comes to pass by means of opposites, for the like is unaffected by the like. He then essays to review each sense separately. Accordingly he maintains that seeing is due to the reflection in the pupil, but that nothing is reflected in what is of like hue, but only in what is of a different hue. Now with most <creatures> this contrast of hue <with that of the pupil> occurs by day, but with some by night, and this is why the latter are keen of vision by night. But, in general, night the rather is of the eye’s own hue. Furthermore, there is reflection by day, he holds, because the light is a contributing cause of reflection, and because the stronger of two colours is regularly reflected better in the weaker.

[. . .]

All sense perception, he holds, is fraught with pain—which would seem in keeping with his general principle, for the unlike when brought in contact <with our organs> always brings distress. This is illustrated by <our experience when an impression> long persists and when the exciting objects are present in excess. For dazzling colours and excessively loud sounds cause pain and we cannot long endure the same objects. The larger animals have more perfect powers of sense, and sense perception varies in general with the size <of the organs of sense>. For animals that have large clear lustrous eyes see large objects and such as are distant; while of animals with small eyes the opposite is true.

- **Hint to Problem 2** Both Empedocles and Anaxagoras pose a number of similar questions—how can vision and the difference between day and night (photopic and scotopic) vision be explained in
terms of the buildup of the eye. Empedocles also investigates the basis of perceiving different colours, while Anaxagoras discusses the difference of the vision of large- and small-eyed animals. The explanatory schemes operate with mechanistically feasible explanations, but are very different. Empedocles employs a concept of pores and the theory of four elements, while Anaxagoras places emphasis on the pupils.

In many contemporary accounts the “pupil” had great importance for vision. These fairly simple theories might seem blatantly false, and even counter to experience—but several phenomena were known that reinforced them. To state, for example, that the eye is made up of water and fire, one can point to the large amount of water that leaves the eye when crying, and draw attention to the little sparks (of fire) we see when our eyes are hit—this fire is even visible from the outside in the flashing light of the eyes of cats and other scotopic animals.

7.5 Text 5: Theophrastus on Empedocles

- The text of DS has been rediscovered in the Renaissance and appeared in print more than once during the sixteenth century. However, as it contained little of Theophrastus’ own views, it has been omitted from some seventeenth century editions of his works. All quotes from the DS are taken from the translation of the American psychologist G. M. Stratton (Stratton 1917).

DS 12-19 Such is Empedocles’ theory of the process both of sense perception and of thought. Yet from his account we might well be at a loss to know, first, wherein animate beings differ from other kinds of beings so far as sense perception is concerned; since particles fit into the minute passages in lifeless objects also. For universally he regards mixture as due to a correspondence with these passages. This explains why oil and water will not mix—in contrast to other fluids and to certain farther substances of which he recounts the peculiar combinations. Wherefore all things would perceive; and mixture, sense perception, and growth would be identical (for he ascribes them one and all to a correspondence with the passages), unless he add some farther difference.

In the second place, with regard even to animate things, why should the fire

---

3The English ‘pupil’ comes from the Latin ‘pupilla’ or ‘pupula’ meaning small doll or puppet. X looking into Y’s eye can see a tiny reflection of himself in Y’s eye; the appearance of this small figure was often taken to be equivalent of Y’s seeing X.
within the living creature perceive, rather than the fire without, if each really fits
into the other? for <on both sides> there is proportion and likeness. And further
there must be some difference between the two if the <fire within> is unable to
fill up the passages, while the <fire> entering from without <has this power>.
Consequently if <this internal fire> were absolutely and in every respect the
same <as the fire without>, there would be no perception. Furthermore, are
these passages empty or full? If empty, Empedocles is inconsistent; for he says
that there is absolutely no void. But if full, creatures would perceive perpetu-
ally; for it is evident that a <substance> similar <to another>—to use his own
expression—fits <into that other>.

And yet doubt might be felt upon the very point—whether it were possible
for diverse elements to be of precisely a size to fit each other; especially if it be
ture, as he says, that eyes with some disproportion in their mixture become
dim of sight by a clogging of their passages, now with fire and now with
air. Granting, however, that there is even here a nice adjustment, and that the
passages are filled by what is alien, yet how and where are these <occluding
particles> to be expelled when perception occurs? Some change must be as-
signed. Thus there is a difficulty in any case: for it is necessary to assume either
the existence of a void, or that creatures are uninterruptedly perceiving things;
or else that an alien substance can fit into <the sensory passages> without
causing perception and without involving the change peculiar to the substances
that do cause <perception>.

But, further, were we to suppose that what is like does not fit <the pas-
sages> but merely touches <there>, perception might reasonably arise from any
source whatever. For he attributes our recognition of things to two factors—
namely, to likeness and to contact; and so he uses the expression “to fit”. 
Accordingly if the smaller <particle> touched the larger ones, there would
be perception. And likeness also, speaking generally, is out of the question,
at least according to him, and commensurateness alone suffices. For he says
that substances fail to perceive one another because their passages are not
commensurate. But whether the emanation is like or unlike <the sensory organ>
he leaves quite undetermined. Consequently either perception is not dependent
on similarity; or else the failure to detect an object cannot be attributed to want
of spatial correspondence, and the senses without exception and all the objects
they perceive must have one and the same essential nature.

Moreover, his explanation of pleasure and pain is inconsistent, for he ascribes
pleasure to the action of similars, while pain he derives from opposites. For
these, he says, are “hostile”, since “ most distant they stand from one another
. . . in source and composition and in their moulded forms.”

Pleasure and pain thus are regarded by them as sense perceptions or as
accompaniments of sense perception; consequently <the perceptive process>
does not in every case arise from similarity. Again, if kindred things especially
cause pleasure by their contact, as he says, things which coalesce in their growth should have the keenest pleasure—and, in general, the keenest perception of one another, for he assigns the same causes for sense perception as for pleasure. And yet when we are perceiving, we often suffer pain in the very act of perception—indeed, Anaxagoras declares, we always do. For all perception, he says, is linked with pain.

A like difficulty appears in connection with the senses severally; for his position is that cognition is due to likeness. Now since, for him, the eye is composed of fire and of its opposite, it might well recognize white and black by means of what is like them; but how could it become conscious of gray and the other compound colours? For he assigns their perception neither to the minute passages of fire nor to those of water nor to others composed of both these elements together. Yet we see the compound colours no whit less than we do the simple.

Odd, too, is his account of the fact that certain animals see better by day, and others by night. For a weaker fire is extinguished by a stronger; and for this reason we find it impossible to gaze at the sun or at anything exceedingly bright. Accordingly animals with less of light in their eyes ought to have had poorer vision by day. Or if what is qualitatively similar does in fact supplement, and what is qualitatively different tends to destroy and thwart, as he says, then all creatures—both those that had less light and those that had more—should have seen white things better by day; and black, by night. Yet in fact all but a few animals see every manner of object better by day. And for these exceptional animals, we may reasonably suppose, the fire inherent in their eyes is peculiarly intense; just as some objects by their own colour glow brighter in the night.

In those cases, moreover, where the blend is in equal measure, each component would of necessity be supplemented in turn. And consequently if an excess of the one element prevented the other from seeing, all creatures would to all intents and purposes be in a like condition.

Although it is a fairly difficult task to explain the facts of vision, yet how could we by likeness discern the objects with which the other senses deal? For the word ‘likeness’ is quite vague. We do not <discern> sound by sound, nor smell by smell, nor other objects by what is kindred to them; but rather, we may say, by their opposites. To these objects it is necessary to offer the sense organ in a passive state. If we have a ringing in the ears, or a taste on the tongue, or a smell in the nostrils, these organs all become blunted; and the more so, the fuller they are of what is like them—unless there be a further distinction of these terms.

DS 23-24 One might likewise have serious misgivings over his doctrine of
thought, if Empedocles actually regards thought as having the same constitution as sense; for then all <creatures> would share in thought. And how can the notion be entertained that thinking arises in a process of change, and at the same time arises by the agency of the like? Since the like produces no change in the like. And it is indeed quite ridiculous to suppose that we think with the blood: for many animals are bloodless; and of those that have blood, the parts about the organs of sense are the most deficient in blood. Furthermore, according to his view, bone and hair ought to perceive, since they too are composed of all the elements. In all consistency, moreover, to think and to perceive and to enjoy would be identical processes; and, on the other hand, to suffer pain and to be ignorant—for these two he ascribes to unlikeness. Accordingly, pain ought to accompany ignorance; and pleasure, the act of thinking.

Again, his idea is odd that the special abilities of men are due to the composition of the blood in their particular members—as if the tongue were the cause of eloquence; or the hands, of craftsmanship; and as if these members did not have the rank of mere instruments! Indeed one might better for this reason assign the shape of the organ as the cause <of talent>; rather than ascribe this to the composition of the blood <in the organ>,—which really has nothing to do with understanding. For this is the case certainly with animals other <than man>. Empedocles thus seems to have gone astray at many a point.

7.6 Text 6: Theophrastus on Anaxagoras

- The following excerpts comprise only a part of Theophrastus’ critique of Anaxagoras in the De Sensibus (Stratton 1917).

DS 36-37 Anaxagoras’ doctrine of the visual image is one somewhat commonly held; for nearly everyone assumes that seeing is occasioned by the reflection in the eyes. They took no account of the fact, however, that the size of objects seen is incommensurate with the size of their reflection; and that it is impossible to have many contrasting objects reflected at the same time; and, farther, that motion, distance, and size are visual objects and yet produce no image. And with some animals nothing whatever is reflected—for example, with those that have horny eyes, or that live in the water. Moreover according to this theory many lifeless things would possess the power of sight; for there is a reflection certainly in water, in bronze, and in many other things.

His own statement is that colours are reflected in one another, but particularly the strong in the weak; consequently each of these—but especially black and the weaker colours generally—should possess the power of sight. For the reason just given, he holds that the organ of vision is of the same colour as the night, and that light is the cause of the visual reflection. But in the first
place, we see light itself, without any image of it whatsoever; and in the second
place, black objects and white objects alike lack light. And furthermore in other
cases we are all the while seeing reflection arise in what is more brilliant and
pure—a fact entirely in keeping with his own statement that the membranes
of the eyes are fine and lustrous. Now most <scientists> assume that the organ
of vision itself is of fire, since colours partake of this element especially. And
Anaxagoras himself, as I have said, upholds this rather common and hoary
doctrine; save that in the case of each and every sense he offers something
original, and particularly of sight when he sets forth the part which size here
plays in perception. But of the senses that have a more material character he
offers no such clear account.

• **Hint to Problem 3** According to (Stratton 1917), Theophrastus
investigated the theories as to how they treated the mutual relation
between sense-organs and sense-objects, whether their implications
were contrary to fact, if their basic premisses were in themselves
contradictory, or whether they would have left obvious facts unex-
plained. For Theophrastus causal relations and simple cooccurrances
or concommitances were separated, and he also kept the economy
and elegance of a theory in high esteem when evaluating them.

A more recent critic highlights two modes of attack (Baltussen 2000),
one focusing on the consistency of the Presocratic arguments, the
other on the general applicability of the early theories. In both cases
the consequences and implications of the theories are investigated,
but here Theophrastus leaves the faithful presentation behind and
applies Peripatetic (i.e. unhistorical) criteria.

There are several differences between his attack on Empedocles and
the one on Anaxagoras. In the first case the critique focuses on the
description of the physiological process leading to sensation. The
general account of perception—effluences fitting into the passages—
turn out to be plagued by difficulties on closer scrutiny. “Fitting into”
is not a *sufficient* explanation or description of perception (DS 12),
and the account entails either that the pores are at times empty (con-
tradictory to Empedocles’s general philosophy), or that perception
occurs at all times (contradicting generally known facts). Both horns
of the dilemma (DS 13) are unacceptable. The problem of the kind
and the size of particles is treated in DS 14, and possibilities of how
they might contact the sense organ in DS 15. Theophrastus raises problems with respect to colours (DS 17), but does not refute the theory. He gives a rival account concerning the intensity and capacity of the senses (DS 18), and in DS 19 he basically states that the starting point “like perceives like” is false. He argues against equating perception and thinking (DS 23), but this is probably a misinterpretation of Empedocles—a conclusion earlier stated by Aristotle.

In case of Anaxagoras, too, the critique exceeds (DS 31-37) the summary (DS 27-30) in length. It focuses mainly on the doctrine that sensation causes pain, and that seeing is caused by a reflection on the pupil. Here only a part of the critique is reproduced, and the critique shows many elements alien to Anaxagoras’ thought.

8 Further Problems

- **Problem 4** Based on the material here and some other sources from the suggested reading, try to present a summary of Presocratic theories on vision.

- **Problem 5** Discuss the statement by Robert Crone (Crone 1999, p. 4):

  Empedocles believes in both extramission and intromission. He is convinced that the eyes shoot fire. The eye can be compared to a lantern: the ocular fire takes the place of the burning oil and the transparent pupil functions as the window of transparent horn. On account of this comparison Empedocles can be regarded as an extramissionist. But on the other hand he also declares that objects produce an emanation. There are pores in the eye, of exactly the right size and shape, which not only allow the ocular fire to leave the eye but also admit the emanations arising from outside objects.

- **Problem 6** Place Empedocles’ theory of perception in the broader context of his philosophy.

9 Suggested Readings

Compared to the amount of text written on the Presocratics, one is surprised to see how meagre the sources themselves are. The most important
collections of fragments is (Diels & Kranz 1922), for an English translation (Fitt & Diels 1962) can be consulted.

For a general introduction see the classic (Kirk & Raven 1957) or (Kirk et al. 1983), and the first three volumes of (Guthrie 1962). For a shorter selection of translations see (Waterfield 2000). The philosophical arguments are investigated in (Barnes 1979). For a first superficial glance an increasing number of internet-sites can be used, like The Internet Encyclopedia of Philosophy (http://www.utm.edu/research/iep/), the Stanford Encyclopedia of Philosophy (http://plato.stanford.edu/), etc.

On the fragments of Heraclitus see (Kahn 1979), for an earlier bibliography (Roussos 1971). On his epistemological views see (Wilcox 1994). For a more general introduction (Geldard 2000).

For the text of Parmenides see (Tarán 1965) and (Sider & Johnstone 1986). For a general introduction to his thought and to Eleatic Monism see (Curd 1998).

The views on Empedocles are expected to change significantly in the coming years, as a new fragment of Empedocles had been published a few years ago (Martin & Primavesi 1999). For a treatment of the fragments see (Wright 1995), and (Kingsley 1995) for a highly stimulating understanding of Empedocles in the context of Sicilian cults. O’Brian gives a reconstruction of Empedocles’ cosmic cycle in (O’Brien 1969), and his (O’Brien 1970) is also important to understand the further development of Empedocles’ famous lantern-simile. See also Laks’s article in (Long 1999) and the earlier (Long 1966).

For Anaxagoras (Cleve 1973) and (Schofield 1980) are good starting points. For his scientific views see (Teodorsson 1982) and (Gershenson & Greenberg 1964).

On Theophrastus’s life see (Sharples 1998), on his work on sense perception (Priscian, Huby, Steel, Lautner, Urmsn & Simplicius 1997). A recent work on the De sensibus is (Baltussen 2000).
1 INTRODUCTION

Plato was probably born in Athens in 428 B.C., into a family of aristocrats with many connections to the ruling oligarchs of the town. He spent much time in the circles of Socrates, whose life and death greatly influenced him. Plato left Athens in 399 B.C. and traveled widely in Italy and Sicily. His exact itinerary is not known. It is reported that Plato went to Megara and visited Cyrene. He left to Italy to study with Pythagoreans (possibly with characters mentioned in his dialogue Phaedo) whose teachings resurface in many of his dialogues. Among others he met the ruler of Syracuse, Dionysius I, and the mathematician Archytas of Tarentum. He also visited Egypt.

Plato established his school the Academy after his return to Athens in 387 B.C., and spent much of his remaining life until 347 B.C. teaching here. His influence on the development of philosophy cannot be overestimated.

When citing his works, generally the title of the dialogue is given, followed by a number and a letter. In case of the Republic and the Laws usually the book number in Latin numbers is also included. The number is the page number of the three-volume edition of Plato’s works printed in 1578 by Henri Estienne (1528-1598), also known as the Stephanus edition. In this edition, the Greek version appears in the right column and the Latin translation by Jean de Serres in the the left. Each page was subdivided into smaller sections marked by letters from a to e—hence the non-numeric part of the Stephanus numbers. As the name of the dialogue identifies the volume, the volume number is not given.

This chapter includes some of Plato’s texts on light, colours and a theory of mirrors, and two short sections connected only loosely to Plato’s work—one on the difficulty of working with Greek colour terms, the other on some aspects of the Pythagorean tradition.
The great teleological cosmology of the *Timaeus* was one of the most important scientific texts throughout the Middle Ages. Introductions to the main topics of this work are readily available, and will not be discussed here.

The *Timaeus* is the best place to look for an articulated Platonic theory of vision, colours, and mirroring—in other works of the corpus only scattered sentences are found (*Philebus* 12E, *Meno* 74C-77A, *Phaedrus* 110B-E, *Critias* 116AB, *Theaetetus* 156DE, *Lysis* 217C-E, *Republic* 507C-508D). Although Plato often mentions colours, these utterances are relatively unsystematic and seem scarce when we take into account that in his youth he probably studied painting. Therefore these texts are not included (except for the quotation in Problem 9), and they are usually fairly accessible, so collecting them calls for little effort.

In the account given in the *Timaeus* (45A-46C, 67C-68D) seeing is described as requiring the coalescence of the mild fire from within the eye and the outer fire. This process takes place outside the body, but the description mixes intromissionist and extromissionist elements, it is a hybrid theory of vision. This unified stream of fire functions as an extended organ. As "like is known by like", when there is no light outside, the inner fire is quenched and no vision takes place. The mechanical effects of particles—all geometrically shaped—are highly important in this theory as well (see also Chapter V). The theory also explains how images are formed in mirrors. This description, however rudimentary, is among the earliest theories of mirrors (see more in Chapter IV, V and VI).

**Problem 1** Image formation is explained by Plato with recourse to the coalescence of the inner and outer fire in Text 1 (46). Summarise his theory of mirror image formation. Explain how side inversion takes place in normal mirrors, and why it does not in concave mirrors.

In Text 2, after investigating taste, odour, and sound, Plato discusses colours. Here particle sizes are of central importance.
Problem 2 Summarise the perception of basic colours in Plato’s account in Text 2. Devise an algorithm (using only dichotomies) which correlates the properties of the particles and the colours perceived.

Problem 3 Collect in a table the mixed colours listed in Text 2. What types of colours can be mixed to give similar results? Do you consider any of the mixtures unusual or unrealistic?

3 Greek Colour Terms

Reading the list of primary colours in Plato’s *Timaeus* might surprise us: how can “brilliant” or “bright” be considered a primary colour, or even a colour? Is this the result of an isolated instance of intranslatability, where no reasonable equivalent can be found for a Greek word in English? It might be so, but Greek colour terms are rather too often similarly surprising. A few examples are given below.

The term *chlôros* is customarily translated as green (as in chlorophyll). It is used in this sense in the *Odyssey* (Od. 9.320) to describe green wood, and in the *Iliad* as the colour of honey. But, surprisingly in Euripides’a drama *Medea* (line 907) tears are *chlôros*, and in his *Hecube* the sacrificed blood of Polyxena is *chlôros*.

The term *kuaneos* is used to signify the strikingly blue *lapis lazuli*, or the blue copper carbonate (the term “cyan” is a derivative) in Theophrastus, but, surprisingly, it is never used to describe the colour of the sea in the Homeric poems. Many epithets are used in these hymns for the sea: white (Od. 10.94), wine-dark (Od. 2.421, etc.), purple (Il. 16.391). Similarly the sky can be bronze (Od. 3.2) or iron (15.328) coloured, but the term that we are accustomed to translate as blue is missing here.

Where the word appears in the *Iliad* is equally surprising: the hair of Hector is *kuaneos* (I. 24.401-03), and there is no darker garment than the veil of Thetis (I. 93-94), and this, too, is *kuaneos*.

This unusual usage of the word *kuaneos* was first recognised by J. W. von Goethe in the early nineteenth century (Goethe 1989b). Since then this and similar phenomena have attracted much attention. These colour terms are

---

1 The use of the two terms *chlôros* and *kuaneos* are analysed in detail in the second and third chapter of (Irwin 1974). Although concentrating on literature before the fourth century B.C. the book also discusses changes and developments in terminology from Homeric hymns to the works of Theophrastus.
used for such an array of referents that it seems to be an impossible task to find terms in English (or in a number of other languages) that faithfully map the Greek usage. Rather than being isolated cases of intranslatability, the group of properties that are mapped by many of the Greek colour terms seems to be significantly different from the referents of colour terms in most modern languages.

Several explanations have been offered for these differences. Nearing the end of the nineteenth century a number of books had been published that argued for the quick evolution of the human visual system. According to this view, the difference of usage implied a difference in the nervous system—an opinion once in favour but now rejected².

Extensive anthropological work with less known languages in the twentieth century had provided a mass of information on colour terms, and in this light Greek colour categories are much less surprising. Although for some time it seemed that colour terms are arbitrary, language-dependent, and this was taken to be a support for cultural relativity, a large-scale study by Brent Berlin and Paul Kay in 1969 had rather surprising conclusions (Berlin & Kay 1969).

The result of the study was that the number of so called basic colour terms (BCT) varies from two (as in the Papuan Dani language) to around eleven. This does not imply that a language has only two to eleven colour terms. The number of these may be well into the hundreds, but they are either not used generally or not known and used by the majority speakers, or are not monolexemic, etc., that is do not qualify as BCTs. For the precise original definition see (Berlin & Kay 1969). Later the writers themselves offered somewhat more vague definitions. Their original study showed that if the languages were ordered according to the number of basic colour terms, the encoding was surprisingly strictly constrained.

There was, it seemed, a restricted universal inventory of basic colour terms. A language acquires basic colour terms in a constrained order. A language with two BCTs distinguishes between black and white. If a third BCT appears, it encodes red. A fourth one encodes either green or yellow. This hypothesis postulates a hierarchy of colour terms, where the presence of any term implies that all terms to its left exist in the given language:

²The “blue-anomaly” of Greek has been recorded in a number of other languages, and explained in evolutionary terms, see (Gladstone 1858, Geiger 1872). For a bibliography of these debates see (Skard 1946), for a recent overview e.g. (Sölch 1998, Irwin 1974).
Black < Red < Green < Blue < Brown < Purple, Pink
Orange, Grey

Could this be the clue to the strange Greek colour terminology? Is it possible that Ancient Greek was at a different level of development (conflating yellow and green, with no BCT for blue, for example)?

Several methods have been developed to establish the BCTs of a language and exclude the non-basic terms, to separate the first six colours, the so called primary basic colours (white, black, red, green, yellow, and blue) from the other, secondary basic colours, and to find regularities in the ordering of BCTs. Measurements of reaction-time, naming experiments, consistency of use, order of occurrence in elicited lists, and frequency of occurrence have all been employed in experiments.

Of course many of these cannot be used in an inquiry about Ancient Greek colour terms, but in a study the occurrence of colour terms in word databases have been found to correlate well with the hierarchy of colour terms (Corbett & Davies 1997). While not always separating BCTs from non-basic terms clearly, it works well in distinguishing between primary and secondary basics.

- **Problem 4** Plato lists four colours as primary, and a number of other, secondary colours (his usage of these terms is obviously not that of Berlin and Kay’s). Were the terms he recognised as primary significantly more often used in Ancient Greek than the secondary terms (and therefore candidates for being BCTs)? Can significant differences in frequency be observed? Investigate the frequency of the Greek colour terms (leukos, melas, lampros, eruthros, xanthos, halourgês, arphinos, áchros, kuaneos, glaukos, purros, phaios mentioned in Text 3, Tim. 68C).

Some help: first a significantly large database is needed, which is readily provided by e.g. the Tufts Perseus Project. The database is regularly extended, and it contains several million words. The English and Greek texts of the *Timaeus* are also accessible at the following the links: http://www.perseus.tufts.edu/cgi-bin/ptext?lookup=Plat.+Tim.+68b and http://www.perseus.tufts.edu/cgi-bin/ptext?-lookup=Plat.+Tim.+68c. Words in the Greek version of the text are directly linked to the corpus and the frequency-charts.
You can view both the English and the transliterated Greek texts. In the Greek every word is tagged and linked with a dictionary entry. This entry, apart from the definition, gives the occurrence of the term in the database, together with the frequency of the word (occurrence per ten thousand words). Collect and compare the data for the colour terms listed in the *Timaeus 68*.

While useful for explaining the small number of frequently used terms, the existence of words that conflate categories (described by separate BCTs in other languages) remains problematic. Other approaches to answer the enigma of the strange distribution of Greek colour terms might prove more fruitful. One such approach is exemplified next.

Harold C. Conklin studying the inhabitants of the small, southern Philippine island of Mindoro in the 1950’s investigated the Hanunóo colour categories (Conklin 1998). His often cited and reprinted study, instead of trying to match a predetermined set of coloured chips with colour terms investigated the actual use of colour categories.

Finding four terms that could be roughly equated with the colour terms “black”, “white”, “red” and “light green”, he noted that the terms might “have certain correlates beyond what is considered the range of chromatic differentiation” (Conklin 1955). He interpreted his findings claiming that, apart from signifying colours, “[f]irst, there is an opposition between light and dark . . . Second there is an opposition between dryness and desiccation and wetness or freshness (succulence) in visible components of the natural environment . . . ”

There seems to be no reason why a set of words should necessarily refer to a specific characteristics that we term hue. This and similar studies suggest that some languages might have terms that are used to characterise not only the colour of surrounding objects but other surface characteristics as well.

But this is not all. With time significant shifts can take place in a language. In English, for example, some studies suggest, “colour terms” of the Anglo-Saxon or Old English period signified predominantly brightness differences, and hue was hardly conceptualised until the Middle English Period (1150-1500). At present, as a result of a gradual semantic shift, English colour terms are almost exclusively hue concepts (Casson 1997).
The seemingly inconsistent use of Greek colour terms therefore might simply reflect that these referred to surface characteristics other than simply hue, and the text studied could represent a specific transitory period.

- **Problem 5** After having answered Problem 4, discuss the difficulties that you encountered while trying to answer the question. What serious methodological flaws can accompany this type of work?

4 **The Pythagorean Influence**

During his travels Plato had probably been influenced by Pythagoreans, followers of the philosopher-sage who flourished around 530 B.C. A short introduction is therefore desirable to some of the doctrines of this school (or sect), as they continued to exert profound influence not only on Plato but on later Platonists, Neo-Platonists, and on several other scientists well into the scientific revolution.

Hardly anything is known about Pythagoras’s life. He left his birthplace Samos in about 520 B.C., settled in Croton, and around 500 B.C. moved to Metapontum. The sect he established had numerous regulations for daily life—many of them dietary—, aimed at the purification of the soul, to escape the cycle of reincarnation to which humans are condemned.

One of the great achievements attributed to Pythagoras was the discovery of the correspondence between musical consonances and simple numerical ratios. If the length of a string is reduced to its half its vibration will produce a sound that is an octave higher than the original. And as the ratio 2 to 1 stands for an octave, similarly 3 to 2 corresponds to a perfect fifth, 4 to 3 to a perfect fourth. By multiplying the ratios corresponding to an octave with the ratio corresponding to a fifth yields 3 to 1, a perfect twelfth. These and other ratios (like 4 to 1 standing for a double octave) highlighted a peculiar correspondence between the most simple ratios between numbers and the most pleasant consonances. These recognitions lead to an attempt at interpreting the entire physical world in terms of numbers.

The study of the harmonic ratios—also called harmonics—continued to be an important subject through the Middle Ages until the seventeenth century. With astronomy, arithmetics, and geometry, it was one of the mathematical sciences dealing with numbers and magnitude. These together were often referred to as the *quadrivium*, an important part of the Medieval university curricula. Combined with the *trivium*—grammar, rhetoric,
Figure III.1: Even in Newton’s *Opticks* we see the influence of the Pythagorean search for ratios in the diagram of the spectrum: 1, 8/9, 5/6, 3/4, 2/3, 3/5, 9/16 (Newton 1730). The ratios in his view not only correspond to musical notes, but also to the boundaries of bands of colours seen in the spectrum (see page 280).

...and dialectic—these subjects constituted the *seven liberal arts* (septem artes liberales)*. 

Apart from astronomy and harmonics only very few of the ancient sciences were mathematisable. But one of these was optics—as we will see in Chapter VI in detail. The connection between these disciplines was significant, as terms like “the music of the spheres”, or “colour harmonies” would suggest. That the simple ratios discovered in music also determine ratios between colours was widely accepted. It is in this light that the significance of Plato’s sceptic stance in the *Timaeus* (68D) can be appreciated: man is not wise and powerful enough to find out the exact ratios of colour-mixing.

### 5 Texts

#### 5.1 Text 1: *Timaeus* 45A-46C

- To help answering the problems of the chapter, the translations are from the Tufts Perseus Project (http://www.perseus.tufts.edu/).

45a Wherefore, dealing first with the vessel of the head, they set the face in the front thereof and bound within it organs for all the forethought of the Soul; and they ordained that this, which is the natural front, should be the leading part. And of the organs they constructed first light-bearing eyes, and these they fixed in the face for the reason following. They contrived that all such fire as

3The divergences from the above, simplified classification are significant. As unusual sciences as architecture, medicine, philosophy or even necromantia (a definitely magical practice with divinatory aims) had been included in the seven liberal arts (Láng 2003).
had the property not of burning but of giving a mild light should form a body akin to the light of every day⁴. For they caused the pure fire within us, which is akin to that of day, to flow through the eyes in a smooth and dense stream; and they compressed the whole substance, and especially the center, of the eyes, so that they occluded all other fire that was coarser and allowed only this pure kind of fire to filter through. So whenever the stream of vision is surrounded by midday light, it flows out like unto like, and coalescing therewith it forms one kindred substance along the path of the eyes’ vision, wheresoever the fire which streams from within collides with an obstructing object without. And this substance, having all become similar in its properties because of its similar nature, distributes the motions of every object it touches, or whereby it is touched, throughout all the body even unto the Soul, and brings about that sensation which we now term “seeing”. But when the kindred fire vanishes into night, the inner fire is cut off; for when it issues forth into what is dissimilar it becomes altered in itself and is quenched, seeing that it is no longer of like nature with the adjoining air, since that air is devoid of fire. Wherefore it leaves off seeing, and becomes also an inducement to sleep. For the eyelids—whose structure the Gods devised as a safeguard for the vision,—when they are shut close, curb the power of the inner fire; which power dissipates and allays the inward motions, and upon their allaying quiet ensues; and when this quiet has become intense there falls upon us a sleep that is well-nigh dreamless; but when some greater motions are still left behind, according to their nature and the positions they occupy such and so great are the images they produce, which images are copied within and are remembered by the sleepers when they awake out of the dream. And it is no longer difficult to perceive the truth about the formation of images in mirrors and in bright and smooth surfaces of every kind. It is from the combination with each other of the inner and the outer fires, every time that they unite on the smooth surface and are variously deflected, that all such reflections necessarily result, owing to the fire of the reflected face coalescing with the fire of the vision on the smooth and bright surface. And left appears as right, because contact takes place between opposite portions of the visual stream and opposite portions of the object, contrary to the regular mode of collision. Contrariwise, right appears as right and left as left whenever the fire changes sides on coalescing with the object wherewith it coalesces; and this occurs whenever the smooth surface of the mirrors, being elevated on this

⁴There is a play here on the words hêmeron ("mild")... hêmeras ("day"); Cf. Cratyl. 418C.
side and on that, repels the right portion of the visual stream to the left and the left to the right. And when this same mirror is turned lengthwise to the face it makes the whole face appear upside down, since it repels the bottom of the ray to the top, and conversely the top to the bottom. Now all these are among the auxiliary Causes which God employs as his ministers in perfecting, so far as possible . . .

5.2 Text 2: *Timaeus* 67c-68D

- The text is also from the Tufts Perseus Project (http://www.perseus.tufts.edu/).

67c . . . We have still remaining a fourth kind of sensation, which we must divide up seeing that it embraces numerous varieties, which, as a whole, we call “colors”. This consists of a flame which issues from the several bodies, and possesses particles so proportioned to the visual stream as to produce sensation; and as regards the visual stream, we have already stated\(^5\) merely the causes which produced it. Concerning colors, then, the following explanation will be the most probable and worthy of a judicious account. Of the particles which fly off from the rest and strike into the visual stream some are smaller, some larger, and some equal to the particles of the stream itself; those, then, that are equal are imperceptible, and we term them “transparent”; while the larger and smaller particles—of which the one kind contracts, the other dilates the visual stream—are akin to the particles of heat and cold which affect the flesh, and to the astringent particles which affect the tongue, and to all the heating particles which we call “bitter”\(^6\) with these “white” and “black” are really identical affections, occurring in a separate class of sensation, although they appear different for the causes stated. These, therefore, are the names we must assign to them: that which dilates the visual stream is “white” and the opposite thereof “black”\(^7\); and the more rapid motion, being that of a different species of fire, which strikes upon the visual stream and dilates it as far as to the eyes, and penetrating and dissolving the very passages of the eyes causes a volume of fire and water to pour from them, which we call “tears”. And this moving body, being itself fire, meets fire from the opposite direction; and as the one firestream is leaping out like a flash, and the other passing in and being quenched in the moisture, in the resultant mixture colors of all kinds are

\(^5\)Cf. 45 C ff.
\(^6\)Cf. 65 E.
\(^7\)Cf. 45 C ff.
produced. This sensation we term “dazzling” and the object which causes it “bright” or “brilliant”. Again, when the kind of fire which is midway between these reaches to the liquid of the eyes and is mingled therewith, it is not brilliant but, owing to the blending of the fire’s ray through the moisture, it gives off a sanguine color, and we give it the name of “red”. And “bright” color when blended with red and white becomes “yellow”. But in what proportions the colors are blended it were foolish to declare, even if one knew, seeing that in such matters one could not properly adduce any necessary ground or probable reason. Red blended with black and white makes “purple”; but when these colors are mixed and more completely burned, and black is blended therewith, the result is “violet”. “Chestnut” comes from the blending of yellow and grey; and “grey” from white and black; and “ochre” from white mixed with yellow. And when white is combined with “bright” and is steeped in deep black it turns into a “dark blue” color; and dark blue mixed with white becomes “light blue”; and chestnut with black becomes “green”. As to the rest, it is fairly clear from these examples what are the mixtures with which we ought to identify them if we would preserve probability in our account. But should any inquirer make an experimental test of these facts, he would evince his ignorance of the difference between man’s nature and Gods—how that, whereas God is sufficiently wise and powerful to blend the many into one and to dissolve again the one into many, there exists not now, nor ever will exist hereafter, a child of man sufficient for either of these tasks.

---

**Figure III.2: Vision according to Plato’s *Timaeus* from (Platon 1992, p. 295).**

---

8I.e., between the kinds of fire which produce “blackness” and “brightness”.
• **Hint to Problem 1** Figure III.2 and III.3 show a contemporary reconstruction by Luc Brisson (Platon 1992).

![Diagram of mirror reflection](image)

Figure III.3: Mirroring in a plane mirror and image formation according to Plato’s *Timaeus* from (Platon 1992, p. 295).

• **Hint to Problem 2** The “flame” meeting the fire from the eye consists of particles that differ in size. Particles that have (i) the same size as those of the inner fire are imperceptible, “transparent”. Contrary to Plato’s earlier account like does not perceive the like here. The particles that have different sizes (ii) can either contract or dilate the visual stream. Larger particles contract the stream and give rise to the perception of black. Smaller particles (iii) either give rise to the perception of white, or—even smaller particles—dilate the visual stream and penetrate the eyes causing tears. If these small particles mixing with fire and water (iv) are very small we perceive “dazzling” or “brilliant” colours, if the particles are somewhat larger we perceive red.

In this account the description can be restructured to follow strict dichotomies, as noted by (Gaiser 1965), a typical hallmark of Platonic classification.

• **Hint to Problem 3** Plato gives a rather detailed list of mixed colours from the primaries white (A), black (B), red (C), and brilliant (D). These—following the notation in (Platon 1992) are:
Yellow($E$) = $A + C + D$
Purple($F$) = $A + B + C$
Violet($G$) = $(A + B + C) + B = A + 2B + C$
Chestnut($H$) = $E + I = 2A + B + C + D$
Grey($I$) = $A + B$
Ochre($J$) = $E + A = 2A + C + D$
Lapis(darkblue)($K$) = $A + B + D$
Lightblue($L$) = $A + K = 2A + B + D$
Green($M$) = $H + B = 2A + 2B + C + D$

In spite of this, and a number of other texts, some art historians, like John Gage, argue that colour-mixing was probably unusual in Ancient Greece, as we have neither surviving palettes or other practical devices for colour-mixing by painters, nor detailed and coherent colour-mixing schemes (Gage 1993).

**Hint to Problem 4** At the time of the writing of this book the frequency differences were more than hundred-fold. A sharp divide was seen between three of Plato’s “primary” colours—frequency more than 1 (in 10 000)—and the other terms—frequency less than 0.3. Three of these terms had a frequency value above 0.1—and these terms, together with the three “primary” terms correlate fairly well with primary BCTs. A “mixed colour term” for Plato, however, had higher frequency than one of the “primary” colours, not only in the overall corpus, but in Plato’s works as well. In general, the hierarchies of the terms in Plato’s account and in the list of frequencies correlate well.

**Hint to Problem 5** While Plato’s “primary” colours are generally more frequent than the “secondary” colours, they are not necessarily the most frequent colour terms. As any English-Greek dictionary testifies, often dozens of Greek colour terms can be equated with an English BCT. Any of these is a candidate for being a Greek BCT (at least at some point in time)—all frequencies should therefore be calculated. Also, the corpus—even if it can be considered as representative for our purposes—consists of Greek texts that have been produced over several centuries in different locations: to use this sort
of database the possibilities of significantly different local dialects have to be excluded, and we have to postulate that within this period no significant change in the structure of Greek colour terms had taken place. This latter assumption is downright false.

Another serious source of failure is to locate frequent terms—and treat them as BCTs—when these cannot be simply considered “colour terms”, but represent other surface characteristics (to give a few English examples: “light”, “blond”, “tan”, etc.). This highlights the problematic nature of finding the boundaries between BCTs, other colour terms, and non-colour terms.

6 Further Problems

• **Problem 6** Of course it is an exciting and worthwhile task to study the Greek colour terms in more detail, and also the relevant new literature from cognitive science. Connecting the two areas might seem a daunting but fruitful task. For some guidelines see the suggested literature below.

• **Problem 7** We have not really analysed in detail Plato’s scattered remarks on colours in *Philebus* 12E, *Meno* 74C-77A, *Phaedrus* 110B-E, *Critias* 116AB, *Theaetetus* 156DE, *Lysis* 217C-E, *Republic* 507C-508D. Sorting these out and thoroughly investigating them is an ideally suited task for students with some knowledge of Ancient Greek.

• **Problem 8** The mapping of the ongoing debate on the ancient colour of the star Sirius is also relevant to several of the questions discussed in this chapter. Contemporary astronomers debate whether to use our present-day knowledge about the colour of the star to determine what the Greeks saw, or to use (an interpretation of) Ancient Greek sources to inform modern science about the earlier colour of the star. For overviews of the literature and some of the current positions see (Bonnet-Bidaud & Gry 1991), or (Bonnet-Bidaud, Colas & Lecacheux 2000), together with (Whittet 1999) and (Chapman-Rietschi 1995).

• **Problem 9** Comment on the following citation from Plato’s *Republic* (trans. by Paul Shorey):
“Well, look at it thus. Do hearing and voice stand in need of another medium so that the one may hear and the other be heard, in the absence of which third element the one will not hear and the other not be heard?” “They need nothing,” he said. “Neither, I fancy,” said I, “do many others, not to say that none require anything of the sort. Or do you know of any?” “Not I,” he said. “But do you not observe that vision and the visible do have this further need?” “How?” “Though vision may be in the eyes and its possessor may try to use it, and though color be present, yet without the presence of a third thing specifically and naturally adapted to this purpose, you are aware that vision will see nothing and the colors will remain invisible”. “What is this thing of which you speak?” he said. “The thing,” I said, “that you call light”. “You say truly,” he replied. “The bond, then, that yokes together visibility and the faculty of sight is more precious by no slight form that which unites the other pairs, if light is not without honor”. “It surely is far from being so,” he said. “Which one can you name of the divinities in heaven as the author and cause of this, whose light makes our vision see best and visible things to be seen?” “Why, the one that you too and other people mean,” he said; “for your question evidently refers to the sun”. “Is not this, then, the relation of vision to that divinity?” “What?” “Neither vision itself nor its vehicle, which we call the eye, is identical with the sun”. “Why, no”. “But it is, I think, the most sunlike of all the instruments of sense”. “By far the most”. “And does it not receive the power which it possesses as an influx, as it were, dispensed from the sun?” “Certainly”. “Is it not also true that the sun is not vision, yet as being the cause thereof is beheld by vision itself?” “That is so,” he said. “This, then, you must understand that I meant by the offspring of the good which the good begot to stand in a proportion with itself: as the good is in the intelligible region to reason and the objects of reason, so is this in the visible world to vision and the objects of vision”.

7 Suggested Readings

The literature on Plato is immense, the selection here is only a somewhat arbitrary cross-section. For further studies one of the critical editions should be consulted. These often contain bibliographies. Also useful as a starting point (Kraut 1992), the classic (Cornford 1987), or other modern compendia for Plato, together with simpler, one volume editions, like (Plato 1961).
On Plato’s works concerning the sciences (Anton 1980), on his philosophy of science (Gregory 2000) on causation (Cheyne 2001) can be consulted. For Plato’s views on sensation see (Silverman 1990).

More specifically on the *Timaeus* one can consult (Ashbaugh 1988, Platon 1992) or the selection of essays in (Wright & Barker 2000). For a reprint of Taylor’s classical 1928 commentary—also reprinted in 1967—see (Taylor 1987). The four element-theory in the work is investigated in (Black 2000). The ideas on sensation are investigated from an unusual perspective in (O’Brien 1984). Aristotle’s criticism of the *Timaeus* is analysed in (Claghorn 1954).

The “evolutionary” development of colour terms has received much support in recent decades, but also serious criticism—see e.g. (Lucy 1997) highlighting the problems from a linguistic point of view, and (Saunders & van Brakel 1997) as a large scale-review of the problem areas, with instructive peer opinions. On colour terms as semantic universals see (van Brakel 1994), on BCTs (Saunders 2000).

For the Early Greek colour terminology you can consult one of the large-scale reviews of single colour terms such as (Maxwell-Stuart 1981a), (Maxwell-Stuart 1981b), or (Stulz 1990), or shorter articles, like (Moonwomon 1994). For a detailed study and a review of the earlier literature consult (Irwin 1974), for the study of coloured drawings (Dürbeck 1977).

It seems that in certain periods languages have quickly expanding colour vocabulary. Examples are cited in (Gage 1993), or in special articles, like (Smith & et al. 1997). In recent years the universality of colour terms has again been doubted—see (Roberson, Davies & Davidoff 2000). On colour language and categorization see (Dedrick 1998).

On sources of colour science a collection is to be found in (MacAdam 1970), but it contains only translations that are—somewhat willfully—edited. For general introductions see (Gage 1993), or the similar but slimmer (Gage 1999) are excellent, but see also (Zollinger 1999), or (Zajonc 1993) for an enjoyable read, and (Thompson 1995) for a more cognitive science-based perspective.

For more on Pythagoras see for example the robust volume of (Burkert 1962, Burkert 1972).
Chapter IV

ARISTOTLE AND THE ARISTOTELIAN CORPUS

1 INTRODUCTION

Aristotle was probably born in 384 B.C. in Northern Greece, in the small town of Stagira. He became a pupil of Plato in 367, and after the death of his master in 347 he left Athens. Having spent some time in Asia Minor he also lived on the island of Lesbos for years.

From the huge Aristotelian corpus (probably around 6000 modern pages) only about a third survives, and this focuses on philosophical and scientific topics (Barnes 1995).

As in the case of many other classical authors, similarly for Aristotle, there exists a “standard edition”. The Greek text edited by Immanuel Bekker and first printed in 1831 is usually referred to by a number (the page number), the letter a or b (standing for the first or second column on the page), and a second number, the line number of the page. This Bekker number is often accompanied by a standard abbreviation of the treatise in question, and the book and chapter number.

This chapter contains various texts that illuminate Aristotle’s views on light, vision, and colours. Their dating is difficult, as most of the surviving texts show signs of revision and updating. It is possible that they were works in progress or lecture-notes, most probably not written for the broad public in the form they have survived, hence their general labeling as the esoteric, or “inner” works. These works are often difficult to understand or interpret. Ironically, apart from a few fragments, none of the “outer” or exoteric works have survived—even though these were considered as written with good style (Barnes 1995, p. 12).

In spite of the possible later amendments, Text 1, the third book of the Meteorology is probably the earliest. Discussing the senses and the sensibles, Text 2 and 3 were written somewhat later, while Text 4 is very likely to have been written by one of the members of Aristotle’s school, possibly after the death of Aristotle.
2 Atmospheric Optics

The *Meteorology* of Aristotle is one of the most detailed ancient discussions of atmospheric phenomena, meteorology, and a number of issues from geology and astronomy\(^1\). The fourth book of the treatise can be considered as one of the earliest textbooks on chemistry. It might not be authentic, and its origin and inclusion in the treatise is obscure, but it is important to note that while the upper bound for meteorological phenomena was clear (the moon's orbit) the lower limit remained vague, and often terrestrial phenomena were included—now treated under geology or inorganic chemistry. For more about this chemical treatise see (Düring 1980).

Written around 350 B.C. the work discusses the structure of the physical world below the domain of heavenly spheres, the sublunary world where change and decay takes place. He defines the subject matter as follows (Meteor. I.1, 338a20-339a10, translated by E. W. Webster):

\begin{quote}
338a20 We have already discussed the first causes of nature, and all natural motion, also the stars ordered in the motion of the heavens, and the physical element—enumerating and specifying them and showing how they change into one another—and becoming and perishing in general. There remains for consideration a part of this inquiry which all our predecessors called meteorology. It is concerned with events that are natural, though their order is less perfect than that of the first of the elements of bodies. They take place in the region nearest to the motion of the stars. Such are the milky way, and comets, and the movements of meteors. It studies also all the affections we may call common to air and water, and the kinds and parts of the earth and the affections of its parts. These throw light on the causes of winds and earthquakes and all the consequences the motions of these kinds and parts involve. Of these things some puzzle us, while others admit of explanation in some degree. Further, the inquiry is concerned with the falling of thunderbolts and with whirlwinds and fire-winds, and further, the recurrent affections produced in these same bodies by concretion. When the inquiry into these matters is concluded let us consider what account we can give, in accordance
\end{quote}

\(^1\)The writings of the Stoic Poseidonios is probably the cream of ancient meteorology, but we do not discuss his views here. Aristotle’s work is a very early and detailed account of several atmospheric optical phenomena and it had an enormous impact on later theories of the rainbow and colours in general—for the Arabic influence see (Lettinck 1999), for the rainbow Chapter VII—this explains the incorporation of this text.
with the method we have followed, of animals and plants, both generally and in detail. When that has been done we may say that the whole of our original undertaking will have been carried out.

Although Aristotle has worked out a detailed philosophy of science, little of that is visible in this predominantly empirical work. The explanations given are often tentative and only consistent with the phenomena (Meteor. I.7). Aristotle himself acknowledges this:

We consider a satisfactory explanation of phenomena inaccessible to observation to have been given when our account of them is free from impossibilities.

The aim is therefore to provide possible (and hopefully probable) explanations to phenomena like earthquakes, rainbows, etc. As meteorology had many practical uses, this area was one of the first where rational explanations started to compete with accounts operating with gods and the like. Aristotle, however, requires more from explanations: they should not just be credible metaphors (Meteor. II.3 357a):

It is equally absurd to suppose that anything has been explained by calling the sea ‘the sweat of the earth’, like Empedocles. Metaphors are poetical and so that expression of his may satisfy the requirements of a poem, but as a scientific theory it is unsatisfactory.

In spite of the fact that meteorology was a major part of the first philosophical-physical systems, the extreme complexity of several phenomena hindered the development its development, and for several centuries it remained a collection of well-known facts mixed with accounts of often greatly distorted rare occurrences. It united plausible assumptions and contingent ad hoc explanations.

Text 1 (not reproduced in the book, but with links provided) includes Aristotle’s explanations of haloes (III.2-3), rainbows (III.4-5), mock suns and rods (III.6). Haloes are very common, often spotted by the trained eye. The most common one is a halo of 22° radius around the sun or the moon. There are, however several other halos.

The HALO3 program can give the reader insight into the formation of several haloes. This program—not discussed here—is freely obtainable
under http://www.sundog.clara.co.uk/halo/halfeat.htm with detailed description of functioning and the physical bases of halo-formation. With the help of the program simulations can be run that take into account several factors: size and orientation of ice crystals, position of the sun, the camera, number of ray-paths calculated, etc. Here only a simple white-on-black image is included (sun in the center) that has been obtained with the help of the program (see figure IV.1).

Figure IV.1: This computer-generated picture of an atmospheric halo has been calculated using the HALO3 program.

On the two sides of the sun (in the center), near the $22^\circ$ halo, the bright patches of the mock-suns or sundogs are discernible—also mentioned in Aristotle’s text. These phenomena are rather less common, and seeing the sun embraced by the two mock suns is a striking experience. Aristotle also discusses rods—these are rainbow-coloured parts of the sky near the sun, usually appearing when thin clouds are present. Rainbows are well-known appearances, and will be investigated in detail in Chapter VII.

- **Problem 1** Collect the characteristic similarities of the listed atmospheric optical phenomena (haloes, rainbows, mock suns, rods). How do the circumstances in which each occurs differ? What influence do these have on the phenomena?
Apart from atmospheric optical phenomena, Text 1 also contains elements of a theory of vision and colours. This theory is most probably an early explanation by Aristotle with close ties to extromissionist theories and the mathematical optical tradition (see Chapter VI). As in other works, here too, Aristotle operates with the concept of the visual ray. In *On the Heavens* (Cael. 2.8 290a17-24, from (Aristoteles 1984), translated by J. L. Stocks) he notices that, in contrast to the stars, the planets don’t twinkle. For him the reason is that they are closer and sight receives them with full strength (the visual ray does not “weaken”):

The visual ray being excessively prolonged becomes weak and wavering. The same reason probably accounts for the apparent twinkle of the fixed stars and the absence of twinkling in the planets.

In his *Meteorology* the modification of the visual ray explains the various colours of the rainbow:

- **Problem 2** Summarise how the colours of atmospheric optical phenomena are explained. Investigate the differences between the colours of the primary and the secondary rainbow and the rods. What pose special difficulties to such explanations?

- **Problem 3** Meteorology has been a science with high practical uses. Collect references where the descriptions in Text 1 can be used for such purposes.

## 3 On the Soul and the Senses

Aristotle’s view on the soul is most thoroughly treated in the *De Anima* (An). This book is closely connected to his treatise on the senses and the objects of sensation *De sensu* (Sens). His general psychology and the interpretation of his views on sensation are heatedly debated even today, and this section can only offer a short introduction to his explanation of the process of vision and to his theory of colours. Text 2 contains chapters 5-7 from the second book of his *De Anima*, while Text 3 reproduces the first three chapters of the *De sensu*.

### 3.1 On the Soul

The fifth chapter (An 2.5) defines sensation as a sort of movement. The word *kinesis*, however, can also refer to qualitative change, and does not necessarily imply a movement of physical bodies.
Like his pupil Theophrastus (see Chapter II), Aristotle is not hostile to theories of perception where like is supposed to be affected by like—but he believes that this statement is in need of qualification. He states that as a result of perception the organ of perception undergoes change, but is altered in a special way. It is assimilated to the thing perceived, but this alteration is a “quasi-alteration”.

- **Problem 4** Reconstruct why the concepts of actuality and the two aspects of potentiality are significant for the explanation of perception in Text 2, chapter 5.

- **Problem 5** Why do the senses not perceive themselves? Investigate Aristotle’s answer in Text 2, chapter 5 and Text 3 chapter 2. (No hint is given to this problem, as the answer is based on the close reading of the text.)

- **Problem 6** List common sensibles, sensibles perceptible by a single sense, and incidental sensibles in Text 2, chapter 6. What are the differences? (No hint is given to this problem, as the answer is based on the close reading of the text.)

- **Problem 7** What is the role of the medium and the transparent in vision according to Text 2 chapter 7?

### 3.2 Alexander of Aphrodisias

This subsection has two aims. One is to give a clearer (or, at least, more elaborate) account of Aristotle’s views on colour and vision. The other is to offer a glimpse of the immense and extremely important corpus of the Aristotle commentators. The large number of commentaries that we possess on Aristotle’s writings luckily facilitates understanding his often difficult texts. While some of these undoubtedly further complicate already confused issues, others are known to be reliable and trustworthy sources that the scholar can turn to when interpreting a difficult passage.

One of the highly renowned commentators is Alexander of Aphrodisias. Among a number of other works he wrote a treatise on certain sections of Aristotle’s *Meteorology* (Alexander 1996), edited and translated into English by Eric Lewis. First we cite his account of the rainbow from (Gage 1993, p. 31). It is, in fact, an extended version of the account we have already seen in Text 1 with Aristotle. Mark the differences in the details of the treatment:
That the ... colours of the rainbow can neither be procured nor
imitated by painters, and that red [phoinikoun, puniceus] is closer to
white than green [prasinon, prasinus] and violet [halourgon, halurgus]
is clear from the following. The natural red pigments are cinnabar
[kinnabari] and dragon’s blood [drakontion] which are made from the
blood of animals;\(^2\) red is also made from a mixture [mixis] of talc
[koapholithos] and purple [porphuron, purpureum], but this is much in-
ferior to the natural colours. Natural green [prasinon] and violet are
chrysocolla and ostrum, the one made from blood and the other sea-
purple.\(^3\) But the artificial colours cannot match them: green is in-
deed made from blue [kuanon] and yellow [êchron], but violet from
blue and red, for the contrasting energies of blue and yellow make
green, but those of blue and red, violet. And in these cases the
artificial colours are far inferior to the natural. ... That red is closer
to white than green and violet is evident from their origin. For red
is made from talc, which is white, but green from ochre, which is
a weaker white, for a [gradually] darkened light appears first to be
changed into this colour, thus red is closer than green to white.
... But again [it is clear] that green is closer than violet to white, since
the former is made from yellow, but violet from red ... and yellow is
closer than red to white...

Next is Alexander’s elaboration of the Aristotelian concept of the trans-
parent or diaphanês (Alexander 1992-, Vol. II, pp. 20-23). This small section
will offer us a glimpse into the enormous literature that mushroomed on the
basis of the Aristotelian corpus and significantly influenced thinking about
light, colours, and vision for nearly two millenia. Reading these passages is
not without difficulties, but it helps to gain a more detailed understanding
of the Aristotelian concept of colour. (The marginal page-numbers of the
English translation are retained.)

Robert Sharples, the translator of the passage, has left the term diaphanês
untranslated. As he explains “it is often rendered by ‘transparent’, but Aris-
totle explains the surface colour of opaque objects as well as translucent
ones in this way. He does so, indeed, because in his view the diaphanês
which is most obviously present in transparent or translucent bodies is also

\(^2\)Cinnabar is the usual term for the red sulphide of mercury (HgS) but Alexander
is probably thinking of vermiculum or kermes made from the dried insect Coccus illicis, as
(Gage 1993, p. 273) suggests (misspelling sulphide for sulphate, and not capitalizing coccus).
Dragon’s blood is a reddish resin produced by a specific palm.

\(^3\)Chrysocolla is probably copper carbonate CuCO\(_3\).
present to some extent throughout opaque ones (cf. *Sens.* 3.439b9-10, see also page 85); but to say 'the boundary of body in so far as it is transparent is colour' may obscure the point that the explanation applies to opaque bodies as well. …Alexander distinguishes between those *diaphanê* bodies that are *diopta* and those that are not; ‘transparent’ seems more suitable as a translation of *diopta* than of *diaphanês*. Since light is the actuality of the *diapahanês*, *diaphanês* might also be rendered by ‘illuminable’; that captures part of the meaning but loses the idea of the *diaphanês* as that through which light extends.” (Alexander 1992-, Vol. II, p. 20)

Alexander starts by connecting two properties of objects: their surface and their colour:

Aristotle defined colour as the boundary of the determinate *diaphanês* *qua* *diaphanês*. For the boundary of body in so far as it is body is surface, but in so far as it is *diaphanês* it is colour. And for this reason colour and surface go together. For since the boundary of every body as body is surface, and the *diaphanês* too is body, the boundary of this too as body is surface. But colour is also its boundary as *diapahanês*, and both its boundaries are together, if they are boundaries, not being the same as one another.

But the relation between surface and colour needs qualification:

For every colour is in a surface and accompanied by a surface, but not every surface is accompanied by colour, because neither is every *diaphanês* body determinate. For of *diaphanês* body some is determinate, and some indeterminate. So just as what is *diaphanês* but indeterminate (like air and water) does not have a surface which is its own and determinate (for fluid and indeterminate bodies are defined by the things that surround them), just so it does not have its own colour, but is a messenger for the colours of other things and receives them, as it also does [their] shapes. And nature made the *diaphanês* that was going to be a messenger for the colours of other things colourless, reasonably; *diaphanês* so that it could receive colours, but indeterminate and colourless so that its own proper colour should not, by mixture with the things seen through it, impede their true revelation. But what is indeterminate and solid body has a determinate shape and boundary both as body and as *diaphanês*, and it is its boundary *qua* *diaphanês* that is colour. For colour is the boundary of the determinate *diaphanês*. For not every body is *diaphanês* (as not every [body] is receptive of colour), nor is everything
that is *diaphanê* determinate. The [parts] in the depth of determinate *diaphanê* things have colour in the same way that they have surface and shape [i.e.: potentially, if the thing should be divided].

Alexander now discusses the types of *diaphanê* things:

Of determinate *diaphanê* things some are more so and some less so. Most *diaphanê* are bright and white bodies, and fire is of this nature (for it appears yellow through the admixture of fumes, which are black). Things are *diaphanê* in the second and third degree according to their distance from this, and as many bodies as are black, being deprived of colour, are also deprived of *diaphaneia*.

The concept of *diaphanê* allows for a localization of colour:

“For since colour is among the things that are and is also among those whose nature is to be in other things, there had also to be some body in which colour [might be], and this is a *diaphanê* body. For this is the proximate matter of colour, what is *diaphanê* in actuality possessing [colour] in actuality, what is [*diaphanê*] potentially being like matter receptive of colours that are opposite to one another and of those that are intermediate between these. And in every body that possesses colour or is receptive of colour there is mixed in also the nature of the *diaphanê*.

The colour of objects can thus be explained:

So in those bodies in which, being *diaphanê* or possessing some illuminability in themselves, the boundary is evident and determinate (and this is so in those that are solid), in these there is also colour that is evident and proper to them, and it is in these alone that there is colour in the strict sense of the term, and more so in those in which this is more so. The indeterminate *diaphanê* has no proper or determinate colour because, through its rarity, it is not able to keep it or retain any. By being receptive of and a messenger for the

---

4In his work on Aristotle’s *On Sense Perception* (trans. A. Towey, 45.11-14), Alexander also mentions the body through which colours appear, the peculiarly *diaphanê*: “Only bodies through which things are seen are peculiarly transparent [*diaphanê*], as they are customarily described, from the fact that they bring <things> to light. That which is apprehensible to sight when it is in the light is described in the proper sense as coming to-light. Bodies whose colour is <light> are peculiarly transparent [*diaphanê*]” (Alexander 2000).
colours of other things it does not take on any of these in a way that involves its being affected. Rather, it is moved by them according to the presence and sort of relation to this [the indeterminate *diaphanêς*] of the bodies that possess their own proper colour, [and in this way] for as long as it is possible for them to be seen through the movement which is brought about in it by the colours that those things possess, it serves as messenger for those of living creatures that are able to see, so that they apprehend the colours through it. This the *diaphanêς* is able to do if it is illuminated; when this is first brought about by the presence of the things which are of such a nature as to illuminate [it], it is affected and receives this as [if it were] its own proper colour. For not even this is [in fact] its own proper colour, but as long as that which is by nature such as to illuminate, [it] is present, it possesses light, but when it departs the light it [the *diaphanêς*], too, ceases. So for as long as it is illuminated it receives the qualities and differentiations of other colours too, just as [it does] the light, and acts as a messenger for the visual senses; these too themselves are composed of a body that is *diaphanêς* in the same way and is, through the affection that is brought about in the sight by the *diaphanêς* that is outside [the eyes], moved in the aforementioned way by the presence of colours.

The colours of objects and the mixing of colours is also connected to properties other than colour:

In the *diaphanêς* that is indeterminate and, on account of its rarity, does not take on any colour in a way that involves its being affected, the presence of that which is of such a nature as to illuminate is the cause of light, its absence darkness. Just so in determinate bodies too the greater admixture and presence of the most *diaphanêς* body (and this is that which has white as its own proper colour; for this is most colour, and most visible) is the cause of bodies being coloured to a greater degree and being more visible; and the complete absence of this is the cause of blackness, which is like the privation of colour rather than colour. The intermediate colours have their origin and their differentiation according to the particular type of mixture of [things] that are *diaphanêς* and those that are not like this.

The colours of simple bodies can also be explained:

Of the simple bodies fire is the most visible in its own nature
and diaphanês in the way I have described; it is such that it can also illuminate air and provide it with perfection according to its being, for light is the actuality and form of the indeterminate diaphanês qua diaphanês. Earth is least like this. It is by a certain sort of mixture of these with one another, and by that of the things produced by the mixture of these things, that [there is brought about] the generation and differentiation of the colours apart from the primary ones (and the primary [ones] are [those] in the simple and primary bodies.)

And, finally, Alexander supports his account with an analogy:

In the case of odour some things have odour in themselves, those that have the sort of origin and constitution which we say is productive of odour, but there is also that which is odourless in itself, but receives and transmits the odours in other things (and this is called ‘transodorant’). Just so determinate diaphanês things possess colour (for this is the nature of colour), but what is colourless in its own nature receives and serves as a messenger for colour, too, and this is the indeterminate diaphanês. When we are asked what surface is, we say the boundary of a solid; just so, when we are asked what colour is, it would be reasonable for us to say that it is the boundary of the determinate diaphanês, since the relation of surface to determinate body is also the relation of colour to the determinate diaphanês.

3.3 ON THE SENSES

Text 3 is from Aristotle’s work on Sense and Sensibilia. The first three chapters from Book I begin the discussion of sensation, attributed to all animals by Aristotle. He groups and classifies the senses, looks at the makeup of the eye, and investigates vision, and the seeing and mixing of colours.

- **Problem 8** Which elements of theories discussed earlier surface explicitly or implicitly in Text 3? Summarise your findings. Compare Aristotle’s evaluation to Theophrastus’s in Chapter II. What are his main arguments for accepting his account of the visual organ? (No hint is given to this problem, as the answer is based on the close reading of the text.)

- **Problem 9** What is the origin of the colours we see? List the ways Aristotle thinks colours can arise. Compare what he writes about the
visual ray and colour mixing with his account in the Meteorology, in Text 1. (No hint is given to this problem, as the answer is based on the close reading of the text.)

4  ON COLOURED BODIES: De Coloribus

One of the most detailed antique works on colours is the pseudo-Aristotelian work On colours, reprinted in Text 4.

The small treatise mainly investigates colour changes and colours of plants, animals, and other, non-animate objects. Accordingly, colours are considered as properties of substances. The elements possess colour, and the colours seen are mixtures of the elemental colours. Colours are seen as intermediaries between white and black. But, unlike in De Sensu and De Anima, but similarly to the Meteorology black is (also?) considered as a failure of light—there is a correlation between the weakening and the colour of light. The medium (even when it is rare, as in the case of air) posits resistance to the light rays—for short distances this resistance can be overcome, but on longer distances the effect affects the colour: hence the blue of the sky. Light is not treated as the colour of the transparent, and the role of the transparent medium as the medium of vision (without which no vision is possible) is also dropped.

The six chapters tackle many of the important questions concerning colour. Goethe in his Theory of Colours gave telling titles to the chapters that are still in use (Goethe 1989b), and this section will shortly introduce them one by one.

1. On simple colours
2. On mixing
3. On the indeterminability of colours
4. On artificial colours
5. On the changes of plant-colours through organic maturation
6. On the colours of hairs, feathers, and skins

The beginning of the book clearly shows connections to the ancient four-element theories (see 5.1 on page 108). But, unlike in earlier writings,
the colour of the elements are either white (air, water, and earth), or yellow (fire or sun). A process of change, transmutation, or ‘burning’ gives black

- **Problem 10** Describe the three conditions under which we see black. What sort of theories of light and vision do these entail?

The second chapter gives an account of the origin of compound colours, but expressly not based on the pigment-mixing practice of painters. The compound colours can be further compounded. But the colour perceived is the result of several factors, and not reducible to any one single property of light, object, medium, or observer.

- **Problem 11** List the conditions that can change or modify colours.

- **Problem 12** Based on Problem 9, list the mentioned compound colours and colour mixing techniques in Text 3 and compare these to the possibilities listed in Text 4. (No hint is given to this problem, as the answer is based on the close reading of the text.)

The fourth chapter tackles a highly technical aspect of colouring: dying and colouring technique. A general mechanism of dying is also summarised, where the transmission of colour is connected to moisture and heat, and the binding of the dye in the pores of the material is a result of drying.

The following chapter investigates colour changes in plants as a function of the level of maturation and the lack or availability of moisture and heat. The Greek word *pepsis* stands for any natural process of maturation in which heat plays a significant role. The process of producing the most important purple dye of the ancients is also discussed in the chapter. The material obtained was a secretion of certain sea animals, in later epochs most importantly the rather large *Murex brandaris*. The obtained white liquid was oxidised after contact with air, and turned yellow. Cooking or storage with salt (and further oxidation) changed the colour to purple. Later in the chapter the purple-fish (*Murex trunculus*) is also mentioned, but the colours used in the description probably do not refer to temporal development of the colours.

---

5 A note on translation. In Chapter 1 the lye-mixture (791a8) can also refer to ashes or burnt chalk.

6 These Latin names as cited in (Aristoteles 1999) are problematic. The species are not clearly identifiable any more. Possibilities include *Murex brandaris, M. trunculus* and *Purpura lapillus*. 
The last chapter also discusses colour changes. The colours of animals (for example red and grey) are between black and white. Maturation and lack of water shifts the colour towards white, abundant and old moisture or water towards black. As opposed to the previous section, no mention is made here of the medium, as an important contributor to colour perception.

- **Problem 13** Summarise the colour changes of plants.

5 Texts

5.1 Text 1: Meteorology

- Aristotle’s explanations of haloes (III.2-3), rainbows (III.4-5), mock suns and rods (III.6) from his *Meteorology* is not reproduced here, as it is readily available in (Aristoteles 1984) or under http://classics.mit.edu/Aristotle/meteorology.mb.txt as translated by E. W. Webster. It appeared in the third volume of the complete Oxford edition of Aristotle, printed from 1908 (Aristotle 1908-). A copy can be reached via the homepage of the WTWG Department of the University of Berne: http://www.philoscience.unibe.ch/bernstudies/titel-.html. Read Book I Part 13 and Book III Parts 2-6.

- **Hint to Problem 1** Aristotle finds a general similarity connecting rods, mock suns, rainbows, and haloes: all are caused by reflection. This seems surprising, as it is generally held that refraction also plays a significant role in coloured atmospheric phenomena, like rainbows, rods. But the Greek term *anaklasis* used in the original can either be translated by reflection and refraction. Used also by Empedocles and Anaxagoras earlier, it was commonly used in Presocratic explanations, conflating the two—for us markedly different—meanings.

It is relatively straightforward to characterise the occurrences with respect to the presence or absence of certain conditions (around the sun, around the moon, full circle, not greater than a semicircle, coloured, only visible at certain parts of the day or year). These neatly separate the phenomena, and provide a classificatory method. This classification, however, is unlike the one seen in the previous chapter by Plato (see Hint to Problem 2 on page 54) as it does not follow a
strict dichotomy, but rather lists a number of factors that all contribute to the classification.

- **Hint to Problem 2** The colour appearances result from a weakening of the visual ray. The example of the weak-sighted man (373b2-10) suggests that the outward motion of the visual ray can be weakened by objects. In case the visual ray is weak, even air can have a big enough resistance for the ray to be reflected. This theory does not need atomistic assumptions and only relies on the qualitative notions of “strength” and “weakness” to explain colour perception.

The colours of the primary bow are explained by the gradual weakening of the visual ray. The colour red is characterised as the effect of darkness (a dark medium or object) on light, the other two colours as gradual further darkening and weakening of the visual ray.

Although apart from the intensity the colours of the primary and secondary rainbow are very similar (thus suggesting a unified explanation), for the first the gradual weakening of the visual ray is employed (as a result of the dark cloud), for the secondary the distance of the observer.

### 5.2 Text 2: On the Soul

- The text of *De Anima*, or *On the soul* was translated by J. A. Smith, and appeared in the third volume of (Aristotle 1908-), the revised edition in (Barnes 1984). © The Jowett Copyright Trustees. (Note that *diaphanês* is here translated as “transparent”.)

**Book II Chapter 5**

*Having made these distinctions let us now speak of sensation in the widest sense. Sensation depends, as we have said, on a process of movement or affection from without, for it is held to be some sort of change of quality. Now some thinkers assert that like is affected only by like; in what sense this is possible and in what sense impossible, we have explained in our general discussion of acting and being acted upon.*

*Here arises a problem: why do we not perceive the senses themselves as well as the external objects of sense, or why without the stimulation of external objects do they not produce sensation, seeing that they contain in themselves fire, earth, and all the other elements, which are the direct or indirect objects is so of sense? It is clear that what is sensitive is only potentially, not actually.*
The power of sense is parallel to what is combustible, for that never ignites itself spontaneously, but requires an agent which has the power of starting ignition; otherwise it could have set itself on fire, and would not have needed actual fire to set it ablaze.

In reply we must recall that we use the word ‘perceive’ in two ways, for we say (a) that what has the power to hear or see, ‘sees’ or ‘hears’, even though it is at the moment asleep, and also (b) that what is actually seeing or hearing, ‘sees’ or ‘hears’. Hence ‘sense’ too must have two meanings, sense potential, and sense actual. Similarly ‘to be a sentient’ means either (a) to have a certain power or (b) to manifest a certain activity. To begin with, for a time, let us speak as if there were no difference between (i) being moved or affected, and (ii) being active, for movement is a kind of activity—an imperfect kind, as has elsewhere been explained. Everything that is acted upon or moved is acted upon by an agent which is actually at work. Hence it is that in one sense, as has already been stated, what acts and what is acted upon are like, in another unlike, i.e. prior to and during the change the two factors are unlike, after it like.

But we must now distinguish not only between what is potential and what is actual but also different senses in which things can be said to be potential or actual; up to now we have been speaking as if each of these phrases had only one sense. We can speak of something as ‘a knower’ either (a) as when we say that man is a knower, meaning that man falls within the class of beings that know or have knowledge, or (b) as when we are speaking of a man who possesses a knowledge of grammar; each of these is so called as having in him a certain potentiality, but there is a difference between their respective potentialities, the one (a) being a potential knower, because his kind or matter is such and such, the other (b), because he can in the absence of any external counteracting cause realize his knowledge in actual knowing at will. This implies a third meaning of ‘a knower’ (c), one who is already realizing his knowledge—he is a knower in actuality and in the most proper sense is knowing, e.g. this A. Both the former are potential knowers, who realize their respective potentialities, the one (a) by change of quality, i.e. repeated transitions from one state to its opposite under instruction, the other (b) by the transition from the inactive possession of sense or grammar to their active exercise. The two kinds of transition are distinct.

Also the expression ‘to be acted upon’ has more than one meaning; it may mean either (a) the extinction of one of two contraries by the other, or (b) the maintenance of what is potential by the agency of what is actual and already like what is acted upon, with such likeness as is compatible with one’s being actual and the other potential. For what possesses knowledge becomes an actual knower by a transition which is either not an alteration of it at all (being in reality a development into its true self or actuality) or at least an alteration in a quite different sense from the usual meaning.

Hence it is wrong to speak of a wise man as being ‘altered’ when he uses
his wisdom, just as it would be absurd to speak of a builder as being altered when he is using his skill in building a house.

What in the case of knowing or understanding leads from potentiality to actuality ought not to be called teaching but something else. That which starting with the power to know learns or acquires knowledge through the agency of one who actually knows and has the power of teaching either (a) ought not to be said ‘to be acted upon’ at all or (b) we must recognize two senses of alteration, viz. (i) the substitution of one quality for another, the first being the contrary of the second, or (ii) the development of an existent quality from potentiality in the direction of fixity or nature.

In the case of what is to possess sense, the first transition is due to the action of the male parent and takes place before birth so that at birth the living thing is, in respect of sensation, at the stage which corresponds to the possession of knowledge. Actual sensation corresponds to the stage of the exercise of knowledge. But between the two cases compared there is a difference; the objects that excite the sensory powers to activity, the seen, the heard, etc., are outside. The ground of this difference is that what actual sensation apprehends is individuals, while what knowledge apprehends is universals, and these are in a sense within the soul. That is why a man can exercise his knowledge when he wishes, but his sensation does not depend upon himself: a sensible object must be there. A similar statement must be made about our knowledge of what is sensible on the same ground, viz. that the sensible objects are individual and external.

A later more appropriate occasion may be found thoroughly to clear up all this. At present it must be enough to recognize the distinctions already drawn; a thing may be said to be potential in either of two senses, (a) in the sense in which we might say of a boy that he may become a general or (b) in the sense in which we might say the same of an adult, and there are two corresponding senses of the term ‘a potential sentient’. There are no separate names for the two stages of potentiality; we have pointed out that they are different and how they are different. We cannot help using the incorrect terms ‘being acted upon or altered’ of the two transitions involved. As we have said, what has the power of sensation is potentially like what the perceived object is actually; that is, while at the beginning of the process of its being acted upon the two interacting factors are dissimilar, at the end the one acted upon is assimilated to the other and is identical in quality with it.

Book II Chapter 6

In dealing with each of the senses we shall have first to speak of the objects which are perceptible by each. The term ‘object of sense’ covers three kinds
of objects, two kinds of which are, in our language, directly perceptible, while the remaining one is only incidentally perceptible. Of the first two kinds one (a) consists of what is perceptible by a single sense, the other (b) of what is perceptible by any and all of the senses. I call by the name of special object of this or that sense that which cannot be perceived by any other sense than that one and in respect of which no error is possible; in this sense colour is the special object of sight, sound of hearing, flavour of taste. Touch, indeed, discriminates more than one set of different qualities. Each sense has one kind of object which it discerns, and never errs in reporting that what is before it is colour or sound (though it may err as to what it is that is coloured or where that is, or what it is that is sounding or where that is.) Such objects are what we propose to call the special objects of this or that sense.

‘Common sensibles’ are movement, rest, number, figure, magnitude; these are not peculiar to any one sense, but are common to all. There are at any rate certain kinds of movement which are perceptible both by touch and by sight.

We speak of an incidental object of sense where e.g. the white object which we see is the son of Diaries; here because ‘being the son of Diaries’ is incidental to the directly visible white patch we speak of the son of Diaries as being (incidentally) perceived or seen by us. Because this is only incidentally an object of sense, it in no way as such affects the senses. Of the two former kinds, both of which are in their own nature perceptible by sense, the first kind—that of special objects of the several senses—constitute the objects of sense in the strictest sense of the term and it is to them that in the nature of things the structure of each several sense is adapted.

**Book II Chapter 7**

The object of sight is the visible, and what is visible is (a) colour and (b) a certain kind of object which can be described in words but which has no single name; what we mean by (b) will be abundantly clear as we proceed. Whatever is visible is colour and colour is what lies upon what is in its own nature visible; ‘in its own nature’ here means not that visibility is involved in the definition of what thus underlies colour, but that that substratum contains in itself the cause of visibility. Every colour has in it the power to set in movement what is actually transparent; that power constitutes its very nature. That is why it is not visible except with the help of light; it is only in light that the colour of a thing is seen. Hence our first task is to explain what light is.

Now there clearly is something which is transparent, and by ‘transparent’ I mean what is visible, and yet not visible in itself, but rather owing its visibility to the colour of something else; of this character are air, water, and many solid bodies. Neither air nor water is transparent because it is air or water; they are transparent because each of them has contained in it a certain substance which
is the same in both and is also found in the eternal body which constitutes the uppermost shell of the physical Cosmos. Of this substance light is the activity—the activity of what is transparent so far forth as it has in it the determinate power of becoming transparent; where this power is present, there is also the potentiality of the contrary, viz. darkness. Light is as it were the proper colour of what is transparent, and exists whenever the potentially transparent is excited to actuality by the influence of fire or something resembling ‘the uppermost body’; for fire too contains something which is one and the same with the substance in question.

We have now explained what the transparent is and what light is; light is neither fire nor any kind whatsoever of body nor an efflux from any kind of body (if it were, it would again itself be a kind of body)—it is the presence of fire or something resembling fire in what is transparent. It is certainly not a body, for two bodies cannot be present in the same place. The opposite of light is darkness; darkness is the absence from what is transparent of the corresponding positive state above characterized; clearly therefore, light is just the presence of that.

Empedocles (and with him all others who used the same forms of expression) was wrong in speaking of light as ‘travelling’ or being at a given moment between the earth and its envelope, its movement being unobservable by us; that view is contrary both to the clear evidence of argument and to the observed facts; if the distance traversed were short, the movement might have been unobservable, but where the distance is from extreme East to extreme West, the draught upon our powers of belief is too great.

What is capable of taking on colour is what in itself is colourless, as what can take on sound is what is soundless; what is colourless includes (a) what is transparent and (b) what is invisible or scarcely visible, i.e. what is ‘dark’. The latter (b) is the same as what is transparent, when it is potentially, not of course when it is actually transparent; it is the same substance which is now darkness, now light.

Not everything that is visible depends upon light for its visibility. This is only true of the ‘proper’ colour of things. Some objects of sight which in light are invisible, in darkness stimulate the sense; that is, things that appear fiery or shining. This class of objects has no simple common name, but instances of it are fungi, flesh, heads, scales, and eyes of fish. In none of these is what is seen their own proper colour. Why we see these at all is another question. At present what is obvious is that what is seen in light is always colour. That is why without the help of light colour remains invisible. Its being colour at all means precisely its having in it the power to set in movement what is already actually transparent, and, as we have seen, the actuality of what is transparent is just light.
The following experiment makes the necessity of a medium clear. If what has colour is placed in immediate contact with the eye, it cannot be seen. Colour sets in movement not the sense organ but what is transparent, e.g. the air, and that, extending continuously from the object to the organ, sets the latter in movement. Democritus misrepresents the facts when he expresses the opinion that if the interspace were empty one could distinctly see an ant on the vault of the sky; that is an impossibility. Seeing is due to an affection or change of what has the perceptive faculty, and it cannot be affected by the seen colour itself; it remains that it must be affected by what comes between. Hence it is indispensable that there be something in between—if there were nothing, so far from seeing with greater distinctness, we should see nothing at all.

We have now explained the cause why colour cannot be seen otherwise than in light. Fire on the other hand is seen both in darkness and in light; this double possibility follows necessarily from our theory, for it is just fire that makes what is potentially transparent actually transparent.

The same account holds also of sound and smell; if the object of either of these senses is in immediate contact with the organ no sensation is produced. In both cases the object sets in movement only what lies between, and this in turn sets the organ in movement: if what sounds or smells is brought into immediate contact with the organ, no sensation will be produced. The same, in spite of all appearances, applies also to touch and taste; why there is this apparent difference will be clear later. What comes between in the case of sounds is air; the corresponding medium in the case of smell has no name. But, corresponding to what is transparent in the case of colour, there is a quality found both in air and water, which serves as a medium for what has smell—I say ‘in water’ because animals that live in water as well as those that live on land seem to possess the sense of smell, and ‘in air’ because man and all other land animals that breathe, perceive smells only when they breathe air in. The explanation of this too will be given later.

- **Hint to Problem 4** For orientation use this quote from Alexander’s *On Aristotle’s “On Sense Perception”*, 42.7-13, (Alexander 2000):

  For the actual colour is not the same as the actual sight nor the actual sound the same as the hearing; for it is possible that these be actual even when they are not seen, but it is not at all possible that there should be actual sensibles without being perceived. Hence the meaning (i.e. of Aristotle’s text) is the following: we have in the *De anima* explained in what sense the actual colour perceived and the actual sound perceived are the
same as the actual sight and the actual hearing respectively, and in what sense they are different…

- **Hint to Problem 7** Without a medium no vision takes place: an object pressed to our eyes is invisible. This medium also has to enter a special state, this is a prerequisite to vision.

Light is a disposition of the medium, in this state the medium is actually transparent. No physical movement characterises this change of state, and its occurrence is also a prerequisite to vision. A citation from (Burnyeat 1995) might enlighten us:

> Fill a transparent glass with water and put it on a table. Hold a red object a short distance away from the glass and look at it through the water. The water in the glass is now serving as a medium within another medium (the surrounding air). You will see a red coloration in the water. But unlike the coloration that ensues if you pour red ink into the water, this coloration is not visible to other observers from other angles of vision. Now let the glass expand in your imagination to meet your eye, on one side, and the red object on the other. The water will become the sole medium and you will see the red object directly through it.

> Such is the effect of a colour on a medium which is actually transparent: the colour appears through it. […] It is evident, I hope, that this appearance or visibility of the colour through the transparent is a static condition, a state of affairs, not an event or process. Nothing happens. Nothing moves from the coloured object. There is not even a real alteration, only the quasi-alteration which consists in the fact that the colour is visible through the transparent.

But not only the space between the observer and the object is a potentially transparent medium. A similar medium is also a constituent of the eye. Here, for the second time, the same changes must take place for vision to occur. See also page 64.

5.3 **Text 3: Sense and Sensibilia**

- The translation of J. I. Beare *On Sense and the Sensible* appeared in (Aristotle 1908-), the revised text reproduced here in (Barnes 1984). ©The Jowett Copyright Trustees.
Having now considered the soul, by itself, and its several faculties, we must next make a survey of animals and all living things, in order to ascertain what functions are peculiar, and what functions are common, to them. What has been already determined respecting the soul must be assumed throughout. The remaining parts of our subject must be now dealt with, and we may begin with those that come first.

The most important attributes of animals, whether common to all or peculiar to some, are, manifestly, attributes of soul and body in conjunction, e.g., sensation, memory, passion, appetite and desire in general, and, in addition, pleasure and pain. For these may, in fact, be said to belong to all animals. But there are, besides these, certain other attributes, of which some are common to all living things, while others are peculiar to certain species of animals.

The most important of these may be summed up in four pairs, viz. waking and sleeping, youth and old age, inhalation and exhalation, life and death. We must examine these, determining their respective natures, and the causes of their occurrence.

But it behoves the natural scientist to obtain also a clear view of the first principles of health and disease, inasmuch as neither health nor disease can exist in lifeless things. Indeed we may say of most physical inquirers, and of those physicians who study their art more philosophically, that while the former complete their works with a disquisition on medicine, the latter start from a consideration of nature.

That all the attributes above enumerated belong to soul and body in conjunction, is obvious; for they all either imply sensation as a concomitant, or have it as their medium. Some are either affections or states of sensation, others, means of defending and safe-guarding it, while others, again, involve its destruction or privation. Now it is clear, alike by reasoning and without reasoning, that sensation is generated in the soul through the medium of the body.

We have already, in our treatise On the Soul, explained the nature of sensation and perceiving, and the reason why this affection belongs to animals. Sensation must, indeed, be attributed to all animals as such, for by its presence or absence we distinguish between what is and what is not an animal.

But coming now to the special senses severally, we may say that touch and taste necessarily appertain to all animals, touch, for the reason given in On the Soul, and taste, because of nutrition. It is by taste that one distinguishes in food the pleasant from the unpleasant, so as to flee from the latter and pursue the former; and savour in general is an affection of the nutritive part.

The senses which operate through external media, viz. smelling, hearing, seeing are found in all animals which possess the faculty of locomotion. To all that possess them they are a means of preservation in order that, guided
by antecedent perception, they may both pursue their food, and shun things that are bad or destructive. But in animals which have also intelligence they serve for the attainment of a higher perfection. They bring in tidings of many distinctive qualities of things, from which knowledge of things both speculative and practical is generated in the soul.

Of the two last mentioned, seeing, regarded as a supply for the primary wants of life is in its own right the superior sense, but for developing thought hearing incidentally takes the precedence. The faculty of seeing, thanks to the fact that all bodies are coloured, brings tidings of multitudes of distinctive qualities of all sorts; whence it is through this sense especially that we perceive the common sensibles, viz. figure, magnitude, motion, number; while hearing announces only the distinctive qualities of sound, and, to some few animals, those also of voice. Incidentally, however, it is hearing that contributes most to the growth of intelligence. For rational discourse is a cause of instruction in virtue of its being audible, which it is, not in its own right, but incidentally; since it is composed of words, and each word is a symbol. Accordingly, of persons destitute from birth of either sense, the blind are more intelligent than the deaf and dumb.

Book I Chapter 2
Of the distinctive powers of each of the faculties of sense enough has been said already.

But as to the nature of the sensory organs, or parts of the body in which each of the senses is naturally implanted, some inquire into them with reference to the elements of bodies. Not, however, finding it easy to coordinate five senses with four elements, they are at a loss respecting the fifth sense. They all hold the organ of sight to consist of fire, being prompted to this view by a certain affection of whose true cause they are ignorant. This is that, when the eye is pressed and moved, fire appears to flash from it. This naturally takes place in darkness, or when the eyelids are closed—for then, too, darkness is produced.

This raises another puzzle; for, unless a man can perceive and see without being aware of it, the eye must see itself. But then why does the above affection not occur also when the eye is at rest? The true explanation of this affection, which will contain the answer to our question, and account for the current notion that the eye consists of fire, must be determined in the following way:

Things which are smooth have the natural property of shining in darkness, without, however, producing light. Now, the part of the eye called the black, i.e. its central part, is smooth. The phenomenon of the flash occurs only when the eye is moved, because one object then becomes as it were two. The rapidity of the movement has the effect of making that which sees and that which is seen
seem different from one another. Hence the phenomenon does not occur unless the motion is rapid and takes place in darkness. For it is in the dark that that which is smooth, e.g. the heads of certain fishes, and the sepia of the cuttlefish, naturally shines, and, when the movement of the eye is slow, it is impossible that that which sees and that which is seen should appear to be simultaneously two and one. The eye sees itself in the above phenomenon as it does so in reflection.

If the visual organ were fire, which is the doctrine of Empedocles, a doctrine taught also in the Timaeus, and if vision were the result of light issuing from the eye as from a lantern, why should the eye not have had the power of seeing even in the dark? It is totally idle to say, as the Timaeus does, that the visual ray coming forth in the darkness is quenched. What is a quenching of light? That which, like a fire of coals or an ordinary flame, is hot and dry is, indeed, quenched by the moist or cold; but heat and dryness are not evidently attributes of light. And if they are attributes of it, but belong to it in a degree so slight as to be imperceptible to us, we should have expected that in the daytime the light of the sun should be quenched when rain falls, and that darkness should prevail in frosty weather. After all, flame and ignited bodies are subject to such extinction, but experience shows that nothing of this sort happens to the sunlight.

Empedocles at times seems to hold, as we said before, that vision occurs when light issues forth from the eye, e.g., in the following passage:

As when one who purposes going abroad prepares a lantern,
A gleam of fire blazing through the stormy night,
Adjusting thereto, to screen it from all sorts of winds, transparent sides,
Which scatter the breath of the winds as they blow,
While, out through them leaping, the fire, i.e. all the more subtle part of this,
Shines along his threshold with incessant beams:
So the primaeval tire, fenced within the membranes.
And delicate tissues gave birth to a round-eyed daughter
Tissues bored through with wonderful channels
And these fended off the deep surrounding flood,
While letting through the fire, i.e. all its more subtle part.

Sometimes he accounts for vision thus, but at other times he explains it by emanations from the visible objects.

Democritus, on the other hand, is right in his opinion that the eye is of water; not, however, when he goes on to explain seeing as mirroring. The mirroring that takes place in an eye is due to the fact that the eye is smooth, and it really has its seat not in the eye, but in that which sees. For the case is one of reflection. But it would seem that in his time there was no scientific
knowledge of the general subject of the formation of images and the phenomena of reflection. It is strange, too, that it never occurred to him to ask why the eye alone sees, while none of the other things in which images are reflected do so.

True, then, the visual organ proper is composed of water, yet vision appertains to it not because it is water, but because it is transparent—a property common alike to water and to air. But water is more easily confined and more easily condensed than air; it is that the pupil, i.e. the eye proper, consists of water. That it does so is proved by facts of actual experience. The substance which flows from eyes when decomposing is seen to be water, and this in undeveloped embryos is remarkably cold and glistening. In sanguineous animals the white of the eye is fat and oily, in order that the moisture of the eye may be proof against freezing. Wherefore the eye is of all parts of the body the least sensitive to cold: no one ever feels cold in the part sheltered by the eyelids. The eyes of bloodless animals are covered with a hard scale which gives the similar protection.

It is, to state the matter generally, an irrational notion that the eye should see in virtue of something issuing from it; that the visual ray should extend itself all the way to the stars, or else go out merely to a certain point, and there coalesce, as some say, with rays which proceed from the object. It would be better to suppose this coalescence to take place in the fundament of the eye itself. But even this would be mere trifling. For what is meant by the coalescence of light with light? Or how is it possible? Coalescence does not occur between any two things taken at random. And how could the light within the eye coalesce with that outside it? For the membrane comes between them.

That without light vision is impossible has been stated elsewhere; but, whether the medium between the eye and its objects is air or light, vision is caused by a process through this medium.

Accordingly, that the inner part of the eye consists of water is easily intelligible, water being transparent.

Now, as vision outwardly is impossible without light, so also it is impossible inwardly. There must, therefore, be some transparent medium within the eye, and, as this is not air, it must be water. The soul or its perceptive part is not situated at the external surface of the eye, but obviously somewhere within: whence the necessity of the interior of the eye being transparent, i.e. capable of admitting light. And that it is so is plain from actual occurrences. It is matter of experience that soldiers wounded in battle by a sword slash on the temple, so inflicted as to sever the passages of the eye, feel a sudden onset of darkness, as if a lamp had gone out; because what is called the pupil, i.e. the transparent, which is a sort of lamp, is then cut off.

Hence, if the facts be at all as here stated, it is clear that—if one should explain the nature of the sensory organs in this way, i.e., by correlating each of them with one of the elements—we must conceive that the part of the eye
which sees consists of water, that what is perceptive of sound consists of air,
and that the sense of smell consists of fire. (For the organ of smell is potentially
that which the sense of smell is actually; since the object of sense is what causes
the actualisation of each sense, so that it must beforehand have been potentially
such and such. Now, odour is a smoke-like evaporation, and smoke-like evap-oration arises from fire. This also helps us to understand why the olfactory organ
has its proper seat in the environment of the brain; for cold matter is potentially
hot. In the same way must the genesis of the eye be explained. Its structure is
an offshoot from the brain, because the latter is the moistest and coldest of all
the bodily parts.)

The organ of touch consists of earth, and the faculty of taste is a particular
form of touch. This explains why the sensory organ of both touch and taste is
closely related to the heart. For the heart, as being the hottest of all the bodily
parts, is the counterpoise of the brain.

This, then, is the way in which the characteristics of the bodily organs of
sense must be determined.

**BOOK I CHAPTER 3**

Of the sensibles corresponding to each sensory organ, viz. colour, sound, odour,
savour, touch, we have treated in *On the Soul* in general terms, having there
determined what their function is, and what is implied in their becoming actual-
ised in relation to their respective organs. We must next consider what account
we are to give of any one of them; what, for example, we should say colour
is, or sound, or odour, or savour; and so also respecting touch. We begin with
colour.

Now, each of them may be spoken of from two points of view, i.e., either
as actual or as potential. We have in *On the Soul* explained in what sense the
colour or sound, regarded as actualised, is the same as, and in what sense it is
different from, the correlative sensation, the actual seeing or hearing. The point
of our present discussion is to determine what each sensible object must be in
itself, in order to produce actual sensation.

We have already in *On the Soul* stated of light that it is the colour of
the transparent incidentally; for whenever a fiery element is in a medium its
presence there is light; while the privation of it is darkness. But what we call
transparent is not something peculiar to air, or water, or any other of the bodies
usually called transparent, but is a common nature and power, capable of no
separate existence of its own, but residing in these, and subsisting likewise in
all other bodies in a greater or less degree. As the bodies in which it subsists
must have some extreme bounding surface, so too must this. Here, then, we
may say that light is a nature inhering in the transparent when the latter is
without determinate boundary. But it is manifest that, when the transparent
is in determinate bodies, its bounding extreme must be something real; and
that colour is just this something we are plainly taught by facts—colour being actually either at the limit, or being itself that limit in bodies. (Hence it was that the Pythagoreans named the superficies of a body its hue.) For it is at the limit of the body, but it is not the limit of the body; but the same natural substance which is coloured outside must be thought to be so inside too.

Air and water, too are evidently coloured; for their brightness is of the nature of colour. But the colour which air or sea presents, since the body in which it resides is not determinately bounded, is not the same when one approaches and views it close by as it is when one regards it from a distance; whereas in determinate bodies the colour presented is definitely fixed, unless, indeed, when the atmospheric environment causes it to change. Hence it is clear that in them which is susceptible of colour is in both cases the same. It is therefore the transparent, according to the degree to which it subsists in bodies (and it does so in all more or less), that causes them to partake of colour. But since the colour is at the extremity of the body, it must be at the extremity of the transparent in the body. Whence it follows that we may define colour as the limit of the transparent in determinately bounded body. For whether we consider the special class of bodies called transparent, as water and such others, or determinate bodies, which appear to possess a fixed colour of their own, it is at the exterior bounding surface that all alike exhibit their colour.

Now, that which when present in air produces light may be present also in the transparent; or again, it may not be present, but there may be a privation of it. Accordingly, as in the case of air the one condition is light, the other darkness, in the same way the colours white and black are generated in determinate bodies.

We must now treat of the other colours, reviewing the several ways in which they can come about.

It is conceivable that the white and the black should be juxtaposed in quantities so minute that either separately would be invisible, though the joint product would be visible; and that they should thus have the other colours for resultants. Their product could, at all events, appear neither white nor black; and, as it must have some colour, and can have neither of these, this colour must be of a mixed character—in fact, a species of colour different from either. Such, then, is a possible way of conceiving the existence of a plurality of colours besides the white and black; and we may suppose that many are the result of a ratio; for they may be juxtaposed in the ratio of 3 to 2, or of 3 to 4, or in ratios expressible by other numbers; while some may be juxtaposed according to no numerically expressible ratio, but according to some incommensurable relation of excess or defect: and, accordingly, we may regard all these colours as analogous to concords, and suppose that those involving numerical ratios, like the concords in music, may be those generally regarded as most agreeable; as, for example, purple, crimson, and some few such colours, their fewness being
due to the same causes which render the concords few. The other compound
colours may be those which are not based on numbers. Or it may be that,
while all colours whatever are based on numbers, some are regular in this
respect, others irregular; and that the latter, whenever they are not pure, owe
this character to a corresponding impurity in their numerical ratios. This then
is one way to explain the genesis of intermediate colours.

Another is that the black and white appear the one through the medium of
the other, giving an effect like that sometimes produced by painters overlaying
a less vivid upon a more vivid colour, as when they desire to represent an object
appearing under water or enveloped in a haze, and like that produced by the sun,
which in itself appears white, but takes a crimson hue when beheld through a
fog or a cloud of smoke. On this hypothesis, too, a variety of colours may be
conceived to arise in the same way as that already described; for between those
at the surface and those in no determinate ratio. To say with the ancients that
colours are emanations, and that the visibility of object is due to such a cause,
is absurd. For they must, in any case, explain sense-perception through touch;
so that it were better to say at once that visual perception is due to a process set
up by the perceived object in the medium between this object and the sensory
organ; due, that is, to contact, not to emanations.

If we accept the hypothesis of juxtaposition, we must assume not only
invisible magnitude, but also imperceptible time, in order that the arrival of
the movements may be unperceived, and that the colour may appear to be
one because they seem to be simultaneous. On the hypothesis of superposition,
however, no such assumption is needed: the stimulatory process produced in the
medium by the upper colour, when this is itself unaffected, will be different in
kind from that produced by it when affected by the underlying colour. Hence it
presents itself as a different colour, i.e. as one which is neither white nor black.
So that, if it is impossible to suppose any magnitude to be invisible, and we must
assume that there is some distance from which every magnitude is visible, this
superposition theory too might pass as a theory of colour-mixture. Indeed, in
the previous case also there is no reason why, to persons at a distance from the
juxtaposed blacks and whites, some one colour should not appear to present
itself as a blend of both. For it will be shown, in a discussion be undertaken
later on, that there is no magnitude absolutely invisible.

There is a mixture of bodies, however, not merely such as some suppose,
i.e. by juxtaposition of their minimal parts, which, owing to sense, are imper-
ceptible by us but a mixture by which they are wholly blent together, as we have
described it in the treatise on mixture, where we dealt with this subject generally
in its most comprehensive aspect. For, on the supposition we are criticizing, the
only totals capable of being mixed are those which are divisible into minimal
parts as men, horses, or seeds. For of mankind as a whole the individual man is
such a least part; of horses the individual horse. Hence by the juxtaposition of
these we obtain a mixed total, consisting of both together: but we do not say that by such a process any individual man has been mixed with any individual horse. Not in this way, but by complete interpenetration must we conceive those things to be mixed which are not divisible into minima; and it is in the case of these that natural mixture exhibits itself in its most perfect form. We have explained already in our discourse on mixture how such mixture is possible. It is plain that when bodies are mixed their colours also are necessarily mixed at the same time; and that this is the real cause determining the existence of a plurality of colours—not superposition or juxtaposition. For when bodies are thus mixed, their resultant colour presents itself as one and the same at all distances alike; not varying as it is seen nearer or farther away.

Colours will thus, too be many in number on account of the fact that the ingredients may be combined with one another in a multitude of ratios; some will be based on determinate numerical ratios, while others again will have as their basis a relation of quantitative excess. And all else that was said in reference to the colours, considered as juxtaposed or superposed, may be said of them likewise when regarded as mixed.

Why colours, as well as savours and sounds, consist of species determinate and not infinite is a question which we shall discuss hereafter.

5.4 Text 4: On Colours

- It has been debated since the Renaissance whether Aristotle or Theophrastus was the genuine composer of the work. It has also been suggested that a student of Theophrastus, Staro of Lampsacus, was the author (Gottschalk 1964). The work could be a part of a larger work, but neither this question nor the question of authorship can be clearly answered (Aristoteles 1999, pp. 42-51). The text has been translated by T. Loveday and E. S. Forster. The text reproduced here is from (Barnes 1984). ©The Jowett Copyright Trustees.

Chapter 1

Simple colours are those which belong to the elements, i.e. to fire, air, water, and earth. Air and water in themselves are by nature white, fire (and the sun) yellow, and earth is naturally white. The variety of hues which earth assumes is due to dying, as is shown by the fact that ashes turn white when the moisture that tinged them is burnt out. It is true they do not turn a pure white, but that is because they are tinged by the smoke, which is black. And this is the reason why lye-mixture turns yellow, the water being coloured by hues of flame and black.
Black is the proper colour of elements in process of transmutation. The remaining colours, it may easily be seen, arise from blending by mixture of these.

Darkness is due to privation of light. For we see black under three different conditions. Either the object of vision is naturally black (for black light is always reflected from black objects); or no light at all passes to the eyes from the object (for an invisible object surrounded by a visible patch looks black); and objects always appear black to us when the light reflected from them is rare and scanty. This last condition is the reason why shadows appear black. It also explains the blackness of ruffled water, e.g. of the sea when a ripple passes over it: owing to the roughness of the surface few rays of light fall on the water and the light is dissipated, and so the part which is in shadow appears black. The same principle applies to very dense cloud, and to masses of water and of air which light fails to penetrate; for water and air look black when present in very deep masses, because of the extreme rarity of the rays reflected, the parts of the mass between the illuminated surfaces being in darkness and therefore looking black. There are many arguments to prove that darkness is not a colour, but merely privation of light, the best being that darkness, unlike all other objects of vision, is never perceived as having any definite magnitude or any definite shape.

Light is clearly the colour of fire; for it is never found with any other hue than this, and it alone is visible in its own right whilst all other things are rendered visible by it. But there is this point to be considered, that some things, though they are not in their nature fire nor any species of fire, yet seem to produce light. So we cannot say that the colour of fire is identical with light, and yet light is the colour of other things besides fire, but we can say that this colour is to be found in other things besides fire, and yet light is the colour of fire. Anyhow, it is only by aid of light that fire is rendered visible, just as all other objects are made visible by the appearance of their colour.

The colour black occurs when air and water are thoroughly burnt by fire, and this is the reason why burning objects turn black, as e.g. wood and charcoal when the fire is put out, and smoke from clay as the moisture in the clay is all secreted and burnt. This is also why the blackest smoke is given off by fat and greasy substances like oil and pitch and resinous wood, because these objects burn most completely, and the process of combustion is most continuous in them.

Again, things turn black through which water percolates if they first become coated with lichen and then the moisture dries off. The stucco on walls is an example of this, and much the same applies to stones under water, which get covered with lichen and turn black when dried.

This then is the list of simple colours.
CHAPTER 2

From these the rest are derived in all their variety of chromatic effects by blending of them and by their presence in varying strengths. The different shades of crimson and violet depend on differences in the strength of their constituents, whilst blending is exemplified by mixture of white with black, which gives grey. So a dusky black mixed with light gives crimson. For observation teaches us that black mixed with sunlight or firelight always turns crimson, and that black objects heated in the fire all change to a crimson colour, as e.g. smoky tongues of flame, or charcoal when subjected to intense heat, are seen to have a crimson colour. But a vivid bright violet is obtained from a blend of feeble sunlight with a thin dusky white. That is why the air sometimes looks purple at sunrise and sunset, for then the air is especially dusky and the impinging rays feeble. So, too, the sea takes a purple hue when the waves rise so that one side of them is in shadow: the rays of the sun strike without force on the slope and so product a violet colour. The same thing may also be observed in birds’ wings, which get a purple colour if extended in a certain way against the light, but if the amount of light falling on them is diminished the result is the dark colour called brown, whilst a great quantity of light blended with primary black gives crimson. Add vividness and lustre, and crimson changes to flame-colour.

For it is after this fashion that we ought to proceed in treating of the blending of colours, starting from an observed colour as our basis and making mixtures with it. (But we must not assign to all colours a similar origin, for there are some colours which, though not simple, bear the same relation to their products that simple colours bear to them, inasmuch as a simple colour has to be mixed with one other colour to produce them.) And when the constituents are obscure in the compound product, we must still try to establish our conclusions by reference to observation. For, whether we are considering the blend which gives violet or crimson, or whether we are considering the mixtures of these colours which produce other tints, we must explain their origin on the same kind of principles, even though they look dissimilar. So we must start from a colour previously established, and observe what happens when it is blent. Thus we find that wine colour results from blending airy rays with pure lustrous black, as may be seen in grapes on the bunch, which grow wine-coloured as they ripen; for, as they blacken, their crimson turns to a violet. After the manner indicated we must treat all differences of colours, getting comparisons by moving coloured objects, keeping out eye on actual phenomena, assimilating different cases of mixture on the strength of the particular known instances in which a given origin and blending produce a certain chromatic effect, and verifying our results. But we must not proceed in this inquiry by blending
pigments as painters do, but rather by comparing the rays reflected from the aforesaid known colours, this being the best way of investigating the true nature of colour-blends. Verification from experience and observation of similarities are necessary if we are to arrive at clear conclusions about the origin of different colours, and the chief ground of similarities is the common origin of nearly all colours in blends of different strengths of sunlight and firelight, and of air and water. At the same time we ought to draw comparisons from the blends of other colours with rays of light. Thus charcoal and smoke, and rust, and brimstone, and birds’ plumage blent, some with firelight and others with sunlight, produce a great variety of chromatic effects. And we must also observe the results of maturation in plants and fruit, and in hair, feathers, and so on.

CHAPTER 3

We must not omit to consider the several conditions which give rise to the manifold tints and infinite variety of colours. It will be found that variations of tint occur either because colours are possessed by varying and irregular strengths of light and shade (for both light and shade may be present in very different strengths, and so whether pure or already mixed with colours they alter the tints of the colours); or because the colours blent vary in fullness and in powers; or because they are blent in different proportions. Thus violet and crimson and white and all colours vary very much both in strength and in intermixture and purity.

Difference of hue may also depend on the brightness and lustre or dimness and dullness of the blend. Lustre is simply continuity and density of light; e.g. we have a glistening gold colour when the yellow colour of sunlight is highly concentrated and therefore lustrous. That explains why pigeons’ necks and drops of falling water look lustrous when light is reflected from them.

Again, some objects change their colour and assume a variety of hues when polished by rubbing or other means, like silver, gold, copper, and iron, when they are polished; and some kinds of stones give rise to different colours, like . . . [the name of the stone has dropped out of the text] which are black but make white marks. This is because the original composition of all such substances is of small dense and black particles, but in the course of their formation they have been tinged, and all the pores through which the tincture passed have taken its colour, so that finally the whole material appears to be of that colour. But the dust that is rubbed off from them loses this golden or copper colour (or whatever the hue may be), and is quite black, because rubbing breaks up the pores through which the tincture passed, and black is the original colour of the substance. The other colour is no longer apparent because the dye is dissipated, and so we see the original natural colour of the material, and this is why these substances all appear black. But when rubbed against a smooth
and even surface, as e.g. against a touchstone, they lose their blackness and get back their other colour, which comes through where the lines of the tincture in the pores are unbroken and continuous.

In the case of objects burning, dissolving, or melting in the fire, we find that those have the greatest variety which are dark in colour and give off a thin hazy smoke, such as the smoke of brimstone or rusty copper vessels, and those which, like silver, are dense and smooth.

Apart from these cases, variety of hue is characteristic of all dark smooth objects, such as water, clouds, and birds’ plumage. For these last, owing to their smoothness and the variety of blends into which the impinging rays of light enter, show various colours, as also does darkness.

We never see a colour in absolute purity: it is always blent, if not with another colour, then with rays of light or with shadows, and so it assumes a tint other than its own. That is why objects assume different tints when seen in shade and in light and sunshine, and according as the rays of light are strong or weak, and the objects themselves slope this way or that, and under other differential conditions. Again, they vary when seen by firelight or moonlight or torchlight, because the colours of those lights differ somewhat. They vary also in consequence of mixture with other colours, for they are coloured by passing through one another. For if light falls on a given object and is coloured by it crimson or herb-green, and then the light reflected from that object falls on another colour, it is again modified by this second colour, and so it gets a new chromatic blend. This happening to it continuously, though imperceptibly, light when it reaches the eye may be a blend of many colours, though the sensation produced is not of a blend but of some colour predominant in the blend. This is why objects under water tend to have the colour of water, and why reflections in mirrors resemble the colour of the mirrors, and we must suppose that the same thing happens in the case of air. Thus all hues represent a threefold mixture of light, a translucent medium (e.g. water or air), and underlying colours from which the light is reflected. A translucent white medium, when of a very rare consistency, looks hazy in colour; but if it is dense, like water or glass, or air when thick, a sort of mist covers its surface, because the rays of light are inadequate at every point on it owing to its density, and so we cannot see the interior clearly. Air seen close at hand appears to have no colour, for it is so rare that it yields and gives passage to the denser rays of light, which thus shine through it; but when seen in a deep mass it looks practically dark blue. This again is the result of its rarity, for where light fails the air lets darkness through and looks dark blue. When densified, air is, like water, the whitest of things.
CHAPTER 4

All dyed things take their colour from the dye. Common sources of such coloration are the flowers of plants and their roots, bark, wood, leaves, or fruit, and again, earth, foam, and inks. Sometimes coloration is due to animal juices (e.g. the juice of the purple-fish, with which clothes are dyed violet), in other cases to wine, or smoke, or lye mixture, or to sea-water, as happens, for instance, to the hair of marine animals, which is always turned red by the sea. In short, anything that has a colour of its own may transfer that colour to other things, and the process is always this, that colour leaving one object passes with moisture and heat into the pores of another, which on drying takes the hue of the object from which the colour came. This explains why colour so often washes out: the dye runs out of the pores again. Furthermore, steeping during the dyeing produces a great variety of hues and mixtures, and these are also affected by the condition of the material itself; as has been said before in the case of blending. Even black fleeces are used for dyeing, but they do not take so bright a colour as white. The reason is that whilst the pores of the wool are tinged by the dye that enters them, the intervals of solid hair between the pores do not take the colour, and if they are white, then in juxtaposition to the colour they make the dye look brighter, but if they are black, they make it look dark and dull. For the same reason a more vivid brown is obtained on black wool than on white, the brown dye blending with the rays of black and so looking purer. For the intervals between the pores are too small to be separately seen, just as tin is invisible when bled with bronze; and there are other parallel cases.

These then are the reasons for the changes in colour produced by dyeing.

CHAPTER 5

As for hair and feathers and flowers and fruit and all plants, it is abundantly clear that all the changes of colour which they undergo coincide with the process of maturation. But what the origins of colour in the various classes of plants are, and what kinds of changes these colours undergo, and from what materials these changes are derived, and the reasons why they are thus affected, and any other difficulties connected with them—in considering all these questions we must start from the following premisses. In all plants the original colour is herb-green; thus shoots and leaves and fruit begin by taking this colour. This can also be seen in the case of rain-water; when water stands for a considerable time and then dries up, it leaves a herb-green behind it. So it is intelligible why herb-green is the first colour to form in all plants. For all water in process of time first turns yellow-green on blending with the rays of the sun; it then gradually turns black, and this further mixture of black and yellow-green produces herb-green. For, as has already been remarked, moisture becoming stale and drying up of itself turns black. This can be seen, for example, on the stucco of reservoirs;
here all the part that is always under water turns black, because the moisture, as it cools, dries up of itself, but the part from which the water has been drawn off, and which is exposed to the sun, becomes herb-green, because yellow mingles with the black. Moreover, with the increasing blackness of the moisture, the herb-green tends to become very deep and of a leek-green hue. This is why the old shoots of all plants are much blacker than the young shoots, which are yellower because the moisture in them has not yet begun to turn black. For, the growth being slow and the moisture remaining in them a long time, owing to the fact that the liquid, as it cools, turns very black, a leek-green is produced by blending with pure black. But the colour of shoots in which the moisture does not mix with the rays of the sun, remains white, unless it has lasted a long time and dried and turned black at an earlier stage. In all plants, therefore, the parts above the ground are at first of a yellow-green, while the parts under the ground, namely the stalks and the roots, are white. The shoots, too, are white as long as they are underground, but if the earth be removed from round them, they turn herb-green; and all fruit, as has been already said, becomes herb-green at first, because the moisture, which passes through the shoots into it, has a natural tendency to assume this colour and is quickly absorbed to promote the growth of the fruit. But when the fruit ceases to grow because the liquid nourishment which flows into it no longer predominates, but the moisture on the contrary is consumed by the heat—then it is that all fruit becomes ripe; and the moisture already present in it being heated by the sun and the warmth of the atmosphere, each species of fruit takes its colour from its juice, just as dyed material takes the hue of the dye. This is why fruits colour gradually, those parts of them which face the sun and heat being most affected; it is also the reason why all fruits change their colour with the changing seasons. This is evident; for all fruits, as soon as they begin to ripen, change from herb-green to their natural colour. They become white and black and grey and yellow and blackish and dusky and crimson and wine-coloured and saffron—in fact, assume practically every variety of colour. Since most hues are the result of the blending of several colours, the hues of plants must certainly also be due to the same blends; for the moisture percolating through the plants washes and carries along with it all the ingredients on which their colours depend. When this moisture is heated up by the sun and the warmth of the atmosphere at the time of the ripening of the fruit, each of the colours forms separately, some quickly and some slowly. The same thing happens in the process of dyeing with purple; when, after breaking up the shell and washing all the moisture from it, they pour it into earthenware vessels and boil it, at first no definite colour is noticeable in the dye, because, as the liquid boils more and more and the colours still remaining in the vessels mix together, each of the hues gradually undergoes a great variety of alterations; for black and white and brown and hazy shades appear, and finally the dye all turns purple, when the colours are sufficiently boiled up together; so as a
result of the blending no other colour is separately noticeable. This is just what happens with fruit. In many instances, because the maturing of all the colours does not take place simultaneously, but some colours form earlier and others later, changes from one to another take place, as in the case of grapes and dates. Some of these are crimson at first; but when black colour forms in them, they turn to a wine colour, and in the end they become a dark-bluish hue when the crimson is finally mixed with a large quantity of pure black. For the colours which appear late, when they predominate, change the earlier colours. This is best seen in black fruits; for broadly speaking most of them, as has already been remarked, first change from herb-green to a pinkish shade and become reddish, but quickly change again from the reddish hue and become dark blue because of the pure black present in them. The presence of crimson is proved by the fact that the twigs and shoots and leaves of all such plants are crimson, because that colour is present in them in large quantities; while that black fruits partake of both colours is clear from the fact that their juice is always a wine colour. Now the crimson hues come into existence at an earlier stage in growth than the black. This is clear from the fact that pavement upon which there is any dripping, and; generally speaking, any spot where is a slight flow of water in a shady place, always turns first from herb-green to a crimson colour, and the pavement looks as though blood had lately been shed over all the portion of it on which the herb-green colour has matured; then finally this also becomes very black and of a dark-bluish colour. The same thing happens in the case of fruit. That change in the colour of fruit occurs by the formation of a fresh colour, which ousts the earlier one, can easily be seen from the following examples. The fruit of the pomegranate and the petals of roses are white at first, but in the end, when the juices in them are beginning to be tinged as they mature, they alter their colours and change to violet and crimson hues. Other parts of plants have a number of shades, for example the juice of the poppy and the scum of olive oil; for this is white at first, as is the fruit of the pomegranate, but, after being white, it changes to crimson, and finally mingling with a large quantity of black it becomes a dark-bluish hue. So, too, the petals of the poppy are crimson at their ends, because the process of maturation takes place quickly there, but at their base they are black, because this colour is already predominant at the end; just as it predominates in the fruit, which also finally becomes black.

In the case of plants which have only one colour—white, for example, or black or crimson or violet—the fruit always keeps a single kind of colour, when once it has changed from herb-green to another colour. Sometimes the blossoms are of the same colour as the fruit—as, for instance, in the pomegranate, the fruit and blossoms of which are both crimson; but sometimes they are of very dissimilar hues—as, for example, in the bay-tree and the ivy, whose blossoms are always yellow, but their fruit respectively black and crimson. The same is
true of the apple-tree; its blossom is white with a tinge of pink, while its fruit is yellow. In the poppy the flower is crimson, but the fruit may be black or white, according to the different time at which the juices present in the plant ripen. The truth of the last statement can be seen from many examples; for, as has been said, some fruits come to differ greatly as they ripen. This is why the odours and flavours of flowers and fruits differ so much. This is still more evident in the actual blossoms. For part of the same petal may be black and part crimson, or, in other cases, part white and part purplish. The best example of all is the iris; for its blossom shows a great variety of hues according to the different states of maturation in its different parts, just as grapes do when they are already ripening. Therefore the extremities of blossoms always ripen most completely, whilst the parts near the base have much less colour; for in some cases the moisture is, as it were, burnt out before the blossom undergoes its proper process of maturation. It is for this reason that the blossoms remain the same in colour, while the fruit changes as it grows riper; for the former, owing to the presence of only a small amount of nutriment, soon mature, while the fruit, owing to the presence of a large quantity of moisture, changes as it ripens to all the various hues which are natural to it. This can also be seen, as has already been remarked, in the process of colour-dyeing. When in dyeing purple they put in the colouring matter from the vein of the purple-fish, at first it turns brown and black and hazy; but when the dye has been boiled sufficiently, a vivid, bright violet appears. So it must be from similar reasons that the blossoms of a plant frequently differ in colour from its fruit, and that some pass to a stage beyond, whilst others never attain to their natural colour, according as they do or do not mature thoroughly. For these reasons, then, blossoms and fruit differ from one another in their colouring. The leaves of most trees turn yellow in the end, because, owing to the failure of nutriment, they become dried up before they change to their natural colour; just as some of the fruits also which fall off are yellow in colour, because here too nutriment fails before they mature. Furthermore, corn and in fact all plants turn yellow in the end. This change of colour is due to the fact that the moisture in them no longer turns black owing to the rapidity with which it dries up. As long as it turns black and blends with the yellow-green, it becomes herb-green, as has already been said; but, since the black is continually becoming weaker, the colour gradually reverts to yellow-green and finally becomes yellow. The leaves of the pear-tree and the arbutus and some other trees become crimson when they mature; but the leaves even of these, if they dry up quickly, turn yellow, because the nutriment fails before
maturity is reached. It seems very probable then that the differences of colour in plants are due to the above causes.

CHAPTER 6

The hairs, feathers, and hides, whether of horses, cattle, sheep, human beings, or any other class of animals, grow white, grey, reddish, or black for the same reason. They are white when the moisture which contains their proper colouring is dried up in the course of maturation. They are black, on the other hand—as was the case in the other form of life—when, during their growth, the moisture present in the skin settles and becomes stale owing to its abundance, and so turns black; in all such cases skin and hide become black. They are grey, reddish, and yellow, and so on, when they have dried up before the moisture in them has completely turned black. Where the process has been irregular, their colours are correspondingly variegated. So in all cases they correspond in colour to the hide and skin; for when men are reddish in colouring, their hair too is of a pale red; when they are black, it is black; and if white leprosy has broken out over some part of the body, the hair on that portion is also always white, like the marking on dappled animals. Thus all hair and feathering follows the colour of the skin, both regional hair and hair which is spread over the whole body. So, too, with hoofs, claws, beaks, and horns; in black animals they are black, in white animals they are white, and always because the nutriment percolates through the skin to the outer surface. A number of facts prove that this is the true cause. For example, the heads of all young children are at first reddish owing to scanty nutriment; that this is so is clear from the fact that the hair of infants is always weak and thin and short at first; but as they grow older, the hair turns black, when the nutriment which flows into them settles owing to its abundance. So, too, with the pubes and beard; when the hair is just beginning to grow on the pubic region and chin, it also is reddish at first, because the moisture in it, being scanty, quickly dries up, but as the nutriment is carried more and more to those regions the hair turns black. But the hair on the rest of the body remains reddish for a considerable time owing to lack of nutriment; for as long as it is growing, it keeps on turning black like the pubes and the hair of the head. This is clear from the fact that hairs which have any length are generally blacker near the body and yellower towards the ends, because the moisture which reaches these parts of them is very scanty and soon dries up. The feathers, too, of black birds are in all cases darker near the body and lighter at the ends. The same is the case with the parts about the neck and, generally speaking, any part which receives scanty nutriment. This can be illustrated by the fact that before turning grey all hair changes colour and becomes reddish, because the nutriment again fails and dries up quickly; finally it becomes white, because the nutriment in it is completely matured before the moisture turns black. This is most evident in the case of beasts of burden; here the hair always
turns white, for in those parts because, owing to the feebleness of the heat, they cannot draw up as much nourishment as the rest of the body, the moisture quickly dries up and turns white. So men tend especially to turn grey in the region of the temples, and generally speaking in any part which is weak and ailing. So, too, white is the colour to which more than any other a change tends to take place in instances of deviation from natural colour. For example, a hare has been known before now to be white—while black hares have also been seen—and similarly white deer and bears have sometimes occurred; similarly white quials, partridges, and swallows. For all these creatures, when weak in their growth, come to maturity too soon owing to lack of nutriment, and so turn white. Similarly some infants at birth have white hair and eyelashes and eyebrows, a circumstance which normally occurs when old age is coming on and is then clearly due to weakness and lack of nutriment. Therefore in most classes of animals the white specimens are weaker than the black; for, owing to lack of nutriment, they mature before their growth is complete, and so turn white, just as does fruit when it is unhealthy; for fruit is still more apt to get mature through weakness. But when animals grow white and at the same time are far superior to the rest of their species, as is the case with horses and dogs, the change from their natural colour to white is due to generous nutriment. For in such animals the moisture, not settling long, but being absorbed in the process of growth, does not turn black. Such animals are soft and well covered with flesh, because they are well nourished, and white hairs, therefore, never change colour. This is clear from the fact that black hairs, when the nutriment in them fails and matures too completely, turn reddish before they grow grey, but finally turn white. Yet some people hold that hair always turns black because its nutriment is burnt up by heat, just like blood and all other substances; but they are in error, for some animals are black from birth—dogs, for example, and goats, and oxen, and, generally speaking, those creatures whose skin and hair get nutriment from the very first—but they are less black as they get older. If their supposition were correct this ought not to be the case, but it would necessarily follow that the hair of all animals would turn black at their prime, when heat predominates in them, and that they would be more likely to be grey at first. For in the beginning the heat is always much weaker than at the time when the hair begins to turn white. This is clear in the case of white animals also. Some of them are very white in colour at birth, those, namely, which at first have an abundance of nutriment, the moisture in which has not been prematurely dried up; but as they grow older their hair turns yellow, because less nutriment afterwards flows into it. Others are yellow at first and are whitest at their prime. Similarly birds change colour when the nutriment in them fails. That this is the case can be seen in the fact that in all these animals it is the parts round the neck, and, generally speaking, any parts which are stinted when the nourishment is scanty, which turn yellow; for it is clear that, just as reddish
colour turns black and vice versa so white turns yellow and vice versa. This happens also in plants, some of which revert from a later stage in the process of maturation back again to an earlier stage. The best illustration of this is to be found in the pomegranate. At first its seeds are crimson, as are also its leaves, owing to the small amount of nourishment which matures completely; afterwards they turn to a herb-green, because a quantity of nutriment flows into them and the process of maturation is less able to predominate than before; but in the end the nutriment does mature and the colour reverts to crimson.

To sum the matter up, in hair and feathers of every kind, changes always occur either—as has already been remarked—when the nutriment in them fails, or when, on the contrary, it is too abundant. Therefore the age at which the hair is at its whitest or blackest varies in different cases; for even ravens’ feathers turn yellow in the end, when the nutriment in them fails. But hair is never crimson or violet or green or any other colour of that kind, because all such colours arise only by mixture with the rays of the sun, and further because in all hairs which contain moisture the changes take place beneath the skin, and so they admit of no admixture. This is clear from the fact that no feathers have their distinctive colouring at first, but practically all gaily coloured birds start by being black—the peacock, for example, and the dove and the swallow; it is only inter that they assume all their varied colours, the process of maturation taking place outside their bodies in their feathers and wattles. Thus in birds, as in plants, the maturation of the colours takes place outside the body. So, too, the other forms of animal life—aquatic creatures, reptiles, and shell-fish—have all sorts and manners of colouring, because in them too the process of maturation is considerable.

From what has been set forth in this treatise one may best understand the theory of colours.

- **Hint to Problem 10** The treatise mentions (i) black light, (ii) lack of light entering the eye, and (iii) little reflection from objects as a source of black. These entail an intromissionist theory. While (i) presupposes only a qualitative difference between the colour black and other colours (and has often been regarded as a mistake in the manuscript), (ii) and (iii) suggest that the difference is quantitative: black is the privation or lack of light.

  Compare also with Aristoteles, de An. II.7, 418b18. It has been suggested that only two separate cases are listed (Aristoteles 1999).

- **Hint to Problem 11** The colour of the light reaching the eye depends on the hue of the light source (sun, moon, fire, lamp, etc.). It
can also depend on the nature and intensity of illumination and shade (dense and continuous light gives lustre). It may undergo further modification in the medium, and the kind of surface from which the reflected light reaches the eye. Several factors affect the surfaces of objects (polishing, rubbing). Even the angle and strength of impact on a coloured body before reflection can contribute to the colour perceived. The colour of the light source, the colour of the medium, and the colour of the object all play an important part. These give rise to nearly endless variations, and signify an attitude that is very different from most other theories, trying to attribute colour exclusively to either a physical property or to physiological states.

- **Hint to Problem 13** The original herb-green *poôdes* in the process of darkening becomes a very deep leek-green *prasoeides*. An example for maturation is water: with the rays of the sun it becomes greenish-yellow *chôron*, which is presumably the colour of algae (but see also chapter III.3), and the darkening of water (dry algea are often very dark or black) produces herb-green *poôdes*. With time moisture blackens if affected by sunlight—young plant organs are lighter (yellower), and subterranean parts can even be white. Sun and heat explain the colouring of flowers and fruits (these are not summarised here).

6 **Further Problems**

- **Problem 14** We have cited Myles Burnyeat’s view on the understanding of the transparent in Aristoteles’ *De Anima*. However, there are many debated issues concerning this question. Summarise and evaluate the views on the physiological basis of perception. For this use the articles by Sorabji, Nussbaum and Putnam, and Burnyeat in (Nussbaum & Putnam 1992), or in the 1995 paperback edition that also contains as an additional essay (Burnyeat 1995)—see also the version (Burnyeat 1993). Against (Burnyeat 1992) the main arguments defending a “functionalist” reading of Aristoteles are exposed in (Nussbaum & Putnam 1992), heavily relying on the perception-account of Richard Sorabji (Sorabji 1992). The debate is also insightfully treated in (Sisko 1998)—an excellent review of (Everson 1997), who does not take (Burnyeat 1995) into consideration. Reconstruct the arguments a) in favour or against a functionalist reading of
Aristotle; b) that try to determine what Aristotle means by ‘alteration’ in the case of vision.

- **Problem 15** There are interesting similarities but also significant differences between the sixth chapter of *On Colours*, and chapters 4-6 in the fifth book of Aristotle’s *Generation of Animals* (GA 5,4-6). Compare the two texts and list common features. Can you find an explanation that accounts for (at least part of) the differences?

## 7 Suggested Readings

The standard edition of Aristotle’s work is (Aristoteles 1984), but for bilingual works see the volumes of the *Loeb Classical Library*. For an introduction to the general thought of Aristotle (Lear 1988) is an excellent and enjoyable starting point. Earlier introductions, like (Ross 1949) also provide a wealth of information, but are harder to digest.

Several books, like (Barnes 1995) provide detailed and excellent bibliography of various areas.

For Aristotle’s views on perception see (Everson 1997). For the *De Anima* see also the translation and notes in (Hamlyn 1968). The commentary of Thomas Aquinas has appeared recently (Aquinas 1999)

About sense organs the PhD thesis re-written in (Johansen 1997) can be consulted together with earlier works, like (Lloyd & Owen 1978). On the ratios of colour mixing see the longish article by (Fine 1996) and (Sorabji 1972).

For a new edition of Alexander of Aphrodisias’ commentary on the rainbow see (Alexander 1996), for the later history of reception see (Schofield 1980), or the recent (Sharples & Wiesner 2001), and (Alexander 2000).

For a new translation of the treatise *On Colours* see (Aristoteles 1999). This edition offers detailed notes, a list of important bibliographical items and the Latin version of the commentary of Michael of Ephesus.
Introduction

Ancient Greek atomism arose in the fifth century BC as a response to problems of the continuum raised by Parmenides (see Chapter II) and other Eleatic philosophers. The original meaning of the word ‘atom’ or *atomon* was undividable, uncuttable and referred to the smallest constituents of matter that are not further divisible.

It was later developed by Epicurus and Epicureans, and also became popular in the Roman world. Although ancient physical atomism won numerous adherents in the Renaissance, its real revival was in the seventeenth century. Several early modern scientists and philosophers, like Pierre Gassendi (1592-1655) embraced atomistic ideas.

Atomism also had several opponents, in ancient times most importantly Sceptics, like Sextus Empiricus, the Stoics, and the Peripatetics, following Aristotle. They pinpointed several conceptual difficulties, but atomism remained a powerful alternative conceptual system.

This chapter investigates an early atomistic theory of sense perception by Democritus, followed by the relevant parts of one of the most detailed extant Epicurean texts on vision written by Lucretius.

Early Greek Atomism

Democritus was born in Abdera around 450 B.C. (see Figure II.1 on page 25). He lived to a very great age (dates unknown). He is said to have met Socrates in Athens. None of his works survived, apart from several quotations and bits and pieces in the doxographic tradition. He was the co-founder of atomism with Leucippus, but their precise relation cannot be now established.

For Democritus vision is the result of the impact of fast-moving films of atoms, also called *eidolon*, in plural *eidola*, or little picture(s). These are constantly emitted in streams by the surfaces of objects around us. Colour is due to the arrangement of atoms in the effluences from the objects.
This account of vision had been extended to explain other sensory processes as well, and, it seems, it also covered thought. The activity of the rational soul (thinking) was also explained on the basis of a series of pictures impinging on the soul.

Text 1 contains Theophrastuts’ reconstruction of the Democritean doctrine in the *De sensibus*.

### 3 Epicurus

Both Plato and Aristotle attacked basic assumptions of atomism in their philosophical writings. To a large extent the writings of Epicurus consists of defences of atomism against Aristotelian and other critiques.

As a development to the original atomistic theories Epicureanism in the third century B.C. posited that the physical particles called atoms (physically indivisible) were further dividable to irreducibly small magnitudes (these are mathematically indivisible).

Epicurus was born at Samos (for the places see figure on page 25, on page 25), an Athenian colony, in 341 B.C. At 18 he went to Athens as an *ephebe*, to spend two years in the national service—a prerequisite to become full citizen of Athens. During this time Samos ceased to be an Athenian colony, and Epicurus joined his parents in Colophon, near Ephesus in Asia Minor. In the following years he was teaching among others in Mytilene on Lesbos, in Lampsacus, and returned in 306 B.C. to Athens to set up his school in his own garden outside the city walls.

With the aim in mind of freeing humanity from anxieties, and especially the fear of death, he wanted to dispose of ignorant fears and superstitions. He gave a voluminous explanation of the origin and structure of the universe in the thirty-seven books of *On nature*. Some of the books of the lost treatise had been recovered in a spectacular archeological find in the eighteenth century—but probably just as many were destroyed when the explorers tried to open up the hundreds of carbonised papyri found in a Herculanean villa (near Naples, see figure II.2 on page 26) flooded by lava from the Vesuvius in 79 A.D.

The books were probably part of a lecture course for early Epicureans, but their length prohibited the wide-scale circulation of all thirty-seven books. Instead, shorter letters by Epicurus, like his *Letter to Herodotus, Letter to Phytocles*, or his *Great Epitome* enjoyed greater popularity.
Epicurus’ teaching consists of three interdependent components: Physics, Ethics, and Canonics (from the Greek kanon, “rule”). While it is not an atheistic philosophy, for Epicureans the gods have no interest in human affairs.

According to his views on physics the infinite universe is made up of matter and void. Imperishable atoms (infinite in number and homogeneous, but with an upper maximum of size) make up matter. They move very swiftly, straight downwards in their natural motion, but at random times they may “swerve”—thus giving rise to a non-deterministic cosmos. As everything in the world is made up of atoms, so are the souls/minds of men.

The Canonics include epistemological views and theories of sensation and perception. By stating that sense data provide a true and accurate picture of external reality, sensation will serve as an ultimate source and criterion of truth. Thus sensory experience (just like e.g. feelings) cannot be overruled. This account, while laying a foundation for knowledge-claims, leaves perceptual errors, illusions, and well-attested weaknesses of the senses in need of explanation (see next section).

A main aim of Epicurus’s ethics is the avoidance of non-necessary pain (like the fear of death) Epicurus among others strongly rejects the notion of eternal torment after death.

To explain sensation Epicurus states that that external objects send off emanations (films of atoms that are like little images or “idols”, eidola) of themselves that travel through the medium and impinge upon our senses. The imprint of these images on the senses gives rise to sensation.

The more specialized works on vision by Epicurus are lost¹, but the basic theory can be reconstructed from his Letter to Herodotus and from the fragments.

Visual perception occurs when, similarly to the general Epicurean account of sensation, the pupil is struck by a film (or films) of atoms (eidolon). The films originate from the outer layer of bodies, which peel off from them. The peeling off results in a train of thin images (eidola) emitted in all directions. They maintain their shape and their atomic arrangements in their travels.

---

¹Books I and III from his Physics, as well as his On Seeing, Visual Forms, Great Epitome, etc.
As to the question of colour Epicurus departs from the road taken by some earlier atomists, and claims that perceptible qualities are real (even if not existing *per se*). He states that certain arrangements of the elementary particles result in the perception of certain colours.

- **Problem 1** Discuss, based on the excerpts in Text 2 why a principally *moral* philosophy, like that of Epicurus emphasizes knowledge about natural phenomena. What is Epicurus’s main argument against scepticism?

- **Problem 2** Compare the treatment of sensation of early atomists, like Democritus, who hold a strict separation of primary and secondary qualities, believing that secondary qualities are “unreal” and are only based on custom (“colour is by convention, sweet by convention, a compound by convention . . . what is real are the void and the atoms”, fr. B125) with that of Epicurus and of Plato, who claims that colour perceived is a function of the shape of the atoms.

## 4 Lucretius

Titus Lucretius Carus, born at the very beginning of the first century B.C. (around 98-95 B.C.), is the author of the most elaborate and poetic exposition of Epicureanism known to us.

The times were very violent throughout his life. After the Social War from 91-88 B.C. the commander Sulla marched on Rome, and a bloody Civil War broke out. He became dictator in 82 B.C. executing thousands of important Roman citizens. The year 71 B.C. marks the outbreak of the famous slave revolt lead by Spartacus.

Epicureanism naturally developed in the interim, and Greek philosophy in the Latin-speaking world significantly proliferated in the first century B.C. Important figures of the Platonic tradition (like Philo of Larissa) worked in Rome, where they influenced people like Cicero. Pythagorianism also experienced increased attention, and manuscripts of Aristotle’s school were brought to Rome. With interests renewed, old philosophical problems and debates also resurfaced—the arguments often buttressed by new empirical findings. But interestingly these new debates hardly ever occur in Lucretius’
It thus seems that while living centuries after Epicurus and writing in a different language, Lucretius is a fairly reliable source of information concerning the views of Epicurus—much more so than concerning the views of the Epicurean school in general or the philosophical environment he lived in.

While most of his philosophical message in the *De Rerum natura* is closely linked to the teachings of Epicurus, a large portion of the stylistic choices and the genre show the influence of Empedocles. In Chapter II Text 3 we have seen the ‘multiple-correspondence simile’ where the eye’s structure (and debatably also function) is treated as analogous to that of a lantern. And just as for Empedocles, for Lucretius as well the listing of correspondences has more than descriptive purposes: the strengthening of the analogy also makes the argument more persuasive.

Greek philosophical schools often employed specific terminology the rendering of which into Latin posed several difficulties. A common practice was *transliteration*, rewriting the Greek term in Latin. While a large number of loan-words in Latin do have Greek origins (like the name of many disciplines: philosophy, rhetorics, physics, etc), in a philosophical epic the amassing of newly latinised words was understandably not the preferred praxis. Another option is to use a Latin equivalent, to translate a technicality with a technicality. But, instead, Lucretius usually employs metaphors, using them to illuminate a single original Greek term.

Apart from Lucretius’ literary influence the effect of his poem on early modern thought can hardly be underestimated. The sixteenth century revival of atomism and corpuscularianism was largely shaped by ancient atomist ideas. Pierre Gassendi (1592-1655) published his *Syntagma Philosophiae Epicuri* in 1649, an elaboration of Epicurean science (Gassendi 1649). Other followers like Walter Charleton (Charleton 1654 (1966)) directly influenced the views of the young Newton. The most significant thinkers of the seven-

---

2 One such debate was the location of the mind. According to Epicurus the rational mind is in the chest (Aristotle and the early Stoics would agree, Pythagoreans and Plato would disagree with this view). But later Alexandrian physicians based on information from the vivisection of prisoners discovered the nervous system and recognised the brain as the location of the mind. To further defend the Epicurean view or to modify it the challenge has to be seen, appreciated, and answered. But, interestingly here and on a score of other occasions Lucretius seems ignorant of the newer problems, and concentrates on the explication of the original Epicurean doctrine in a fundamentalist manner (Sedley 1998).
teenth century, Spinoza, Leibniz, and others were also influenced by these views.

4.1 The Poem De Rerum Natura

The poem *De Rerum Natura* (DRN) consists of six books, in different stages of completion. The first two lay out the basic doctrine of Epicurean atomism. In the first book, after an invocation to Venus, Lucretius starts the explication of the Epicurean system (atomism, infinity of the universe, existence of bodies and void) and refutes the systems of other philosophers. Beginning with Parmenides he states that nothing comes out of nothing (see Chapter II.), and the book ends with the description of the collapse of the world. The next book after lyric passages celebrating philosophy explains atomic motion and shapes and argues that the atoms do not have secondary qualities. The Epicurean doctrine of the infinite number of worlds is also treated in this book.

The two middle books concentrate on man, the soul, sensation, and passions, proving the materiality and mortality of the mind and soul. The third book argues that the human soul is mortal, and that as the soul dies, death should not be feared. The fourth book is a detailed account of perception, especially of vision and argues against scepticism, ending with an analysis of sexual love—seen as a mental delusion.

The last two books describe the world, the origin of life, civilisation, and religion, cosmic phenomena. The fifth book concentrates on Epicurean cosmology and sociology, with the (evolutionary) stages of life on earth and the origin and development of civilization. The sixth book, apart from treating phenomena belonging to meteorology gives the grim details of the great plague of Athens (430 BC), a memento of mortality.

Text 3 is the first part of the fourth book of the poem. After the poetic introduction Lucretius discusses the existence and character of the images. The Greek *eidolon* is most commonly translated as *simulacrum* (plural as *simulacra*), but at times also as *imago, figura*, or *effigies*. These words together with the rich metaphors collectively convey the meaning of the Greek term.

- **Problem 3** Describe the multiple correspondence simile that supports the existence of the images. What characterises the thin layers of atoms (do they have e.g. colour)? No hint is given to this problem.
The ingenious model of thin films of atoms necessarily faces a number of problems: how can people look into each other’s eyes, how can the simulacra fit the pupil, etc. The physiology and psychology of vision, and well-known optical phenomena all have to be accounted for within this framework. Lucretius selectively tackles these.

- **Problem 4** When discussing mirrors and mirroring (from Text 3), which aspects of catoptrics are investigated and which neglected?

The images have power to travel inconceivable space within a point of time—as they are a) small and b) pushed from behind. Small things pushed move far away, and the extreme rarity of the images means that they can propagate in the air—even exceeding the speed of light.

Upon entering the eye the images move through the pupil (for Democritus the pupil was the place of seeing) and transmit the form and hue of the object.

- **Problem 5** What is the connection of visual and tactile stimuli? What properties of objects are perceived? How is distance perceived? How is a person seen through a window? Answer these questions using Text 3, without further hints.

Bright light sources, like the sun emit strong effluences with fire particles that can even cause pain. Physiological conditions can also affect things seen: Lucretius mentions the millennium-old mistake that people with jaundice see everything in yellow. After some remarks on how perception occurs, the poem discusses misperceiving and gives a detailed description of illusions.

- **Problem 6** List the causes and the types of illusions.

Our eyeballs cannot know the nature of reality—so Lucretius states that in controversial cases the reasoning of mind (using non-controversial sense-impressions) serves to interpret the seen. The sceptical arguments are rejected in spite of the numerous illusions.

- **Problem 7** How does Lucretius argue that the sceptic stance involves a vicious circle? What is his comment on the possible contradiction between sense-data from different sense organs? Why is
perception to be ultimately trusted? Evaluate his argument. No hint is given to this problem.

5 Texts

5.1 Text 1: Democritus on Vision

The text of Theophrastus’ reconstruction of the Democritean theory of vision is from (Stratton 1917).

DS 49-50 Democritus in his account of sense perception does not make it entirely clear whether it is due to contrast or to similarity. For in so far as he ascribes the action of the senses to an alteration, it would seem to depend on contrast; for the like is never altered by the like. On the other hand, sense perception would seem to depend on similarity in so far as he ascribes the perceptive process and, in a word, alteration to the fact that something is acted upon. For things that are not the same cannot be acted upon, he says; but even when things that are different do act, <their action is> not due to their difference but to the presence in them of something identical. Upon such matters as these he may consequently be understood either way. He now undertakes to discuss the <senses> each in turn.

Vision he explains by the reflection <in the eye>, of which he gives a unique account. For the reflection does not arise immediately in the pupil. On the contrary, the air between the eye and the object of sight is compressed by the object and the visual organ, and thus becomes imprinted; since there is always an effluence of some kind arising from everything. Thereupon this imprinted air, because it is solid and is of a hue contrasting <with the pupil>, is reflected in the eyes, which are moist. A dense substance does not receive <this reflection>, but what is moist gives it admission. Moist eyes accordingly have a better power of vision than have hard eyes; provided their outer tunic be exceedingly fine and close-knit, and the inner <tissues> be to the last degree spongy and free from dense and stubborn flesh, and free too, from thick oily moisture; and provided, also, the ducts connected with the eyes be straight and dry that they may “perfectly conform” to the entering imprints. For each knows best its kindred.

DS 73-78 The simple colours, he says, are four. What is smooth is white; since what neither is rough nor casts shadows nor is hard to penetrate,—all such substances are brilliant. But brilliant substances must also have open passages and be translucent. Now white substances that are hard have the structure just described, - for instance, the inner surface of cockle shells; for the substance here would be shadowless, ‘gleaming’, and with straight passages. But the
white substances that are loose and friable are composed of round particles, yet with these placed oblique to one another and oblique in their conjunction by pairs, while the arrangement as a whole is uniform in the extreme. With such a structure these substances are loose because their particles are in contact only over a small portion of their surface; friable, because their composition is so uniform; shadowless, because they are smooth and flat. But those substances are whiter, compared with one another, in which the figures are more exactly as described above and are freer from admixture with other figures and whose order and position more nearly conform to the given description. From such figures, then, is white derived.

Black is composed of figures the very opposite to those of white,—figures rough, irregular, and differing from one another. For these cast shadows, and the passages amongst them are not straight nor easy to thread. Their effluences, too, are sluggish and confused; for the character of the effluence also makes a difference in the inner presentation, as this emanation is changed by its retention of air.

Red is composed of figures such as enter into heat, save that those of red are larger. For if the aggregations be larger although the figures are the same, they produce the quality of redness rather than of heat. Evidence that redness is derived from such figures is found in the fact that we redden as we become heated, as do other things placed in the fire until they have a fiery colour. Those substances are redder that are composed of large figures—for example, the flame and coals of green wood are redder than those of dry. And iron, too, and other things placed in fire become redder. Those are most luminous, however, that contain the most fire and the subtilest, while those are redder that have coarser fire and less of it. Redder things, accordingly, are not so hot; for what is subtile is hot.

Green is composed of both the solid and the void,—the hue varying with the position and order of these constituents.

Such are the figures which the simple colours possess; and each of these colours is the purer the less the admixture of other figures. The other colours are derived from these by mixture.

Golden and copper-colour and all such tones, for instance, come from white and red, their brilliance being derived from the white, their ruddiness from the red component; for in combination the red sinks into the empty spaces of the white. Now if green be added to white and red, there results the most beautiful colour; but the green component must be small, for any large admixture would not comport with the union of white with red. The tint will vary according to the amount of green that is introduced.

Crimson comes from white, black, and red,—the largest ‘portion’ being red, that of black small, and of white midway; for thus it makes an appearance delightful to the sense. That black and red are present in it is patent to the eye:
its brilliance and lustre testify to the presence of white; for white produces such effects.

Woad hue is composed of deep black and golden green, but with the major ‘portion’ black. Leek green is of crimson and woad, or of golden green and purplish. . . . For sulphur colour is of this character, with a dash of brilliance. Indigo is a mixture of woad and fiery red, with round figures and figures needle-shaped to give a gleam to the colour’s darkness.

Brown is derived from golden green and deep blue; but if more of the golden green be mixed, flame-colour is the result; for the blackness is expelled because <the golden green> is shadowless. And red, too, when mixed with white, gives almost a ‘pure’ golden green, and not a black; which accounts for the fact that plants at first are of such a green before there is a heating and dispersion.

This completes the tale of colours he recounts; although he holds that the colours, like the savours, are endless in number according to their combinations,—according as we remove some and add others and ‘combine’ them in varying proportion. For no one of these colours would be the same as another.

5.2 Text 2: Epicurus: Principal Doctrines

- The Principal Doctrines by Epicurus are translated by Robert Drew Hicks. The forty maxims were preserved in the Lives of the Philosopher of Diogenes Laertius, one of the most important sources concerning early Greek philosophy. All maxims can be read under: http://classics.mit.edu/Epicurus/princdoc.html.

1. A happy and eternal being has no trouble himself and brings no trouble upon any other being; hence he is exempt from movements of anger and partiality, for every such movement implies weakness.

3. The magnitude of pleasure reaches its limit in the removal of all pain. When pleasure is present, so long as it is uninterrupted, there is no pain either of body or of mind or of both together.

11. If we had never been molested by alarms at celestial and atmospheric phenomena, nor by the misgiving that death somehow affects us, nor by neglect of the proper limits of pains and desires, we should have had no need to study natural science.

22. We must take into account as the end all that really exists and all clear evidence of sense to which we refer our opinions; for otherwise everything will be full of uncertainty and confusion.

23. If you fight against all your sensations, you will have no standard to which to refer, and thus no means of judging even those judgments which you pronounce false.
24. If you reject absolutely any single sensation without stopping to discriminate with respect to that which awaits confirmation between matter of opinion and that which is already present, whether in sensation or in feelings or in any immediate perception of the mind, you will throw into confusion even the rest of your sensations by your groundless belief and so you will be rejecting the standard of truth altogether. If in your ideas based upon opinion you hastily affirm as true all that awaits confirmation as well as that which does not, you will not escape error, as you will be maintaining complete ambiguity whenever it is a case of judging between right and wrong opinion.

• **Hint to Problem 1** Even though the Epicurean philosophy is principally a moral one, to reach its aim (see maxims 1. and 3.), uninterrupted pleasure and happiness, it needs to get rid of fears. As celestial and atmospheric (in general natural) phenomena can disturb us and give rise to unfounded fear (see maxim 11.), understanding them is important for a fearless life—hence the the great detail of natural philosophy in the Epicurean system.

If we were sceptical about our sensations, we would loose the certainty (maxim 22.) we need for living happily in life. 23. and 24. argue that scepticism undermines our standards of truth. To be unable to choose between right and wrong opinion, for Epicurus, is an obstacle to reaching *ataraxia*, or peace of mind.

• **Hint to Problem 2** Epicurus strongly reacted against the earlier atomist distinction between primary and secondary qualities. The world of senses was seen as a guarantee of reality. Through sensory experience one can even understand the laws of nature (hidden behind her “face”). Even the inaccessible world of atoms can reveal their true behaviour. As Lucretius suggests in his poem (II. 116-124), the motion of dust particles in a ray of light reveals (models) the movement of atoms:

> An image, a type goes on before our eyes  
> Present each moment; for behold whenever  
> The sun’s light and the rays, let in, pour down  
> Across dark halls of houses: thou wilt see  
> The many mites in many a manner mixed  
> Amid a void in the very light of the rays,
And battling on, as in eternal strife,
And in battalions contending without halt,
In meetings, partings, harried up and down.
From this thou mayest conjecture of what sort
The ceaseless tossing of primordial seeds
Amid the mightier void - at least so far
As small affair can for a vaster serve,
And by example put thee on the spoor
Of knowledge...

However, in general, the senses can testify against a certain, faulty conception— their role is more falsificatory than verificatory (Lucretius Carus 1997).

5.3 TEXT 3: LUcretius: De Rerum Natura IV.

- The poem can readily be downloaded from the internet: http://classics.mit.edu/Carus/nature-things.html (translated by William Ellery Leonard). This file is based on (Lucretius Carus 1921). A number of other, newer translations are also available, see the Suggested Readings. A copy can be reached via the homepage of the WTWG Department of the University of Berne: http://www.philoscience.unibe.ch/bernstudies/titel.html. To answer the questions in the chapter, please read lines 1-719 of Book IV, with special attention to lines 26-519.

- Hint to Problem 4 Shape and colour-constancy are mentioned in case of the mirrors, together with the swiftness of the appearance of the mirror image, and the existence of concave mirrors. The reversal of the right and left sides is explained through the example of inverting a mask, in which case, seen from the other side, the left eye becomes the right, etc. (but why the top and bottom does not, has puzzled many for centuries). The equality of the angles of incidence and reflection is also mentioned, but not explained. The behaviour of reflected images from curved (i.e. concave) mirrors is either a case of double reflection (not easily tenable), or arises from the spinning of the image (also far from satisfactory). Convex mirrors are not mentioned.
• **Hint to Problem 6** Angles of travelling *eidola* are blunted by long travel and continuous collisions, so they can appear as obtuse or might not even be perceived, thus giving rise to shape-distortion. The movement of our shadow is understood if we realise that it is not an entity in itself, but simply air without light. The trust in senses does not rule out that in certain cases the mind has to decide which of the interpretations equally well fitting certain sense-impressions is “real”. If in motion, other stationary objects seem to be moving, but this is explainable as the correct impression one has if one is moving. Similarly the displacement of far-away objects is so small that for the casual observer they seem stationary.

Lucretius also mentions mirages, the oculogyral illusion (see also pages 153 and 159), and the distance illusion of the seeming closeness of the horizon sun. Other traditional topoi are mentioned like the depth illusion of water mirroring the sky, the relative motion illusion of objects in a water current, the perspectival illusion, the peculiar setting and rising of the sun “from water”, the illusion of the broken oar, when immersed in water, and a number of other instances including disturbed stereoscopic vision and even dreams.

6 **Further Problems**

• **Problem 8** Discuss the system of colour by Democritus in Text 1. What are the characteristic conditions where they appear? Compare this list and the constitution of mixed colors with the system Plato discussed in Chapter III. (See also Problem 3 of Chapter III.)

• **Problem 9** The Lucretian (Epicurean) theory of colours is found in earlier sections of the poem. Discuss the account in Book II (especially lines 757-809!).

7 **Suggested Readings**

A classical study on the subject is (Bailey 1928), or the later (Bailey 1964). For more on the atomic theory as it appeared in Plato’s *Timaeus* and in Democritus’ works, see (Nikolaou 1998)

For a reproduction of the 1887 edition of Epicurean writings by Hermann Usener, see (Epicurus 1967), for an English translation by Cyril Bailey,
(Epicurus 1926) or (Epicurus 1970). For an introduction to the Epicurean philosophy, see e.g. (Sharples 1996), for Epicurus’ teachings on the peace of mind, or ataraxia, see (Warren 2002). For his “scientific” method see (Asmis 1984), and on the connection between him and Lucretius, (Clay 1983).

The translation by William Ellery Leonard can be read and downloaded from http://classics.mit.edu/Carus/nature−things.html. A number of recent translations are available, like the classic (Lucretius Carus 1975), the one by Sir Ronald Melville (Lucretius Carus 1997), Anthony M. Esolen’s work (Lucretius Carus 1995), or the translation by Martin Ferguson Smith (Lucretius Carus 2001). An exciting and new approach to the text of Lucretius’ poem is outlined in (Sedley 1998). On the dating of the poem see (Hutchinson 2001).

Chapter VI

THE MATHEMATICAL TRADITION

1  INTRODUCTION

In his *Physics* Aristotle described the subject matter of optics: “while geometry investigates natural lines but not *qua* natural, optics investigates mathematical lines, but *qua* natural, not *qua* mathematical.” (Phys. II,2 194a 7-12)

Optics was considered as one of the more natural branches of mathematics, and for centuries was subdivided into three main areas: *optics* investigated vision and the behaviour of unaltered visual rays and rays of light, *catoptrics* dealt with reflection and mirroring, while *dioptrics* analysed the behaviour of rays in different media, in air, water, glass; phenomena that we would generally consider as cases of refraction. The optical treatises, however, often incorporated very practical elements of knowledge, like the production of burning mirrors, building instruments that display illusions, or can be used for military, religious, or other purposes.

There is meagre direct textual evidence about mathematical treatments of optics for the period before 350 B.C. but, surprisingly, also for the years between 300 and 100 B.C.

This chapter investigates some of the basic principles of Euclid’s *Optics*, the theory of mirroring in Hero’s *Catoptrics*, and segments from the culmination of the Greek science of mathematical Optics, the *Optics* of Ptolemy.

A common fundamental starting point for all the investigated Greek theories is that some kind of (physical) contact has to occur between the object of vision and the eye. For the mathematical optical tradition this contact is not inside the eye, but at the object. There an interaction takes place between the visual ray and the surface of the object. This establishes a strong parallel between tactile and visual sensations.

Another common feature of the works belonging to this tradition is that rectilinearly travelling rays leave the eye, thus enabling a geometrical analysis of vision. The rays form a cone, the apex of which is at the eye. For Euclid the cone is a collection of these discrete rays, for Ptolemy it is continuous, and the individual ray is non-existent, but can serve as a useful simplification.
2 The Beginnings of Mathematical Optics

Democritus (460-357 B.C.) is said to have written—from his near seventy treatises—at least four on optics, but none of these survive. From the four optical treatises by Theophrastus (out of the nearly 230 titles) only the *De sensibus* quoted in Chapters II and V survives—containing a critical commentary and not the original views of Theophrastus. Surprisingly enough the earliest extant Greek optical texts are found scattered in Plato’s dialogues and the Aristotelian corpus (Smith 1999). The most detailed geometrical proof from this time concerning the shape and size of the rainbow was cited in Chapter IV Text 1, and will be analysed in detail in Chapter XII.
3 Euclid

Euclid of Alexandria (flourished around 300 B.C., see figure VI.1) is the author of our first detailed geometrical optical treatise. Very little is known about his life. Probably after having studied at Plato’s Academy, he left Athens for Alexandria following the call of Ptolemy I to teach at the newly founded university in Alexandria. He had enormous influence on the later development of optics and mathematics. The third century B.C. was Alexandria’s intellectual heyday. Hosting the largest library of the world at that time, it continued to be an important intellectual and cultural center throughout the period investigated in this chapter.

His Optics is structured as other mathematical treatises, like his famous Elements. It sets out with a number of definitions based on an extromissionist starting point. These are followed by propositions that are proved using the definitions. Most of the propositions investigate the apparent size of geometrical figures and objects.

- **Problem 1** Discuss to what extent is Euclid’s treatment “mathematical”. If other factors play an important role, which are these? Why?

4 Hero

Finding the dates for the life of Hero of Alexandria (also called Heron), writer of numerous treatises from geometry to mechanics, is a long-debated issue. Opinions range from 150 B.C. (as Hero does not quote from works written later than Archimedes) to 250 A.D. (some sources suggest that he lived after Ptolemy), and as the name was very common, sorting out the relevant references is a very difficult task. The generally accepted opinion today is, following Otto Neugebauer, that he witnessed an eclipse in Alexandria in 62 A.D.

A large number of his works survived, but the authorship of some are disputed. The books include studies of puppet theatres, war-engines, catapults, pneumatic devices, measurement of two- and three-dimensional objects, geometry, and, among a number of other treatises, a work on mirrors and mirroring, the Catoptrica.

---

¹Some attribute it to Ptolemy, but here I discuss it as a work of Hero.
Figure VI.2: The plate is from the first German edition of Vitruvius: De Architectura Libri Decem, entitled: Des allernamhaftigsten und hocherfahrensten Römischen Architecti und Kunstreichen Werck oder Bawmeysters und künstlichem Bawen (1548). Refraction in water, the rainbow, burning lens are shown, together with inverting mirrors.

His work is very often what we would now call “applied science”, and as Pappus (appr. 290-350), a later Alexandrian describes in his Mathematical Collection (Thomas 1941):

The mechanicians of Heron’s school say that mechanics can be divided into a theoretical and a manual part; the theoretical part is composed of geometry, arithmetic, astronomy and physics, the manual of work in metals, architecture, carpentering and painting and anything involving skill with the hands.

…the ancients also describe as mechanicians the wonder-workers, of whom some work by means of pneumatics, as Heron in his Pneumatica, some by using strings and ropes, thinking to imitate the movements of living things, as Heron in his Automata and Balancings,
...or by using water to tell the time, as Heron in his *Hydria*, which appears to have affinities with the science of sundials.

Dioptrics remained a significant part of optics, but even in the Renaissance the stock examples continued to be the ones that puzzled and bemused the Ancient world—see figure VI.2.

- **Problem 2** Why is catoptrics an important discipline for Hero? And why is it a part of the study of vision?

- **Problem 3** How does Hero argue for the equality of the angles of incidence and reflection?

5 PTOLEMY

Very little is known about Claudius Ptolemy’s life. He lived in the first three quarters of the 2nd century A.D. Probably born in upper Egypt he lived most of his life, like the previous two mathematicians, in Alexandria, and became one of the most gifted scientists of the period. His works had profound influence on the development of later Arabic and European thought. His *Mathematike Syntaxis*, also known as *Almagest*, a part of the title of the Arabic translation, became the most important source book for mathematical astronomy until the scientific revolution. His *Tetrabiblos* was the most popular astrological treatise in Antiquity, and his *Planetary Hypotheses* described a universe, which dominated Islamic and western medieval astronomical thought. His *Geography* also had enormous impact on Islamic geographers, and it would be natural to think that his *Optics*, undisputably the best account up to his time, was equally successful.

But, surprisingly enough, this is far from the case. It was hardly ever mentioned in Antique sources, although it exerted significant influence on later Arabic opticians, like Alhazen. But the translation of the single known copy to Latin (see also Chapter I) was soon overshadowed by the translation of Alhazen’s work, and was hardly known in the seventeenth-eighteenth centuries. Rediscovered in the nineteenth century it was finally printed only in 1885!

Ptolemy probably had access to many of the sources cited in this and previous chapters: Euclid’s and Hero’s works, the account of vision in Plato’s

---

2So scant is the evidence on the Greek original, that it has been seriously questioned whether the author was in fact Ptolemy (Knorr 1985).
Timaeus (which he rejected), and Aristotle’s De Anima. He amalgamated many aspects of this composite tradition.

The aim of the book is to explain visual perception, and, following Euclid it is an extromissionist account. It tackles several problems using geometry, like reflection and refraction, but a significant part of the work is on aspects of vision that are non-mathematical, like the perception of colour, visual acuity, and many illusions.

5.1 The Optics

The lost first book is about the interaction of light and the visual flux, as the beginning of the second book tells us (Text 3, para. [1]). The second book analyses vision, both its objective and subjective mode—that is both the objects of sight and the perception of colour, size, shape and motion. The third book is a detailed study of reflection from plane and convex mirrors, while the fourth is on reflection from concave mirrors. After these catoptrical studies, the fifth book tackles dioptrics, the study of refraction.

Our selection of texts includes the introductory remarks from the second book [II. 1-29] in Text 3, a detailed study of illusions [II. 83-142] in Text 4, and some of Ptolemy’s work on refraction from the fifth book [V. 1-18, 31]. To make the navigation in the texts easier, a detailed table of contents of the relevant parts are included before each text.

- **Problem 4** The first 29 paragraphs of Book II refer to a number of theories mentioned earlier in the book. Try collect these implicit references, and note how Ptolemy supports or criticizes them.

- **Problem 5** What are visible properties (intrinsically, primarily, secondarily) for Ptolemy? How does this taxonomy effect the traditional distinction between primary and secondary qualities, and special and common sensibles?

A significant part of the second book is concerned with the systematic study of illusions. Ptolemy follows a taxonomy in grouping the illusions: illusions are either common to all senses, or special to the visual faculty, in which case they can arise either in the sense itself, or in the judgement of the mind [II. 84]. Although a number of these misperceptions or illusions would not be called such, some of the observations and their explanations are remarkable. The colours of quickly rotating discs are investigated [95],
and the effect of quick movement on colour-perception is paralleled with
the effect of high spatial frequencies of colour [96]. Coloured afterimages
are also mentioned. The oculogyral illusion is described [121]—the con-
tinuous movement of objects that is experienced after spinning in one di-
rection and stopping, or as a result of vertigo or before fainting. Also a
number of correlations are set up between size-distance-movement per-
ception and the visual faculty. To name just a few: with age we see nearer,
while deeply set eyes can see further [86]. More moisture in the eye also
helps farsight. We tend to see closer movement [100] and the movement
of smaller objects [101] as faster. Brighter objects are seen as closer [124],
while dimmer objects look more remote, and also larger [126]. More remote
objects are also seen as larger [126].

While a number of these are considered unscientific or even false, oth-
ers have stood the test of time and are still important in our investigation
of the visual system, as Chapter X will show.

- **Problem 6** Give a systematised list of the optical illusions Ptolemy
  lists. Use the short introduction above, and the detailed table of con-
  tents. Once you have the list circle the illusions that would count even
today as illusions. Find contemporary names for these. (No further
  hint is provided for this problem.)

- **Problem 7** By studying Ptolemy’s analysis of visual illusions, what
can we infer about his understanding of visual perception? How
  many stages does it have?

The fifth book of the *Optics* is the study of *dioptrics*, i.e. image-displace-
ment as a result of refraction. Some such phenomena were sources of
wonder and also concern in Antiquity, like the seemingly broken oar sub-
merged into water [II. 120]. Similarly to reflection, the apparent position
of an object has to be explained in terms of the real position of the object
and the circumstances that deflect the visual ray. The location of the image
depends in both cases on the relation of the angles of incidence and of
reflection or refraction. But while in the first case the rather simple law
of the equality of angles had been established centuries before Ptolemy
(and we see a discussion of it in Hero’s work in Text 2), the mathematical
relationship has not been clarified for the latter.
Refraction was often attributed to imperfect reflection. Reflection takes place from hard, resisting surfaces, while refraction takes place if the surface is penetrable [V. 1-2].

- **Problem 8** List the principles that both reflection and refraction share in the comparison of Text 5.

- **Problem 9** What is the quantitative relationship between the angles of incidence and refraction in the two cases (air to water, air to glass)? How general are the conclusions drawn?

The rest of the fifth book investigates the problems of image location (also its importance in astronomy), and of image-distortion.

6  TEXTS

6.1  TEXT 1: EUCLID’S *Optics*

- As (Cohen & Drabkin 1966) states: “the *Optics* of Euclid is extant in two versions, of which the earlier form is thought to be Euclid’s own arrangement and the later that of Theon of Alexandria (latter part of the fourth century A.D.). The work consists of definitions (or rather assumptions) followed by 68 theorems geometrically demonstrated and constituting a treatise on perspective. It is the earlier version, as edited by Heiberg, from which the present translations have been made.” The text is from (Cohen & Drabkin 1948), maintaining some of the footnotes and the original notation (like “∴” for *therefore*).

*Let it be assumed*

1. That the rectilinear rays proceeding from the eye diverge indefinitely;
2. That the figure contained by a set of visual rays is a cone of which the vertex is at the eye and the base at the surface of the objects seen;
3. That those things are seen upon which visual rays fall and those things are not seen upon which visual rays do not fall;
4. That things seen under a larger angle appear larger, those under a smaller angle appear smaller, and those under equal angles appear equal;
5. That things seen by higher visual rays appear higher, and things seen by lower visual rays appear lower;
6. That, similarly, things seen by rays further to the right appear further to the right, and things seen by rays further to the left appear further to the left;
7. That things seen under more angles are seen more clearly\(^3\).

\(^3\) The angle referred to is that at the vertex of the cone (Definition 2.)
PROPOSITION I

No visible object is seen completely at one time.

Let $AD$ be a visible object, $B$ the eye, $BA$, $BG$, $BK$, and $BD$ visual rays from $B$ to the object. Then, since the incident rays move at an interval from one another, they cannot fall continuously over $AD$. Hence there are intervals along $AD$ upon which the rays will not fall. The whole of $AD$ will, therefore, not be seen at one time. We think that we see the whole of $AD$ at one time because the rays move along the object very quickly.

PROPOSITION II

Of equal magnitudes situated at a distance those that are nearer are seen more clearly. Let $B$ be the eye, and $GD$ and $KL$ the visible objects, which we are to consider as equal and parallel, $CD$ being nearer the eye. Let $BG$, $BD$, $BK$, and $BL$ be incident visual rays. The visual rays to $KL$ will not pass through points $G$ and $D$. For if they did, in the resulting [hypothetical] triangle, $BDLKGB$, $KL$ would be larger than $GD$. But they were assumed to be equal. Therefore, $GD$ will be seen by more visual rays than will $KL$. $GD$ will, consequently, be seen more clearly than will $KL$. For objects seen under a larger number of angles are seen more clearly.

PROPOSITION III

For every object there is a distance at which it is no longer seen. Let $B$ be the eye and $GD$ the visible object. I say that at a certain distance $GD$ will no longer be seen. For suppose that $GD$ is situated in an interval, $K$, between visual rays. Hence none of the visual rays from $B$ will fall upon $K$. But an object upon which visual rays do not fall is not seen. Therefore, for each object there is a distance at which it is no longer seen.

PROPOSITION IV

Of equal intervals on the same straight line those seen from a greater distance appear smaller.
Let AB, BG, and GD be equal intervals on the same straight line. Draw AE perpendicular to this line; let the eye be at E. I say that AB will appear larger than BG, and BG larger than GD.

Let EB, EG, and ED be incident visual rays. Draw BZ through B parallel to GE. AZ = ZE, for since BZ was drawn parallel to one side, GE, of △ AEG, it follows that EZ:ZA = GB:BA.

Hence, as we have said, AZ = ZE.

But BZ > ZA.

Therefore BZ > ZE.

and \( \angle ZEB > \angle ZBE \).

But \( \angle ZBE = \angle BEG \).

Therefore \( \angle ZEB > \angle BEG \).

Consequently \( AB \) will appear larger than \( BG \).

Similarly, if a parallel to \( DE \) be drawn through \( G \), it may be shown that \( BG \) will appear larger than \( GD \).

**Proposition V**

*Equal magnitudes situated at different distances from the eye appear unequal, and the nearer always appears larger.*

**Proposition VI**

*Parallel lines when seen from a distance appear to be an unequal distance apart.*

Let AB and GD be two parallels, and E be the eye. I hold that AB and GD seem to be an unequal distance apart, and that the interval between them at a point nearer the eye seems greater than at a point more remote from the eye.

Let EB, EZ, ET, ED, EH, and EK be visual rays. Draw BD, ZH, and TK.

Now since \( \angle BED > \angle ZEH \), BD appears greater than ZH.

Again, since \( \angle ZEH > \angle TEK \), ZH appears greater than TK.

That is, BD > ZH > TK in appearance.

The intervals, then, between parallels will not appear equal but unequal\(^4\).

\(^4\)The theorem of convergence is fundamental in the theory of perspective. There follows a proof of convergence for the case where the eye is not in the same plane as the parallels.
PROPOSITION VII

*Equal but non-contiguous intercepts on the same straight line if unequally distant from the eye appear unequal...*

PROPOSITION XLV

*There is a common point from which unequal magnitudes appear equal.*

Let BG be greater than GD. About BG describe a segment of a circle greater than a semicircle, and about GD describe a segment of a circle similar to that about BG, i.e., a segment containing an angle equal to that contained in segment BZG. The segments, then, will intersect, let us say at Z. Draw ZB, ZG, and ZD.

Since angles inscribed in similar segments are equal, the angles in segments BZG and GZD are equal. But things seen under equal angles appear equal. Therefore, if the eye is placed at point Z, BG will appear equal to GD. But BC > GD. There is, then, a common point from which unequal magnitudes appear equal.

PROPOSITION XLVIII

*To find points from which a given magnitude will appear half as large or a fourth as large, or, in general, in any fraction in which the angle may be divided.*

Let magnitude AZ be equal to BC. Describe a semicircle about line AZ and inscribe right angle K therein.

Let line BC be equal to AZ and around BC describe a segment of a circle such that an angle inscribed therein will be half of angle K. Then angle K is double angle D, and AZ will, therefore, appear twice as large as BC when the eye is on circumferences AK2 and BCD, respectively.

---

5The process may be repeated infinitively, the magnitude appearing 1/4, 1/8, etc. as large.
• **Hint to Problem 1** From the first assumption it is clear that Euclid does not give a purely “mathematical” treatment of optics. Although the anatomy of the eye and other physiological considerations are not discussed, optics was seen as a discipline that included the explanation of visual sensation.

As can be seen, Euclid adopts an *extromissionist* theory of vision—but this does not affect the geometric development of the main object of Euclid’s work, a theory of perspective.

Other elements of the work show a conflict between geometrical assumptions and perceptual experience. The discontinuity of visual rays in Proposition I., II., III., for example, is in line with perceptual experience (not everything is seen that is in our visual field—one can look for an object long before one recognises that it was right in front of him).

6.2 **Text 2: Hero’s *Catoptrics***

• The *Catoptrics* of Hero of Alexandria is our earliest extant work on mirrors. Plato’s *Timaeus* contains a short description of mirroring, and a work ascribed to Euclid is probably a later compilation by Theon of Alexandria. The original Greek is lost, but we have a thirteenth century Latin translation which is probably Hero’s work, even though for centuries it was ascribed to Ptolemy. The selection of texts (ch. 1-6, 7, 10, 15, 18) is from (Cohen & Drabkin 1948), maintaining some of the footnotes.

1. The science of vision is divided into three parts: optics, dioptics and catoptrics. Now optics has been adequately treated by our predecessors and particularly by Aristotle, and dioptics we have ourselves treated elsewhere as fully as seemed necessary. But catoptrics, too, is clearly a science worthy of study and at the same time produces spectacles which excite wonder in the observer. For with the aid of this science mirrors are constructed which show the right side as the right side, and, similarly the left side as the left side, whereas ordinary mirrors by their nature have the contrary property and show the opposite sides. It is also possible with the aid of mirrors to see our

---

6See texts 2 and 3 from Chapter IV. (G.Z.).
own backs\textsuperscript{7}, and to see ourselves inverted, standing on our heads, with three eyes, and two noses, and features distorted, as if in intense grief. The study of catoptrics, however, is useful not merely in affording diverting spectacles but also for necessary purposes. For who will not deem it very useful that we should be able to observe, on occasion, while remaining inside our own house, how many people there are on the street and what they are doing? And will anyone not consider it remarkable to be able to tell the hour, night or day, with the aid of figures appearing in a mirror? For as many figures appear as there are hours of the day or of the night, and if a [given] part of the day has passed a [given] figure will appear. Again, who will not be astonished when he sees, in a mirror, neither himself nor another, but whatever we desire that be seen? Such, then, being the scope of the science, I think it necessary and proper to describe the views held by my predecessors, that my account may not be incomplete.

2. Practically all who have written of dioptrics and of optics have been in doubt as to why rays proceeding from our eyes are reflected by mirrors and why the reflections are at equal angles. Now the proposition that our sight is directed in straight lines proceeding from the organ of vision may be substantiated as follows. For whatever moves with unchanging velocity moves in a straight line\textsuperscript{8}. The arrows we see shot from bows may serve as an example. For because of the impelling force the object in motion strives to move over the shortest possible distance, since it has not the time for slower motion, that is, for motion over a longer trajectory. The impelling force does not permit such retardation. And so, by reason of its speed, the object tends to move over the shortest path\textsuperscript{9}. But the shortest of all lines having the same end points is the straight line.

That the rays proceeding from our eyes move with infinite velocity may be gathered from the following consideration. For when, after our eyes have been closed, we open them and look up at the sky, no interval of time is required for the visual rays to reach the sky. Indeed, we see the stars as soon as we look up, though the distance is, as we may say, infinite. Again, if this distance were greater the result would be the same, so that, clearly, the rays are emitted with infinite velocity. Therefore they will suffer neither interruption, nor curvature,

\textsuperscript{7}The meaning might possibly be “to see those who are behind us”, but see below, point 15.

\textsuperscript{8}The meaning seems to be that as long as an object moves very swiftly it retains its rectilinear motion. A similar notion with respect to projectile motion prevailed until the time of Galileo.

\textsuperscript{9}The point seems to be that if visual rays were not propagated in straight lines they would not traverse a given distance as swiftly as possible.
nor breaking, but will move along the shortest path, a straight line.

3. That our vision is directed along a straight line has, then, been sufficiently indicated. We shall now show that rays incident on mirrors and also on water and on all plane surfaces are reflected. Now the essential characteristic of polished bodies is that their surfaces are compact. Thus, before they are polished, mirrors have some porosities upon which the rays fall and so cannot be reflected. But these mirrors are polished by rubbing until the porosities are filled by a fine substance; then the rays incident upon the compact surface are reflected. For just as a stone violently hurled against a compact body, such as a board or wall, rebounds, whereas a stone hurled against a soft body, such as wool or the like, does not (for the projecting force accompanies the stone and then, in the case of the hard obstacle, gives way, not being able to accompany the stone any further or move it forward, while in the case of the soft obstacle, the force merely slackens and is separated from the stone), so the rays that are emitted by us with great velocity, as we have shown, also rebound when they impinge on a body of compact surface. Now in the case of water and glass not all such rays are reflected since both these substances have irregularities, composed as they are of units having minute parts, and of solid particles. For in looking through glass and water we see our own reflection and also what lies beyond the surface of the glass or water. That is, in the case of standing water, we see what is at the bottom, and in the case of glass, what lies beyond its surface. For those rays which fall upon solid bodies are themselves turned back and reflected, while those which penetrate through porous bodies enable us to see that which lies beyond. Hence images reflected from such bodies are imperfectly seen because not all the visual rays are reflected to the objects, but some of them, as we have indicated, are lost through the pores.

4. That rays incident upon polished bodies are reflected has, then, in our opinion, been adequately proved. Now by the same reasoning, that is, by a consideration of the speed of the incidence and the reflection, we shall prove that these rays are reflected at equal angles in the case of plane and spherical mirrors. For our proof must again make use of minimum lines. I say, therefore, that of all incident rays [from a given point] reflected to a given point by plane

---

10 The reference is not to reflection or refraction but to a change in the continuous rectilinear character of the motion.

11 A type of atomism, not very clearly defined, is invoked here to aid the explanation.

12 That is, the angle of incidence will equal the angle of refraction.

13 That is, the sum of the incident ray from eye to mirror and the reflected ray from mirror to object must be a minimum.
and spherical mirrors the shortest are those that are reflected at equal angles; and if this is the case the reflection at equal angles is in conformity with reason.

Consider AB a plane mirror, G the eye, and D the object of vision. Let ray GA be incident upon this mirror. Draw AD, and let \( \angle EAG = \angle BAD \). Let another ray GB also be incident upon the mirror. Draw BD. I say that \( GA + AD < GB + BD \).

Draw GE from G perpendicular to AB, and prolong GE and AD until they meet, say at Z. Draw ZB.

Now \( \angle BAD = \angle EAG \), and \( \angle ZAE = \angle BAD \) (as vertical angles).

\[ \therefore \angle ZAE = \angle EAG. \]

And since the angles at E are right angles,

\[ ZA = AG \]
and \( ZB = BG \).

But \( ZD < ZB + BD \)
and \( ZA = AC, ZB = BG \).

Therefore \( GA + AD < GB + BD \).

Now \( \angle EAG = \angle BAD \),
and \( \angle EBG < \angle EAG \),
and \( \angle HBD > \angle BAD \).

Therefore \( \angle HBD \) is, a fortiori, greater than \( \angle EBG \).

5. . . . In general, then, in the case of mirrors [both plane and spherical], one must consider whether there is or is not a point from which incident rays may be reflected at equal angles in such a way that the ray incident from the organ of vision and the ray reflected to the object of vision, when added together, make a sum less than that of all other pairs of rays similarly incident and reflected.

6. In the case of plane mirrors there is a place at the covering of which an image will no longer be seen.

Let AG be a plane mirror . . . , B the eye, and D the visible object. Draw AD and BG perpendicular to the mirror, and divide AG at H in such a way that \( AD:BG = AH:HG \). I say, then, that if H is covered, the image of D is no longer seen.

\[ ^{14} \text{That is, not only is the path of the ray shortest when the angles of incidence and reflection are equal, but there is only one incident ray which can be reflected at equal angles.} \]

\[ ^{15} \text{But not reflected at an angle equal to the angle of incidence. There are probably lacunae in the text as we have it, but the reference seems to be to a test whether the image of a given object will be seen in a mirror (plane or spherical) from a given fixed point.} \]
For if BH and HD are drawn, the triangles, because of the proportionality of their sides, will be similar. Hence \( \angle E = \angle Z \) and D will be visible through point H. Therefore, if this point is covered with wax or some other material, D will no longer be visible. If, however, the covering at H is removed from the mirror, the image will again appear in the mirror. For all rays incident upon a mirror will be reflected at equal angles.

7. In the case of plane mirrors the reflected rays neither will converge nor are parallel. . . .

10. In case of concave mirrors, when the eye is situated at the circumference the reflected rays converge. . . .

Let BGA be a concave mirror, and let the eye be placed at B. Let BG and BA be incident rays and GX and AN reflected rays. I say that GX and AN will meet on the side of X and N.

For since arc AB > arc GB, \( \angle Z > \angle T \).

\[ \therefore \angle E > \angle H. \]

\[ \therefore \angle L > \angle K \text{ (as remainders)}. \]

But \( \angle M > \angle L \). \[ \therefore \angle M > \angle K, \] and, consequently, GX and AN will meet on the side of N and X.

15. It is desired to secure the same effect [of surprising the observer] by another construction. Let ABG be a right triangle. Bisect BG at T. Let ZH and DE be plane mirrors on lines AG and AB, respectively. Consider TF as an observer with the eye at point T capable of looking into either mirror as desired. And so the problem will be solved.

If one mirror (ZH into which the observer looks) is kept unmoved while the other (DE, behind the observer) is moved up and down, the ray will reach a point where the image of the heel\(^{16}\) of the observer will appear in the mirror and he will think that he is flying.

\(^{16}\)K2 in figure VI.3.
To place a mirror so that one approaching it sees neither his own image nor that of another but only the image which we select.

Let $AB$ [figure VI.4] be the wall where the mirror is to be put and let the mirror be inclined to it at a given angle. If this angle is one-third of a right angle the measurements will be suitable. Let $BG$ be the surface of the mirror and let $BD$ be perpendicular to $AB$. $D$, the point on $BD$ at which the eye is, is so situated that a perpendicular drawn from it to $BG$ falls outside $BG$. Let this perpendicular be $ED$. Draw $DG$ to the end $G$ of the mirror and let $\angle EGD = \angle BGH$. If then, a visual ray from the eye $D$ falls on $G$, the end of the mirror, it will be reflected to $H$. Now let $HN$ be drawn from $H$ at right angles to $DB$. Now let $DT^{17}$ be another incident ray and draw $HT$.

$\therefore \angle BTH > \angle ETD^{18}$ and $\angle BTK = \angle GTD$.

Therefore $TK$ intersects $HN$, as do all rays incident upon the mirror when reflected.

Now let a plane [mirror] $LM$ be drawn parallel to mirror $GB$ and intersected by a ray reflected from that mirror. Clearly, then, the eye will see only that which lies within $HN$, since all the reflected $WS$ fall within $HN$. Therefore, if we place whatever object we wish near plane $LM$, those approaching will see not their own image but merely that of the aforesaid object. It will consequently be necessary, as we have said, to place $LM$ within $HN$ so that the object in question may be between the parallel plane mirrors$^{19}$ . . .

---

$^{17}$T being any point on $BG$.

$^{18}$For $\angle BTH > \angle BGH$, and $\angle ETD < \angle EGD$. But $\angle BGH = \angle EGD$. Therefore $\angle BTH > \angle ETD$.

$^{19}$The rest of the paragraph, which is not given here, deals with the placing of the mirrors in a temple so that the apparition may be seen by someone approaching it. The figure shows
• **Hint to Problem 2** Hero considers *catoptrics* a topic not adequately treated before. He states that Aristotle wrote extensively on *Optics*. This could either refer to the Texts 2 and 3 from Chapter IV, or additionally to a treatise on optics that ancient lists of Aristotle’s works mention, but is not extant today.

The subject matter of *catoptrics* is, apart from providing diverting spectacles—an aspect that has remained to characterise the study of optical phenomena—to enable the construction of highly utilitarian devices. The list Hero gives in the first chapter shows how technically motivated the study of this mathematical science was.

As the ancient mathematical tradition in optics subscribed to an *extramissionist* account of vision, the study of the properties of mirrors necessarily contributed to the understanding of how visual rays behave, and formed an integral part of the study of vision. This means that one of catoptrics’ fundamental tasks was to account for image displacement—why an object appears where it is not.

• **Hint to Problem 3** Hero claims that the equality of angles has not been explained previously. To do so, he sets out to prove that the visual ray travels in a straight line. One possible way of reconstructing the argument is as follows:

Figure VI.4: Hero’s *Catoptrics*: seeing selected images
P1: Impelling force on an object ⊃ object has no time for slower motion
P2: Object has no time for slower motion ⊃ no time for longer trajectory
(Implicit Premise 1: if an object with constant velocity travels on a longer trajectory between two points, then its motion will be slower, i.e. it passes from A to B in longer time.)
P3: No time for longer trajectory ⊃ the object will travel the shortest possible distance.
(Implicit Premise 2: A very swiftly moving object is impelled by a force.)
C1: A very swiftly moving object will travel the shortest possible distance.
P4: The visual rays proceed with infinite velocity (as experience teaches us)
C2: The visual rays will move along the shortest path, a straight line.

Of course other reconstructions are equally possible. This argument will provide the rationale for maintaining that light travels in a straight line, but will not be sufficient to show why the angles of incidence and reflection are equal. This latter would only be the case if the equal angles would always correlate with the shortest distance travelled. Hero states in point 4.: “I say, therefore, that of all incident rays [from a given point] reflected to a given point by plane and spherical mirrors the shortest are those that are reflected at equal angles; and if this is the case the reflection at equal angles is in conformity with reason.”

In some concave mirrors, however, this will not always be the case: the shortest rays will not be the ones reflected at equal angles. An example is shown on the left: a very deep elliptical mirror, where Hero’s explanatory principle seems to fail (Park 1997).

6.3 Text 3: Ptolemy’s Optics, Book II, Introduction
- The American Philosophical Society has kindly permitted to reprint substantial sections from Mark A. Smith’s recent translation of Ptolemy’s Optics (Smith 1996). From Text 3-5 below many of the original footnotes have been omitted or abridged without special note. Some new ones have been included with (G.Z.) at the end to show their origin.
In the preceding book we have explained everything that one can gather about what enables light and visual flux to interact, how they assimilate to one another, how they differ in their powers and operations, what kind of essential difference characterizes each of them, and what sort of effect they undergo. In this book, and in those that follow, however, we shall assemble the facts that pertain to the sensible action of the visual faculty according, as is fitting, to an order and sequence of propositions, first discussing in particular what the visible properties are and what characterizes each of them, along with their first and last distinctions, so that how the sense [of sight] operates may become more obvious.

Accordingly, we say that the visual faculty apprehends corporeity, size, color, shape, place, activity, and rest. Yet it apprehends none of these without some illumination and something [opaque] to block the passage [of the visual flux]. We need say no more on this score but must instead specify what characterizes each of the visible properties.

We contend, therefore, that these visible properties exist in two ways, one of which depends upon the disposition of the visible property [itself] and the other upon the action of the visual faculty. So let us speak separately of each, and let us start with the way that is determined by the disposition of the visible properties, some of which are intrinsically visible, some primarily visible, and some secondarily visible.

Now luminous compactness is what is intrinsically visible, for objects that are subject to vision must somehow be luminous, either in and of themselves or from elsewhere, since that is essential to [the functioning of] the visual sense; visible objects must also be compact in substance in order to impede the visual flux, so that its power may enter into them rather than pass through without incident effect. Thus, it is impossible for anything to be seen without these two conditions being met, nor [can anything be seen] when one of them is met without the other.

On the other hand, colors are primarily visible, because nothing, besides light, that does not have color is seen. Still, colors are not intrinsically visible, since colors are somehow contingent on the compactness of bodies and are not visible per se without light. Indeed, colors are never seen in darkness, except for [the color of] an object that shines from inherent whiteness or that is exceedingly polished, for each of these is a case of brightness, and brightness is a kind of luminosity.

All the rest of the aforementioned visible properties are secondarily visible, because the visual faculty apprehends things as bodies by means of...
their [inherent] colors and characteristics, whereas objects that have no compactness, but are exceedingly tenuous and have no color, are neither sensed nor perceived as bodies by the visual faculty. Furthermore, size, place, and shape are perceived only through the mediation of bodies’ surfaces, which coincide with the colors upon which external light falls. Activity and rest, as well, are apprehended by means of an alteration, or lack thereof, in any of the aforementioned visible properties.

[7] While the faculty of sight apprehends illuminated colors immediately, it apprehends the rest of the visible properties by means of such illuminated colors, not insofar as they are colors but only insofar as they have boundaries. For the visual faculty apprehends shapes and dimensions by means of the boundaries of the colored object, while place is apprehended by means of its location. The visual faculty also apprehends the motion or rest of these same colors by means of their change or lack thereof. And the motion or rest of shapes, dimensions, and location is perceived by means of the motion or rest of the boundaries or places of the colored object. For instance, an object that appears white certainly does not appear round, small, near, or stationary on account of its whiteness, just as the characteristics of a resisting object that the hand feels are not grasped through its hardness. On the contrary, these properties are discerned by means of either the boundary of whiteness or the boundary of hardness—the boundary in this case being an essential characteristic [of that hardness] or its size according to its extended nature.

[8] And that is why the hand apprehends differences among the resisting objects it feels and why the eye apprehends differences among colors in and of themselves without something else’s intervening. However, what characterizes the remaining [visible properties] is only perceived by means of these two [primary qualities], since those [remaining] properties are not apprehended except through something that is contingent on something else. For the apprehension of those contingent properties depends on a perception of sensibles that pertains specifically to the sense. Indeed, because they are inseparable from the bodies [in which they inhere], shapes and sizes are perceived by means of boundaries; but place, activity, and rest are perceived because they are separable [from] and conveyed [by such basic properties].

[9] This is why, when we are surrounded by clear air and open our eyes, we do not perceive the shape of the nearby air enveloped by the visual cone, because the air’s superadded color is extremely subtle and does not have enough intensity to render it sensible. But when the nearby air enveloped by the visual cone is condensed, then we perceive its color and shape, because its color extends far into its interior and becomes more substantial and evident, as is the case with water.

[10] Furthermore, if we take a liquid of some color and add to it another [medium of] identical color for the purpose of using that identical color to draw
images and shapes in the liquid, then the colors of the drawings will be seen together with the surrounding liquid of the same color, but their shapes will not be seen, nor their distances, nor the relative differences in size among them. Now, everything pertaining to colors appears in some way or another by means of the adjoining boundaries of dissimilar[ly colored] objects, just as the location of defining outlines appears by means solely of the lines separating them; and that [apparent location] is created either by contrasts among all the colors or by the correct imitation [of separation] that appears through [linear] drawings.

[11] Likewise, the outlining of shapes [within the liquid] causes a change in the uniformity of the representations, but we do not on that account say that the representations are actually seen. As we have said, first, since the continuous color of the defining lines, the color of the edges, and the colors of all the drawings in no way differ [among each other] through any change in the quality of their color, they happen to lack [distinction] by virtue of sameness. Also, as far as the observation of their shape is concerned, since no defining line appears in any of these figures, and because of the absolute uniformity [of color] that obtains among these different shapes, some color is needed to disclose their outlines. Finally, nothing about them will appear clearly without color. On the contrary, according to differences among colors alone, the shape of dissimilar colors, being intrinsic to them, must appear along with them. The same condition holds in each of the remaining objects that are seen along with a corresponding one—all of which involves a false perceptual inference.

[12] So too, since light and visual flux strike the surfaces of bodies together, it is quite appropriate that the first thing to be sensed in all visible objects is a characteristic of their surfaces. And color is more properly attributed to the surface than to the interior of things. For this reason, the ancients used to equate surface and color, because color is a certain property affixed to the substance of an illuminated thing, and the genus “surface” is like that; and so it is an apt designation for it. As far as the remaining visible properties are concerned, corporeity is not surface, because surface is its boundary, yet all the remaining visible properties depend upon something in bodies having to do with surface. For instance, size is the boundary of a surface’s quantity, whereas shape is a qualitative arrangement of surface, and place is the boundary of the location of a surface. Activity, though, depends upon surface insofar as it is attributed to any of those properties [i.e., size, shape, and place]—for instance, the activity of alteration, or of growth, or of diminution, or of locomotion.

[13] A [sole] proper sensible can be found that is appropriate to each of the senses; e.g., the quality of “resisting the hand” for touch, savors for taste, sounds for hearing, and odors for smell. But among the things that are common to the senses according to the origin of nervous activity, sight and touch share in all except color, for color is perceived by no sense but sight. Thus, color must be the proper sensible for sight, and that is why color is taken to be what is
primarily visible after light.

[14] In view of this fact [that color requires light to render it actually visible], it seems that color is not really a proper sensible, as certain people have supposed, claiming that color is something accidental to the visual flux and to light and that it has no real subsistence, because none of the [proper] sensibles needs anything extrinsic [to make it sensible], whereas colors need light. Hence, it seems that this missing subsistence is provided by the visual flux, not by the visible objects. For objects that are seen and that are affected by the visual flux are not of such a nature as to appear to the visual faculty without light.

[15] However, [if we grant] that the subsistence of different colors, as well as their generation, depends solely upon a change in light and in the visual flux, it follows that [all] bodies that maintain precisely the same location with respect to a given luminous source and a given viewpoint should appear to be of the same color. Yet we find the majority of such bodies to be of various colors. We find this to be so not only in the case of several objects, but also in the case of a single object that remains perfectly fixed in relation to both the luminous source and the viewpoint. [Take], for example, the animal called a chameleon, or the redness that arises in certain people from blushing, or the pallor that overtakes others from fear. Furthermore, what happens in these cases occurs without any noticeable change appearing in the things themselves or in the external conditions, except for a change in color.

[16] It is therefore obvious from what we have said that color truly inheres in these objects and belongs to them by nature, and it is seen only when light and visual rays combine to make it effective. And that is what has prompted the claim that nothing proper to any color is ever seen, since no color may be seen without light somehow shining [upon it]: either something that is self-luminous or something that is rendered luminous in some similar way should highlight colors by mixing with them at the surface, because [light] is generically related to none of the visible properties but color. It relates to itself, however, as [if providing] “form” to the “matter” of color.

[17] It is for this reason that objects are seen more or less clearly, not only because of variation in the condition of the visual flux, but also because of internal variation and, even more important, because of variation in the condition of the objects that illuminate the visual field. In each of the two types [of radiation], moreover, this variation is a function sometimes of the intensity of their powers and sometimes of the quality of their operations.

[18] As far as variations in intensity are concerned, an object is seen more clearly when more visual flux impinges upon it or when more light shines upon it—e.g., what is looked at directly is seen more clearly than what is looked at by means of reflection or refraction. Even so, an aggregation of [uninterrupted] visual rays is weakened when they are extended out to a great distance. Also,
what is looked at with both eyes is seen more clearly than what is looked at with either eye alone. And what is self-luminous is seen more clearly than what is illuminated by something else. Also, the larger or more numerous the luminous sources, the more clearly the object upon which their light shines is seen.

[19] On the other hand, among objects whose appearance depends upon the quality of [radiative] effects, those that lie directly in front of, and at right angles to the rays are seen more clearly than those that do not. For everything that falls orthogonally strikes its subjects more intensely than whatever falls obliquely. Also, what is polished is seen more clearly than what is rough, because there is disorder in a rough object resulting from the fact that its parts are not arranged in a regular way. But the parts of a polished object have a certain regularity, and [so] brightness is inherent to it. Dense objects, as well, are more clearly seen than rare ones, because rare bodies give way to the impinging [ray], whereas dense bodies resist it. Objects that radiate [light] by themselves are also seen more clearly than those that are lit by something else, and so, for example, is an object that is seen in illuminated, rarefied air. Furthermore, objects that lie a moderate distance from the viewpoint appear more clearly [than those that do not], because objects that are [too] near the eye are enveloped by visual cones that fall within the internal humor of the eye and that impede the visual flux. On the other hand, objects that lie far away [from the eye] appear less clearly, since the visual rays, as they stream outward, take on some of the blackness of the air through which they pass. Thus, distant objects appear nebulous, as if seen through a veil.

[20] And since [each] visual ray terminates at its own unique point, what is seen by the central ray—i.e., the one that lies upon the axis [of the visual cone]—should be seen more clearly than what is viewed to the sides [of the visual axis] by lateral rays. The reason is that those rays lie nearer to [the edge of the visual cone where there is an increasing] absence [of rays], whereas those rays that approach the [visual axis] lie farther from [such an area of] absence. The same holds for objects that lie toward the middle of spherical sections whose centerpoint is the apex of the visual cone, because the generating point of the sphere itself and powers that approach their generating sources are more effective. The farther such powers extend from their sources, then, the weaker they become—as, e.g., [the power of] projection [in relation to] the thrower, or of heat in relation to the heater, or of illumination in relation to the light-source. Therefore, since the visual ray within the cone has two primary referents, one being the centerpoint of the [ocular] sphere where the vertex of the visual cone lies and the other being the straight line that originates at this point and extends the whole length [of the cone] to form its axis, it necessarily follows that the visual perception of what lies far from the vertex of the cone is carried out by a more weakly-acting ray than the visual perception of something lying at a moderate distance. The same holds for objects that lie far from the visual axis
in comparison to those that lie near it.

[21] We have therefore accounted for the properties that are intrinsically visible, those that are primarily visible, and those that are secondarily visible. We have also discussed how objects are seen more or less clearly. Darkness, however, is never seen; instead, we apprehend it through a deprivation that occurs in the visual faculty, just as we in no way hear silence but apprehend it by means of the absence of sound. And, generally speaking, we recognize various types of [sensible] deprivation through an absence of sensation. We apprehend the extent of those [missing sensibles], though, by means of the [contrasting] limits of the actual [sensible qualities bracketing them], just as we apprehend the size of a dark spot by means of the boundaries of light surrounding it, and just as we apprehend the duration of silence by means of the end and beginning of sound.

[22] We said before that there are two basic modes [determining] how visible properties are seen, and now that we have sufficiently explained the first of these modes, which depends upon the disposition of the visible properties themselves, we must explain the second mode, which depends upon how the sense of sight itself functions. And we should say, first, that all the intrinsically or primarily visible properties are in fact seen by means of a passion that arises in the visual flux, whereas the secondarily visible properties are only seen through accidents conveyed by that passion. In every visual perception, neither that passion nor whatever accident it conveys owes its existence to the Governing Faculty [alone], nor in fact does it stem from the visible objects [alone]. Rather, it depends upon a relationship or rational interchange that is established between those objects and what pertains to and arises from the Governing Faculty. Nor do we apprehend any of these passions and accidents according to such relationships alone, but also according to a moderate sensibility that the [perceptual] source has of objects under its scrutiny. That point will be demonstrated when we have analyzed it according to particular visible properties.

[23] According to what we have presupposed, we see any luminosity or color by means of a passion arising in the visual flux, while we see the secondarily visible properties that remain through the accidents that this passion conveys. Indeed, the passion arising in the visual flux is [called] “illumination” or “coloring.” Illumination by itself, however, is a sort of excess-condition in luminous objects, so it hurts and offends the [visual] sense. Illumination is also created along with coloring in objects that are struck by light from outside. Light and color are also transformed into one another by a transition of one into the species of the other, since luminosity provides the genus for both. And if light falls upon it, color becomes luminous, while light, if it is colored, is obviously altered. The visual flux, on the other hand, provides nothing qualitative to either of them, for it is necessary that the sense [of sight, which is]
perspicuous, should have no qualification but should be pure and should suffer
the qualification [passed to it] by light and color, because it shares their genus.
Nevertheless, it does undergo a straightforward qualitative alteration from all
colors and light. And this [alteration] is not always sensible, except when the
intensity of the passion [it arouses in the visual faculty] is adequate to the
perceptual capacity of the Governing Faculty. So too, differences among colors
struck by the visual flux are not apprehended if the colors of the objects in
question are merely separate while the rest of their [qualifications] are identical;
but [they will be apprehended] if there is a sensible degree of difference among
those colors. Moreover, objects whose colors differ by some [small] amount
appear distinct from one another [when viewed] from nearby, but not from afar,
because vision becomes weak in distinguishing among remote objects.

[24] From what we have said, then, it is clear that the visual flux appre-
hends color by the accident of coloring. For instance, it apprehends whiteness
because it is whitened, whereas it recognizes blackness because it is blackened,
and the same holds for each of the intermediate colors. Some have thought
that whiteness is perceived through a spreading out of the visual rays, while
blackness is perceived through a constriction of them, but that is not so. In the
first place, we find no necessary reason why the rays should be spread out by
the one and constricted by the other, nor [do we know] how the intermediate
colors—i.e., red, rose, and blood-red—that are composed of these [extremes]
might be apprehended. Besides, the larger of identically white objects ought
to appear larger yet, because it is apprehended by many [more] rays, not only
on account of a [greater] aggregation of [incident] rays, but also because the
angle [at the vertex] of the visual cone does not remain constant but dilates if
the objects are white, given that such objects spread the rays out. And if these
objects are black, that [visual angle] is reduced on account of the constriction
[of the rays]. The visible portion of the sky ought therefore to appear larger in
daytime than at night, but nothing of this sort seems to happen.

[25] Accordingly, the visual flux apprehends these colors by means of the
quality of the passion that is aroused in it, and it apprehends bodies by means
of the passion aroused in it of their simply being colored and impeding the
passage [of visual rays]. Moreover, it recognizes them in a general way insofar
as surface is an intrinsic property of bodies. Thus, bodies that do not arouse a
passion of this kind in the visual faculty (such as those that we said are not
compact but subtle) are in no way seen. For they are not perceived by the
visual faculty, much less perceived as bodies by it. Yet whatever has a more
concentrated color appears to be denser in substance, even though it may not
actually be, as [in the case of] milk in comparison to glass.

[26] The visual faculty also discerns the place of bodies and apprehends
it by reference to the location of its own source-points [i.e., the vertices of the
visual cones], which we have already discussed, as well as by the arrangements
of the visual rays falling from the eye upon those bodies. That is, longitudinal distance [is determined] by how far the rays extend outward from the vertex of the cone, whereas breadth and height [are determined] by the symmetrical displacement of the rays away from the visual axis. That is how differences in location are determined, for whatever is seen with a longer ray appears farther away, as long as the increase in [the ray’s] length is sensible. (We should bear this point in mind in the case of all types of illusion, for a certain kind of illusion is created on the basis of sensible differences other than those [objective ones] arising from the actual visible properties—but we will deal with these later.) Objects that appear higher are seen with rays that incline more toward the tops of our heads, whereas objects that appear lower are seen with rays that are lower and more inclined toward the feet. So too, whatever is seen with right-hand rays appears to the right, and whatever is seen with left-hand rays appears to the left. Therefore, objects are more clearly perceived through the ordered disposition of rays, because [those rays] are not directed everywhere [indiscriminately] toward all the particular points that are seen from the vertex [of the visual cone]. Otherwise, in fact, it would happen that up and down or right and left would appear the same, and the position of a thing would never appear definite, or else every position would appear undifferentiated.

[27] Furthermore, every body appears at a single location to the [left or right] sides of anyone who looks with one eye alone, and it is seen at a single location by anyone looking with both eyes if he apprehends it with rays that are correspondingly arranged—that is, rays that have an identical and equal position within both visual cones in relation to their own axis. This happens when the axes of both cones converge at a given [spot on a] visible object, as is the case when we fix on visible objects with the normal glance that, by nature, lets us scrutinize them carefully.

[28] It seems, moreover, that nature has doubled our eyes so that we may see more clearly and so that our vision may be regular and definite. We are naturally disposed to turn our raised eyes unconsciously in various directions with a remarkable and accurate motion, until both axes converge on the middle of a visible object, and both cones form a single base upon the visible object they touch; and [that base] is composed of all the correspondingly arranged rays [within the separate visual cones].

[29] But if we somehow force our sight from its accustomed focus and shift it to an object other than the one we wanted to see, and if the [new] object toward which our sight is directed is somewhat narrower than the distance between our eyes, and if the visual rays [that] fall together from our eyes on that [new] object are not correspondingly arranged, then that same object will be seen at two places. But when we close or cover either of our eyes, then the image in one of the two locations will immediately disappear, while the other
will persist, [and the image that persists is] sometimes the one directly in front of the covered eye and sometimes the one directly in front of the other eye . . .

• **Hint to Problem 4** Ptolemy is heavily drawing from Aristotle’s account of vision, which he seems to accept. He uses a dynamic model to explain reflection and refraction, and the intensity-changes of illumination. This is reminiscent of Hero’s dynamic (and implicitly corpuscular) model he used to account for the rectilinear propagation of light. He is critical of Plato’s theory from the *Timaeus* [14, 24] and Democritus’ ideas from DS [14]. Although a conception of “visual rays” characterised Euclid’s theory as well, their treatment of visual acuity differs greatly. For Euclid at greater distances the discrete rays are dispersed (Propositions II. and III.), for Ptolemy (again following Aristotle) a weakening of the visual ray takes place [20].

• **Hint to Problem 5** Visible properties exist in two ways: some depend on the “objective” visibility conditions others on the “subjective” visual faculty [3]. In Ptolemy’s taxonomy what is intrinsically visible is the density or luminous compactness [4]. However, what is seen is colour (and light), and so colours are primarily visible. Colour, together with boundary [8] will be the “proper” (special) sensibles concerning the faculty of sight [13]. All other sensibles, like size, place, shape, motion, rest, etc. will be secondarily visible. These secondary qualities are thus relegated to be intentional entities.

An important feature of Ptolemy’s list of visible properties is that it includes a number of spatial characteristics that are the Aristotelian “common sensibles”: size (magnitude), shape (figure), activity (movement), and rest (Smith 1999). This taxonomy correlates well with the Aristotelian distinction of special and common sensibles (but Ptolemy adds “place” to the list), but for Ptolemy the separation of objective and subjective aspects of sight are stressed. This concept is also in marked contrast with atomistic conceptions of primary and secondary qualities, where extension is usually considered as primary quality (size, shape), while colour as secondary.
6.4 TEXT 4: PTOLEMY’S Optics, BOOK II, ILLUSIONS

- From the text below from (Smith 1996) many of the original footnotes have been omitted or abridged without special note. Some new ones have been included with (G.Z.) at the end to show their origin.

[83] We have thus demonstrated how each of the visible properties is apprehended through passions arising in the visual faculty, as well as through the accidents that are conveyed by those passions, and we are satisfied with what has been said on that score. However, it follows that we should differentiate among the illusions that pertain to these [visible properties] so that, on this basis, we can resolve the issues that arise from the scientific investigation of optics.

[84] Of the illusions that involve the properties themselves, some involve causes that are common to all the senses, and some involve things that are specific to the visual faculty in the actual visible properties; some, moreover, arise in the sight, and some in the mind. It is therefore worthwhile to differentiate among them and to explain which illusions arise in the sense itself and which ones arise in the judgment of the mind.

[85] Now in the illusions that are common to all the senses, misperception should not be ascribed to the eye, for in these cases one of the variables that naturally bear on every such phenomenon shows up. Moreover, we find that such illusions stem from a variety of causes [such as]: variation in the intensity of the visual power or differences among sensible objects themselves, when they are [mentally] compared to one another, or when one is [physically] placed beside the other. This happens in the science of optics with regard to the comparison of one [visual] power to another, as, for example, [when we address] the issue of why some people do and some do not see the same objects at the same distance, and the same holds for objects that are nearer or farther away. In all these cases, it is a matter of greater or less [intensity of power].

[86] It is because of an abundance of visual power that objects are seen at a distance. Hence, older people always look at an object up close, because, along with the rest of their faculties, the visual power is produced more weakly in them\(^{20}\). On the other hand, those who have deep-set eyes see farther than those who do not have such eyes; and the reason for this is that their visual power is compressed, for when it emanates from narrow places, the visual flux is stretched and elongated.

[87] The cause of farsightedness, however, is the eye’s [humoral] moisture, part of which is taken away along with the visual flux. Thus, when there is little moisture, the visual flux leaves the accompanying moisture behind at once as

\(^{20}\) This reasoning seems to contravene the evidence: it is generally the case that the older one gets, the more farsighted, not nearsighted one gets.
it radiates outward and [so] close things are seen clearly. But when there is a
great deal of moisture, an object will be seen at a great distance, so that one
who wishes to see with certainty must look from far off.

[88] What happens in vision, moreover, is like what happens in touch. For
a given object feels softer to one whose body is harder than to one whose body
is softer. Furthermore, it feels cooler to one whose body is hotter than to one
whose body is cooler.

[89] In addition, since the whole of a given magnitude appears larger, while
its parts appear smaller, when the sense grasps the magnitude as a whole, it
apprehends the quantity of the whole better [than it does that of its parts], but
when it grasps one part at a time, it apprehends the whole less [well] than the
part. And when each of the objects that are apprehended is smaller than the
whole that is composed of them, then the magnitude, if it is continuous, will
have a certain disposition with respect to its particular elemental parts.

[90] So too, nearly the same as what we have described arises from the
comparative difference of objects according to “greater” or “less”—e.g., what
happens when objects are seen next to something larger and are judged to be
smaller [than they really are], or when objects are seen along with more brightly
colored ones and are judged to be duller [than they really are]. This of course
stems from a defect in judgment when one [sense-impression] is overshadowed
by another according to the perception proper to objects that are overshadowed
by others and the [perceptual comparison] is not made proportionately.

[91] In fact, this affects all the senses, for an impression sometimes di-
minishes and sometimes augments not just in vision, but also in smell, and in
taste as well as in sound. For at the same time some things are sensed more
[intensely] than others of the same kind. Not only should accidents of this kind
be segregated from those about which vision errs, but also accidents that depend
upon a change in the visible properties that becomes so extreme as to cause an
object to appear other than it would if the visual faculty and mind were not then
somehow disturbed and their state were not changed.

[92] Now when particular colors are not seen according to their proper
nature but are mixed together, no one would say that this [mixed appearance]
is due to an illusion occurring in the visual faculty either on account of their
multiplicity or on account of the place they occupy, because the images of
colors do not appear other than they were on account of the way light falls
on them. Instead, their appearance varies in terms of more or less for the same
reason that what is not illuminated does not appear.

[93] The moon, however, has its own color, which appears during eclipse,
when light is absent, but does not appear at other times. This problem is solved
insofar as, during the time of eclipse, the moon lies in a sort of shadow. The
earth, by means of which the blocking [of sunlight] occurs, is quite far from the
moon at that time. At other times, however, the moon lies in darkness, since the
portion that creates the blockage—namely, half the lunar sphere—is contiguous with the portion that is thereby blocked. But more light falls on what lies in shadow than on what lies in darkness.

[94] Likewise, we ought not to assume that our sight is deceived when the image of colors changes in some way, as happens in the case of light that is colored by certain flowers or other coloring agents and then strikes visible objects. For everything that appears to be the same color by dint of the color of something else shining on it should not be deemed to have changed except through an effect created in the visible objects themselves from a mixture stemming from both [agent and patient] that produces a common state in the eye, in the visible object itself, and in the sensation arising from them.

[95] Under the same classification are grouped those objects whose color appears homogeneous not because of a mixture of another color with its own but because of the various colors that it possesses. They appear this way on account of distance or quickness of motion, because in either of those cases the visual power is weakened in its capacity for seeing and discerning individual constituents. For, if the distance of the visible objects is such that the [visual] angle subtended by the entire object is sufficiently large while the constituent angles subtended by different [constituent] colors are imperceptible, then, on the basis of an apprehension of the undiscerned constituent parts, when the overall sense-impression is assembled, the color of the object will appear uniform rather than composed of individual constituents.

[96] The same happens in the case of extremely quick motion, for instance, the motion of a potter’s wheel daubed with several colors. For a given visual ray does not stay fixed on one particular color, since that color leaves it behind on account of the speed of rotation. And so, in falling on all the [constituent] colors, the same ray cannot distinguish between first and last nor among those that [now] occupy different locations. In fact, all the colors appear simultaneously throughout the disc as a single, uniform color that is the same as the color that would actually be formed from a mixture of the constituent colors. By the same token, if spots of a color different from that of the disc are marked on it (provided they are not on its very axis), they will appear to form circles of the same color [as the given spot] when the disc is rapidly spun. On the other hand, if [differently colored] lines are drawn on the disc’s surface through the axis, then the entire surface of the disc will appear to be of a uniform color when it is spun. For, since the color spins through a perceptible distance in equal perceptible time-intervals, it is adjudged to touch all the locations through which it passes. The visual impression that is created in the first revolution is invariably followed by repeated instances that subsequently produce an identical impression. This also happens in the case of shooting stars, whose light seems distended on account of their speed of motion, all according to the amount of perceptible distance it passes along with the sensible impression that arises in
the visual faculty.

[97] In addition, the illusions that involve size and shapes are created in much the same way as we have already said, [for instance,] when the same object appears smaller and seems to be more rounded [than it actually is] at a great distance, because the subtended angles diminish and become tiny. Also, when an object similar to a disc revolves swiftly about one of its diameters, it appears egg-shaped to anyone viewing it from the side rather than along the axis of its rotation. The reason is that during each revolution its surface is sometimes perpendicular to the visual ray falling on its center, at which time it appears circular; but sometimes that same surface will be aligned [lengthwise] along the line whose extension intersects the vertex of the visual cone, in which case it appears as a straight line. More often than not, however, it is obliquely disposed toward the viewpoint and thus appears distended. Since the longer shape of the object predominates over the others in continuous motion, then, the shape of the entire object is noticeably distended throughout because of the speed of rotation. And so it appears to the observer to have the shape of an egg.

[98] An illusion can also arise regarding the motions of visible objects without any deception of the visual sense. But this is due to the Differentiating Faculty, when something like what we said happens in the case of the rapid motion of potter’s wheels occurs by means of the aforementioned [visual] effects. When [such wheels] revolve, it is assumed that their motion goes unperceived because of the brevity of time within which they rotate. This happens as well in the case of moving objects that are seen from afar, for they are adjudged to be stationary on account of the immobility of the visual ray, since the rays do not in this case move perceptible distances in a [sufficiently] short time.

[99] If, indeed, we move some given distance that does not bear a sensible proportion to the distance of a visible object, then the visible object will seem to travel along with us in the same direction and at the same speed. This often happens with the moon and all the other celestial bodies on account of their brightness and considerable distance [from us]. And so, when we move and direct our eyes toward such a visible object, our sight does give us the impression of a perceptible displacement, because the distance we move is as a point to the distance of the object away from us. On the other hand, objects that a given eye apprehends as moving move with it. It necessarily follows that, as long as they travel through a perceptible distance that is detectable to the sense, and as long as the distance [between eye and object and the distance of the object’s travel] are not [sensibly] proportional, those objects will be adjudged to move with the same speed and in the same direction as we do.

[100] Likewise, too, of objects that have the same speed, those that are closer appear to move faster. Among those objects whose cross-sections are equal, the ones that are nearer to the viewpoint subtend a greater visual angle.
Those things, moreover, that mark out greater arcs in equal times appear to travel more swiftly.

[101] ...even though they are smaller, [certain objects], such as small boats and small arrows, that are moving at the same rate as larger ones in fact appear to move more rapidly. Indeed, putting it another way, objects that traverse [incremental] spaces longer than themselves in several stages during equal times are adjudged to move faster than those that traverse [incremental] spaces equal to their own length, since a small object measures a given distance several times more than a large one. It therefore follows that, when objects of unequal size travel at the same speed over given distances, the smaller ones appear to move in more stages. Accordingly, then, since the object passes over the distance apprehended by the visual flux in equal times in more stages than [does the other object passing] over the same distance, it appears to move faster.

[102] As we have said, these difficulties and their ilk must be imputed to the passions and natural accidents that befall the visual sense, for in those features through which the reality [of visible objects] is perceived, there arises a misperception based on variations that are due to the visible objects themselves. There are, however, cases where this is not the cause. In such cases, where the properties of the visible objects are unaltered and there arises a certain perception, the sense does not grasp the visible properties according to its custom and nature. There is, instead, another passion or accident in them that creates an illusion, in which case the cause lies in the visual faculty itself. We must present this case in the present chapter and discuss it, distinguishing first among the things responsible for the illusion, some of which affect the visual faculty from the visible objects themselves, others of which arise from elsewhere.

[103] Now a visual illusion springing from the visible object arises in the case of colors only by the action of something besides the visible object, whereas for all the remaining visible properties an illusion stems from the various defining characteristics that pertain to those objects themselves. For these visible properties are perceived medially, while colors are perceived immediately and by themselves\(^{21}\). As far as characteristics that seem to pertain to the visible object are concerned, the cause of the visual illusion itself is that the visual sense does not grasp the features that provide the intrinsic, defining characteristics for [visual] perception. Rather, it grasps features of something else that has a mediate status [with respect to those primary features]. But the reason for illusion in the case of those properties that create visual passions apart from the visible object is that visible objects cannot be sensed without an accompanying sensation of neighboring bodies. This does not happen because

\(^{21}\text{In other words, altering the secondarily visible properties of visible objects will not alter their color, whereas such alteration can change other secondary properties. Thus, for instance, change in shape will not affect colour, but it can affect size.}\)
their action is such in all rare and subtle bodies but because what does the first sensing must to some extent not be sensible, just as the prime mover must by no means itself be mobile; so we ought not to impute a passion to everything and [to suppose] that nothing remains according to its own nature. For things that are assumed to be of such character are found to be of a different kind, and thus the first effect will be destroyed, since the end in these cases would lead to infinity.

[104] Since, therefore, we have already discussed visual illusion and categorized it according to a) the kind arising in the passion that we have said is proper to the visual faculty itself and b) the kind that arises during interpretation, let us explain each of them separately according to the particular visible properties that are subject to them. But let us speak first of those things that arise in the passion; and let us assert that, among the passions that arise in the visual flux, one is called “coloring”, another “breaking”, and another “diplopia”.

[105] It is natural and customary for visual flux to fall in a pure and direct fashion upon visible objects while maintaining a corresponding arrangement of the two visual cones. On the one hand, coloring involves the privation of purity, while, on the other, breaking involves the privation of a direct line of sight to an object. Diplopia, finally, involves a lack of corresponding arrangement [of the two visual cones]. Each of these passions is of two kinds. Coloring, in fact, can be either anterior or posterior. Anterior [coloring] is the kind that occurs in front of the visible object, whereas posterior [coloring] is the kind that occurs behind the visible object. Breaking, for its part, involves either refraction or reflection. Refraction is the sort of breaking that happens [during the visual ray’s passage] through a resisting medium, whereas reflection entails a [complete] breaking [and rebound of the ray] from [the surface of] the resistant medium. Diplopia, finally, is either anterior or posterior. Anterior diplopia involves focusing of the [visual] axes in front of the visible object, whereas posterior diplopia involves focusing of the [visual] axes behind the visible object.

[106] An illusion arises in the case of colors by the visual flux’s being colored by something other [than the visible object itself]. Anterior coloring is generated by itself and also by the two types of breaking. Posterior coloring, on the other hand, is generated only by the two types of breaking.

[107] Anterior coloring is generated by itself when we have looked for a long time at some very bright color and then look away to something else. For in that case what is last looked at seems to possess something of the color of what was first observed because the impression of bright colors lasts a long time in the visual faculty. And so, after we have looked at such [bright colors],

[22] Perhaps we are dealing here with another transposition of text.
we see neither clearly nor without some impairment\textsuperscript{23}. Anterior coloring also happens when we look at something through thin, threadbare cloths of a red or purple hue. The visual flux passes through the weft and warp of the cloths without breaking and, in the process, takes on something of the color of the threads it brushes by. Thus the visible object appears to be tinged with the color of media traversed by the visual flux.

[108] In addition, anterior coloring arises in reflection when the mirror is immobile, dense, and colored. For the colors of objects that are seen by reflection from such mirrors appear to be commingled with the color of the mirrors\textsuperscript{24}. Anterior coloring also arises in refraction when the transparent media are not particularly rare but, rather, cause significant refraction yet are neither very weakly nor very intensely colored. For if they are intensely colored, they do not allow the colors of the objects seen by means of refraction to mingle with their own after the visual flux has passed through; in effect, they predominate and color the visual flux by the force of their own coloring. And if the transparent media are only tenuously colored, the visual flux does not undergo coloring [from them]. If, however, the mingling is proportionate, then some sort of coloring occurs, and the visual flux takes on incidental color, as happens, for instance, [when it passes] through a thin cloud that is neither white nor excessively tinged, or through sheets of unblemished horn, or through the subtle vapors above the earth’s surface, or through glass fragments, or through other lightly colored objects\textsuperscript{25}. In such cases, the colors of visible objects will appear to be commingled with the color of the transparent medium through which the visual flux passes, [the resulting color] being a composite of both colors.

[109] Posterior coloring, as well, arises in the case of refraction, for instance, when transparent media take on something of the colors of the objects that lie within them or that appear behind them. The same holds for reflection as it occurs from mirrors. For, if mirrors are so disposed that the images seen in

---

\textsuperscript{23}Ptolemy is of course referring to after-images here; understandably enough he fails to realize that, depending on certain conditions, such images can be either positive (i.e. the same color of the original stimulus) or negative (i.e. complementary color of the original stimulus).

\textsuperscript{24}The most common material for mirrors in Ptolemy’s time was bronze, although other materials, such as tin, lead, gold, silver, polished stone (e.g. obsidian), and even dark glass were used. Glass mirrors with metal backing were probably only used after the third century A.D. Modern glass mirrors backed with mercury-tin amalgam were developed during the sixteenth century.

\textsuperscript{25}Thin plates of horn, as well as of marble and other less durable materials (e.g. parchment), were often used in the place of glass in Antiquity. Depending on where it was manufactured and whether (and what) coloring agents were used, glass of that era tended to take on a variety of shadings from light yellow to dark blue.
them lie on the surfaces of the mirrors themselves, then they are deemed to have
the color of the object that the visual faculty apprehends after reflection. This
sort of thing also happens in another sort of disposition, when the reflection
takes place at extremely acute angles to the surface of the mirrors. For the place
where the object[’s image] appears is determined by where the ray emanating
from the eye to the mirror’s surface meets the orthogonal dropped from the
visible object to that surface [i.e., the cathetus of reflection], given that each of
these two lines lies in the same plane, which is normal to the mirror’s surface.
Therefore, when the angles at which the reflection takes place from the surface
are small, the cathetus of reflection will be short, and so the location of the
image will nearly coincide with the surface itself.

[110] These facts, as well their logical implications, will be demonstrated
in due time in a later discussion that will include appropriate details, so that,
in attempting throughout to demonstrate specific points about individual in-
stances, we do not create confusion in our account. Let this suffice for now,
therefore, and let us limit ourselves to discussing only what is pertinent here:
i.e., to account for the [type of] illusion that is due to the visual faculty.26

[111] According to the preceding account, if a fire or light lies slightly
above the horizon and there is a pond or pool of water near the observers,
and if that water is somewhat roiled, then, to the visual flux that is reflected
from the water’s surface to the fire or light, the image of the light in that
water’s surface will appear distended; and [that image] will seem to move in
the same direction as the observers so as to lie in a direct line between them
and the visible object. For that reason, there appears to be light where in fact
no light exists, but this appearance is due to reflection. For the light of the
luminous object is considerable and is diffused in various directions, but the
portion of it that is seen in the water represents a mere fraction that is projected
only longitudinally. Moreover, the luminous body itself does not shift with the
movement of those viewing it but lies in the same place for all who look at it
from different directions. The image in the water does shift with the movement
of those viewing it, though, and when there are several observers, each will see
it in a different location: i.e., where it appears [to lie] in a direct line with the
generating object.

[112] These, then, are the characteristics of objects that appear by them-
selves and of those that are seen by means of images cast in other objects.
As long as it is moderate, any unevenness in the water’s surface causes light to
appear elongated and dispersed. When the surface is smooth and absolutely flat,
then the image seen in it is identical in shape to the actual objects [generating
it], because in that case reflection can take place at only one spot on the plane
surface, the angles being equal when there is reflection from one place to an-

---
26See II. 84.
other. And when the [reflecting] surface is uneven and unpolished, it is possible for a visible object to appear at several spots on the composite surface defined by the [waves and troughs of the] surfaces under discussion. That is how the visible object's image becomes distended, since the image will lie at those spots from which reflection at equal angles occurs, and those spots vary according to the juxtaposition of convex and concave surface-segments. Moreover, the location of the [composite] image that appears at common spots will coincide with the [reflecting] surface, since the visible object will lie outside the center [of concave curvature]. On the other hand, each of the images that appear below the surface will appear only slightly removed from it on account of the obliquity of the visual ray\textsuperscript{27}. But when the water is violently roiled, the [composite] image is broken up and dispersed according to the crests and troughs, whose unevenness is considerable. This is why each particular image produced by such surfaces can be seen clearly, some of them appearing higher than others and some appearing lower. When, however, the surface is only slightly agitated, the image appears more continuous because of the smallness of the reflecting surface-segments that are contiguous to one another. And when the differences in height between such contiguous surface-segments is undetectable, the entire [composite] image will appear as a single, elongated object and will seem to lie at the same level as far as the sense is concerned.

[113] From what we have said, then, it is obvious that if some portion of the sun overlooks a relatively calm sea at dawn or dusk, there will appear on the water an image that is not only continuous with, but also of the same size as the portion of the sun appearing above the horizon. If, however, the water is somewhat roiled, that image will appear longer [than the portion of the sun generating it].

[114] So any illusion involving colors that is due to a passion arising in the visual faculty is created at most according to the types discussed earlier. However, the illusion involving location that is due to breaking and diplopia is created in either of the following ways.

[115] In the case of diplopia, the same object appears in two places, or two objects appear in one place, just as we showed earlier. In the case of reflection and refraction, though, the object appears [to lie] directly in line with the [incident] visual ray, even though this is not actually the case insofar as the ray is broken. Moreover, among objects that are seen in this way, some appear to lie closer than they actually are the true measure being the distance between the objects and the viewers in combination with the distance between the objects and the surface at which the breaking occurs. Other objects, however, appear to lie at the same distance as the true one, and others yet appear to lie at a greater distance [than the true one]. In addition, certain objects are seen by those visual

\textsuperscript{27}See II. 109.
rays that lie on the same side with them, whereas others are seen by different rays, depending on the surface-shapes of the bodies that cause the breaking; this is the case, for example, when we see left-hand objects with right-hand rays or higher objects with lower rays, and vice-versa.

[116] It is for this reason that elevated islands appear lower than they actually are when seen at sea, because the rays that strike the sea below them are raised higher by the reflection while, meantime, the air that is perceived directly by the visual flux is rendered higher than the islands. This point becomes even clearer when a red cloud lies above the island, for in that case an identical red cloud appears below the island.\(^{28}\)

[117] So too, it is possible in reflection for the same object to be seen in several places, as happens in the case of concave mirrors and in mirrors that display several images. They do so because they are set up in such a way that certain of their constituent parts reflect rays to the visible object while others reflect them elsewhere. Accordingly, several disparate rays show the object according to their particular direction, and the number of locations where the visible object appears is equal to the number of [such disparate] rays.\(^{29}\)

[118] Therefore, in the case of all visible objects, an illusion arises from the passion that affects the visual flux. In the case of sizes and shapes, this is due to either kind of breaking, whereas in the case of motion it is due to radial sweep [of the visual flux].

[119] In reflection and refraction, which define the kinds of breaking, an illusion comes about when the surfaces of reflection or refraction are not plane. In fact, this is due to the convexity and concavity of the [given] surface, because, if the visual angles subtended by the images are larger or smaller than the angles subtended by the [generating] objects as seen [directly] and properly, and if the distance [between eye and image] is the same as that [between eye and] actual object, then the image is rendered larger or smaller than the object itself as seen directly.

[120] Furthermore, when individual rays distributed over the image of the object are longer or shorter than the rays that fall upon the actual object itself when it is directly seen, then, for the previously-given reason, the shape of the image appears different from that of its generating [and directly viewed] object. For that reason, too, straight lines that appear behind a transparent object whose surface is not plane do not appear straight, because, in that case, the broken visual ray does not maintain its proper [spatial] arrangement, so that it will not strike the visible object from the point directly facing it but, rather, from a point located elsewhere. Thus, even though the line is actually straight, its image does not appear so. On the other hand, if the [surface of the] transparent object is curved, then the line is not bent, but the image is bent in the same way.\(^{29}\)

---

\(^{28}\) Ptolemy seems to be referring to mirages.

\(^{29}\) Ptolemy is describing composite mirrors made up of several plane mirrors.
body is plane and the object seen [immersed] in it is straight, then, when part
of that immersed object lies outside [the medium] while the other lies within it,
as happens in the case of oars, the resulting perception is that the object itself is
bent. In fact, the part outside [the medium] is seen in its true location, whereas
the part lying within the transparent medium appears closer to the refracting
surface [than it really is]. Therefore, the two aforementioned segments of the
visible object will not appear to lie in a straight line with one another; instead,
the object will look broken.

[121] A particular kind of continuous sweep of the visual ray gives the
impression that the visible object is moving. This sort of sweep sometimes
occurs at the very source of the visual flux, as happens, for instance, in the case
of vertigo and fainting, whose effect reaches to the eye itself. For, while striking
the visible object, the ray continues to follow the motion of its source, which
has changed the direction of its focus, but the sense does not detect this sort of
[internal] motion. Consequently, the perceiver assumes that, in passing over
the object seriatim, the visual ray is moving. This illusion is created at times
because of visual contact with transparent media, as happens in the case of an
object seen in moving water. It actually seems to move with the flow of the
water, because, as the water’s surface moves, [its constituent parts continually]
take up different locations. Therefore, several visual rays view the visible object
[seriatim], and, being seen at several [successive] locations [in terms of the
moving portions of water], the object seems to move.

[122] This, then, is all we have to say concerning the types of illusion that
are due to the passions aroused in the visual faculty from the visible prop-
erties taken generally. We are accordingly obliged to turn our attention to the
phenomena that arise from the perceptual scrutiny of those properties.

[123] But let us begin with illusions that have to do with location. Now
illusions that involve colors come about solely through a passion of the visual
sense. But the illusions we just mentioned that involve location and all [the rest
of] the visible properties appear in two ways at most. On the one hand, such
an illusion arises from the colors inherent in a variety of visible objects, those
colors being misapprehended by the visual faculty when it is unable to carry
out its proper perceptual scrutiny and is led to apprehend the object by means
of its initial sense-impression. On the other hand, such an illusion arises from
the actual arrangement of the visual rays, in which case what is discerned about
the visible properties is not perceptually scrutinized as it should be in terms of
variations [in the proper arrangement of the visual rays] but in some easier way.

[124] Of objects that lie in the same region, those that are brighter seem
closer. But in these sorts of illusion, visual perception of [how far away] the
dim object [lies] depends not on the length of the visual ray when the distance is

30The reference might be to what is today called the oculogyral illusion.
considerable, but, rather, on the difference [in brightness] of colors. Likewise, the locations of bright objects, such as the sun and the moon, are judged to be nearer, whereas dim objects appear to be remoter, even if they are actually nearer. Thus, mural-painters use weak and tenuous colors to render things that they want to represent as distant⁴¹.

[125] Furthermore, when we look in some given direction from elevated places without [also] looking downward, we judge that the terrain that is far away from us lies below our level. This illusion arises from the fact that the basis upon which the judgment should be made is the length of the rays. In that case [i.e., of properly based judgment], though, the [central rays of the] visual flux fall on the actual ground, so the illusion does not arise. However, since the object [i.e., the distant terrain] has been sensed by the lower of the visual rays or by a ray that is similarly oblique, and since things that are seen by such rays are wont to appear below our feet, and since the viewpoint rarely happens to be in an elevated place, the object is judged to lie below us.

[126] The same sort of illusion arises in the case of size according to both of the ways that have been discussed. For example, when objects subtend equal [visual] angles and lie at equal distances [from the eye], the one that is less brightly colored appears to be larger⁴². But if the distances are unequal, the more remote one seems larger than it did at equal distance insofar as [judgment of] distance now enters into the account. This sort of illusion arises because, instead of judging [the sizes of] these things by means of the size of [subtended] angles, one judges in this case on the basis of a perception of the increased length of the visual ray. According to that criterion, objects that are the same size appear smaller when they move away from the viewpoint. From this it necessarily follows that, among those objects that should be [judged] equal in size according to the subtended visual angles, the ones that lie at the greater distance should appear larger. The same illusion also stems from differences in colors, for a object whose color is dimmer seems farther away and is therefore immediately assumed to be larger, just as happens with objects that actually are—i.e., when objects are seen under equal angles while some of them lie at a greater distance.

[127] A similar sort of illusion occurs in the case of shapes when the shape of an object is perceived and recognized not by means of the shape of the [base of the] visual cone as defined by the visible object with which it is in contact, but by means of one of the aforementioned characteristics. For, according to

---

⁴¹This technique of rendering so-called atmospheric perspective by the use of weaker colors was well established among painters of Ptolemy's time.

⁴²The Helmholtz irradiation illusion states the opposite: the brighter object will appear larger than the dimmer one. Perhaps Ptolemy's misapprehension stems from his interpretation of the moon illusion (III. 59, cited in the chapter on the illusion, G.Z.).
the colors applied to them, surfaces sometimes appear convex and sometimes concave. Thus, a painter who wishes to represent these two shapes by means of colors paints the part he wants to appear higher a bright color, whereas the part he wants to appear concave he paints with a weaker and darker color.

[128] This is why we judge a concave veil to be convex when we view it from afar. The reason is not that the wind disposes it in such a way that sunlight and the visual flux reach the area of concavity [blown inward by the wind]. Rather, the reason is that the [relatively] orthogonal rays strike the middle of the veil so that it shows forth vividly, whereas at its outer edges either no ray at all or a somewhat oblique one strikes it, which is why it appears dark [toward the edge]. Accordingly, then, the edges of the veil appear depressed while the middle appears elevated, and this is how something that is actually convex appears [to the viewer].

[129] In the case of transparent objects, such as a glass [plate] etched on one of its faces, the surface itself does not appear flat to us when we view it from the unmarked face. The section [of the plane face] that lies over a raised portion of the etching will appear indented, and the section [of the plane face] that lies over an indented portion of the etching on the opposite face will appear raised. For in this case the sense does not interpret the shape of the first surface it encounters by means of the shape of the base of the visual cone that strikes the surface. Instead it does so by means of the shape [of the base] formed by the flux as it passes out of the transparent body. When the visual flux passes through a raised portion of the etching, its shape [at the base of the visual cone] will be raised, whereas when it passes through an indented portion of the etching, its shape [at the base of the visual cone] will be indented. But since the sense of sight concludes that the shapes of such objects are contrary to the shape at the base of the visual cone (indeed, it detects concavity by the convexity of the base of the visual cone and convexity by the concavity of the base of the visual cone), it arrives at the same conclusion in this case too. For, since the base of the visual cone striking a raised portion of the etched side of the transparent body is raised, it concludes that the bodies themselves are indented. On the other hand, it concludes that an indented portion is raised when the base of the visual cone is indented at that point.

[130] Something like this also happens in certain types of visual impressions involving motion. We can understand this on the basis of what we will [now] show: namely, that objects that do not move swiftly but nonetheless disappear quickly from sight are judged by the viewer to travel swiftly. Examples

---

33 Well before Ptolemy’s time, painters in the Greco-Roman world had perfected techniques of chiaroscuro using both “linear” and “painterly” techniques to give the illusion of relief in their depictions.  
34 Alexandria was an important centre for glassmaking at the time.
are to be found in a fire that moves for a brief time, a spark, and [luminous] bodies that pass by windows and narrow openings. This illusion [of speed] comes about because the sense does not judge how much the visual flux has been affected [by the passing object] in terms of the time in which the motion occurs but rather in terms of what appears and is perceived by means of color. For, when any object passes over the entire visual field in a brief, perceptible period of time, it appears to move swiftly and disappears at the edge of that field. But, in the case of moving objects, if any of them happens to disintegrate or vanish for some other reason—e.g., if it disintegrates before leaving [the visual field] or vanishes before reaching either extremity of that field—then such objects are judged to move swiftly because, in that case, the visual faculty grasps the speed in terms solely of the speed with which the object is lost to sight or vanishes.

[131] Furthermore, when a boat stands still in a calm, waveless river that flows swiftly, anyone in the boat who does not look at the shoreline beyond [but focuses on the river] judges that the boat is moving swiftly upriver while the water is standing still. The reason for this illusion is that the motion of the water sensed by the visual flux, being opposite to that with which the boat is assumed to move, is manifested by the contrast between the color of the boat and the color of the water. Now the contrast created by the motion of the parts of the water’s surface alone is not clear to the sense because of the uniformity of the parts of the [water’s] surface and the similarity [throughout] of its color. Yet, according to the motion of the visual flux upon the parts of the visible object’s surface, it is necessary that either the water or the boat appear to move. Thus, since the water will appear calm, the motion must appear to belong to the boat. On the other hand, if we look at the water, the shoreline, and the boat all at the same time, and if we take cognizance of the fact that the shoreline is stationary, then we will see that the boat is stationary, since the boat is seen by the same rays that see the shoreline. We will also see the water moving since we will have realized that the boat and the shoreline are stationary.

[132] Likewise, if we sail in a boat along the shore during twilight, or if we move in something other than a boat, and if we do not sense the motion of the thing carrying us, then we judge the trees and topographical features of the shoreline to be moving. This illusion stems from the fact that, when the visual rays are displaced [laterally], we infer that the visible objects are moving because of the displacement of the visual ray. Although the visible objects are stationary, then, it is assumed that the apparent motion belongs to them.

[133] It is also assumed that the image of a face painted on panels follows the gaze of [moving] viewers to some extent even though there is no motion in the image itself, and the reason is that the true direction of the painted face’s gaze is perceived by means only of the stationary disposition of the visual cone that strikes the painted face. The visual faculty does not recognize this, but the
gaze remains fixed solely along the visual axis, because the parts themselves of
the face are seen by means of corresponding visual rays. Thus, as the observer
moves away, he supposes that the image’s gaze follows his.

[134] It must be borne in mind, though, that in all the visual perceptions
that arise inferentially from sensation and perceptual scrutiny, several visible
features in one and the same object are involved. And those features that are
not properly judged in and of themselves cause a false perception of those that
are properly judged in and of themselves. The kind of misperception formed in
this way need not always arise in the cases of illusion that we have applied to
the particular visible properties. It may arise, instead, when the apprehension
of the characteristics to be inferred is not clear but is derived through differences
among characteristics that are not supposed to be inferred and that are irrelevant
to those that are. This is the case with the variations in the visual flux by which
location, motion, size, or shape are judged, since these properties cannot be ap-
prehended in and of themselves by sense [alone], whereas those characteristics
by which the colors of the objects themselves are judged are more manifest.

[135] When visible objects of the same size and equidistant [from the
viewer] are seen from nearby, the visual sense does not judge the brighter of
these objects to be nearer or smaller, nor does it judge that some flat objects
are convex or that others are concave on account of differences in color. Nor
does it judge that smooth, flowing water is static or that the depiction of a
face follows the gaze of the viewer. The reason for this is that, on the basis
of its appropriate passion, the visual faculty is able to discern differences in
color at a greater distance. However, because they are apprehended by means
of accidents that are conveyed incidentally by that passion, differences among
the remaining visible properties are apprehended when they are close, but how
[they are actually disposed] is not [apprehended]. For in each of these visible
properties the accident that occurs at the base of the visual cone relates to
[radial] length, which represents the distance [between viewpoint and visible
object].

[136] Yet, if that relation is beyond the sense’s capacity to grasp, it induces
an imperfect perceptual apprehension. Therefore, since it cannot see the visible
object in the way properly suited to it, the sense apprehends it through the evi-
dence of other differences. Accordingly, the visible object sometimes appears to
it properly and sometimes through a false perception—one that is false, like the
one arising from the illusion in the previously discussed cases, yet true insofar
as it is of a smaller, more oblique object. For visual perception should naturally
take place at a fairly short distance, and when the object is properly sensed
because the eye is near and the visual flux strikes that object in the direction
suited to it, then in such circumstances perceptual inference is not led astray in
the visual faculty. And all the characteristics of objects that are perceived by
means of angles will be no less evident than the others.
[137] At this juncture, though, we ought to point out that everything we have said about illusion applies not only to an illusion due to the sense of sight itself but also to the perception that arises from it. And since we are deceived in several cases when the visual sense impinges on visible objects according to its nature and habit, while the mind, remaining in continuous apperceptual touch with such objects, judges them to be abnormal, we must reiterate that this is an illusion involving mental inference.

[138] This is what happens in [the perception] of position when we look into plane mirrors and the visible object [i.e., the viewer himself] faces the mirror directly. In that case our sight shows us our [right-hand and left-hand] sides in the way that is natural for it to show objects viewed directly: i.e., what is seen by right-hand rays appears to the right, while what is seen by left-hand rays appears to the left. Our mind, however, shows us right as left and left as right, because objects that actually face us are so disposed that their right is opposite to our left, while their left is opposite to our right. And this is why, when we move one of our hands [in front of a mirror] our sight tells us that the hand that moves [in the mirror] is the one facing it [i.e., right to right or left to left], while our mind tells us the opposite.

[139] An inferential error about distances and their amount also arises, as happens in daylight when we look through the air that surrounds us. Since this air is in fact thicker and more colored than that which is higher up because it lies at [the surface] where plentiful vapor arises from land and water, it is more apt to take on the light that infuses it and to absorb the visual flux. We therefore suppose that we see the sky to be of the color that is common to both the vapors and the sky. The same holds generally for all extended bodies that are rare and humid and that lie at a fairly remote distance from the air that surrounds us; the exceeding rarity of such bodies precludes the visual flux entirely from seeing them, even though the light does not impede their apprehension. This is what happens when the viewer stands in the dark and looks at the stars but does not see what surrounds them, in spite of the fact that light impinges upon it. On the other hand, when the viewer stands in a lighted place, he does not see the stars, because the light intervening between him and them weakens the outgoing visual flux. The air that is seen in daylight is judged to be more remote than everything else, insofar as nothing else appears more distant than it, but the sun and moon are judged to be nearer [than its outer reaches] because of their brightness. However, in seeing the sky as higher than anything else, the mind arrives at a false conclusion and supposes that the mere visual impression is correct and represents what it perceives; and [so] it judges that what seems very remote to it is larger than something else that is naturally and truly more remote and larger than everything else.

---

35 This explanation is confusing at best. See II. 90. and II. 19.
[140] Something like this also happens in the case of shapes: namely, that in buildings whose walls are parallel along the sides, the upper edges appear wider, as do high door-lintel. This is due to a perception created in the sense, even though the upper parts of those buildings do not actually have a narrower interval or lie nearer [one another] than the lower parts. Indeed, men have been accustomed to building in such a way that the result is well disposed and solid. Therefore, because it normally seems to the mind that they are wider, even though they are not, it judges them to be truly so, since it assumes of such edifices that they have parallel sides. Now, because these buildings are erected vertically, they stand perpendicular to the horizon, but things that are perpendicular to the horizon are parallel to one another. However, since this is not how it looks, the mind judges that one of the two opposite [horizontal] sides is longer than it actually is, as is the case with things that do not actually have parallel sides\textsuperscript{36}.

[141] A false inference can also be made in the case of movements, just as happens in the case of horse-drawn chariots, when their motion is not fast. For the sense does not grasp their individual constituents [separately] but takes in the horses and the wheels at the same time. When we look at them, though, we assume that the horses are going quickly, because in things of this sort that pass over equal distances, the motion [of each constituent part] takes place in equal times, and that holds for both the horses and the wheels. As far as the wheels are concerned, because of their small size, they make several revolutions in the course of their motion, and that rotary motion is fast and continuous. On the basis of a general extrapolation, the mind imputes an equality between [linear] motion and rotation and concludes that the passage of the horses is swift.

[142] Here, then, is the end in the present book of our promised discussion of illusions and the cases in which vision is deceived.

• **Hint to Problem 7** A contemporary account answers the question the following way (Smith 1999, pp. 45-6):

Ptolemy’s analysis of visual illusions reflects his understanding of visual perception as a three-stage process running from sensation, through perception, to apperception. The first type of illusion, which occurs at the level of brute sensation, involves “feeling” rather than judgment. This characterization is evident, for instance, in the oculogyral illusion, the result of which is a “feeling” that, even when the viewer stops spinning, his surroundings continue to rotate for a while. Likewise, in the

\textsuperscript{36}A possible mistranslation from the Greek might be the result of the obscurity of this passage (G.Z.).
case of reflection, the object appears to be where it is not (i.e.,
directly along the line of sight) because the flux fails to “feel”
the impedance of the mirror-surface that breaks it. The second
type of illusion, for its part, involves the common sensibles
and, therefore, the perceptible characteristics of things. The
example of illusionism in painting illustrates this type of mis-
perception perfectly. The very fact that we are able to discern
spatial characteristics in two-dimensional representations (e.g.,
trompe l’oeil paintings) is a clear indication that it is not our
eyes, but our brains that are deceiving us. Finally, as to the
third type of illusion, involving false mental inference, it is
evident that we are dealing not with common sensibles but
with incidental sensibles. Thus, in the case of image-reversal,
the representation that we initially “see” is not of ourselves but
of someone else facing us as if through a window. The crucial
feature of Ptolemy’s analysis of visual illusions is therefore not
to be found in detail or specification; indeed, many of the
texts cited by him were part of a common stock gathered
together long before he wrote. It is, rather, to be found in the
systematic way he fit that analysis into the broader framework
of his theory of visual perception.

6.5 Text 5: Ptolemy’s Optics, Book V, Refraction

- From the text below from (Smith 1996) many of the original foot-
notes have been omitted or abridged without special note. Some new
ones have been included with (G.Z.) at the end to show their origin.

[1] There are two ways in which the visual ray is broken. One involves
rebound and is caused by reflection from bodies that block the [visual ray’s]
passage and that are included under the heading of “mirrors”. The other way,
however, involves penetration and is caused by a deflection in media that do
not [completely] block the [visual ray’s] passage, and those media are included
under the single heading “transparent”. In the preceding books we have dis-
cussed mirrors; and, insofar as it is possible for it to be demonstrated, we have
explained not only variations in the images of visible objects according to the
principles laid out for the science of optics, but also what happens with each
of the visible properties [in reflection]. It thus remains for us at this point to
analyse what sorts of variations occur in such objects when we look at them
through transparent media.

[2] It has been claimed earlier that this sort of bending of the visual ray does
not occur [the same way] in all liquids and rare media; what happens, instead, is
that in each one of these [media] the amount of deflection is determined solely by the way in which the medium allows penetration. It has also been claimed that the visual ray radiates rectilinearly, and such rays break only because of an impedance posed by the surfaces separating media of different consistency. It has also been claimed that refraction occurs not only in the passage from rarer and more tenuous to denser media—as happens in the case of reflections—but also in the passage from a denser to a rarer medium. And it has been claimed that this breaking does not take place at equal angles; however, the angles [of incidence and refraction] do bear a certain consistent quantitative relation to one another with respect to the normals.

[3] At this point we ought to investigate the quantitative relationship between the angles [of incidence and refraction] according to specific intervals. But we should start by discussing the phenomena that such refractions have in common with reflections. First, in either case, whatever is seen appears along the continuation of the incident ray—i.e., along the continuation of the ray that emanates from the eye to the surface at which it is broken—[second, the object appears] on the straight line dropped perpendicularly from the visible object to the surface where the breaking occurs. It therefore follows that, just as was the case for mirrors, so in this case, the plane containing the broken ray-couple must be perpendicular to the surface where the breaking occurs.

[4] We have already shown in the place where we laid out the principles governing mirrors that the above points are in the nature of observable phenomena and that what happens [in the breaking of rays] is quantifiable.

[5] That this is clear and indubitable we can understand on its own terms by means of a coin that is placed in a vessel called a baptist37. For, if the eye remains fixed so that the visual ray passing over the lip of the vessel passes above the coin, and if water is then poured slowly into the vessel until the ray that passes over the edge of the vessel is refracted toward the interior to fall on the coin, then objects that were invisible before are seen along a straight line extended from the eye to a point higher than the true point [at which the coin lies]. And it will be supposed not that the ray is refracted toward those lower objects but, rather, that the objects themselves are floating and are raised up to [meet] the ray. For this reason, such objects will be seen along the continuation of the [incident] visual ray, as well as along the normal dropped [from the visible object] to the water’s surface—all according to the principles we have previously established.

37 A hollow semicylinder whose ends are closed off so that it can hold water.
[6] [EXAMPLE V.1] Now, let us suppose that point A [see figure] is the eye, ZHE the common section of the plane containing the refracted ray-couple and the surface [of the water] in the vessel, and ABD the ray passing over the vessel’s lip at B. Let us also suppose that there is a coin at G, which lies toward the bottom of the vessel. Then, as long as the vessel remains empty, the coin will not be seen, because the body of the apparatus at B blocks the visual ray that could proceed directly to the coin. Yet, when just enough water is poured into the vessel so that its surface reaches line ZHE, ray ABH is deflected along line GH, compared to which AH[D] is higher. In that case, then, the coin will appear to be located along the cathetus\(^{38}\) dropped from point G to EH—i.e., cathetus LKG, which intersects line AHD at point K. Moreover, its image-location will lie on the radial line passing from the eye and continuing rectilinearly to point K, that radial line being higher than the actual ray [HG] and nearer the water’s surface; so the image will appear at point K.

[7] The amount that the ray is refracted in water below the [original line of] sight is determined according to the following experiment, which is conducted by means of the bronze plaque that we constructed for analysing the phenomena of mirrors.

[8] [EXPERIMENT V.1] Let circle ABGD [in figure] be described on that plaque about centerpoint E, and let the two diameters AEG and BED intersect one another at right angles. Let each of the [resulting] quadrants be divided into 90 equal increments. At the centerpoint let a small marker of some color or other be attached, and let the plaque be stood upright in the small vessel [discussed in the previous experiment]. Then let a suitable amount of water that is clear enough to be seen through be poured into that vessel, and let the graduated plaque be placed erect at right angles to the surface of the water. Let all of semicircle BGD of the plaque, but nothing beyond that, lie under water, so that diameter AEG is normal to the water’s surface. From point A, let a given arc AZ be marked off on either of the two quadrants that lie above the water. Furthermore, let a small, colored marker be placed at Z.

\(^{38}\) Or perpendicular (G.Z.).
Now, if we line up both markers at Z and E along a line of sight from either eye so that they appear to coincide, and if we then move a small, thin peg along the opposite arc GD under water until the end of the peg, which lies upon that opposite arc, appears to lie directly in line with the two previous markers, and if we mark off the portion of the arc GH that lies between G and the point at which the object would appear unrefracted, the resulting arc will always turn out to be smaller than AZ. Moreover, if we join lines ZE and EH, angle AEZ > angle GEH, which cannot be the case unless there is refraction—that is, unless ray ZE is refracted toward H according to the excess of one of the opposite angles over the other.

Furthermore, if we place our line of sight along normal AE, we will find the image directly opposite along its rectilinear continuation, which will extend to G; and this [radial line] undergoes no refraction.

In the case of all the remaining positions, when arc AZ is increased, arc GH in turn will be increased, and the refraction will be greater. When arc AZ is 10 degrees out of the 90 into which quadrant [AB] is divided, then arc CH will be around 8 degrees. When AZ is 20 degrees, then GH will be 15.5. When AZ is 30 [degrees], then GH will be 22.5. When AZ is 40 [degrees], then GH will be 29. When AZ is 50 [degrees], then GH will be 35. When AZ is 60 [degrees], then GH will be 40.5. When AZ is 70 [degrees], then GH will be 45.5. And when AZ is 80 [degrees], then GH will be 50 [see the table below for a synopsis].

<table>
<thead>
<tr>
<th>incidence</th>
<th>refraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>8.0</td>
</tr>
<tr>
<td>20</td>
<td>15.5</td>
</tr>
<tr>
<td>30</td>
<td>22.5</td>
</tr>
<tr>
<td>40</td>
<td>28.0</td>
</tr>
<tr>
<td>50</td>
<td>35.0</td>
</tr>
<tr>
<td>60</td>
<td>40.5</td>
</tr>
<tr>
<td>70</td>
<td>45.5</td>
</tr>
<tr>
<td>80</td>
<td>50.0</td>
</tr>
</tbody>
</table>

We find the amounts of refraction created in water to be just as shown on the understanding that there is no sensible difference in density and rarity among [various] waters.

Now, if we look toward a rarer medium from a dense one, such as plain water, there will appear a considerable alteration in the difference between the angles as well as in the amount of angular deflection that occurs in the course of the ray’s passage from water, which is denser, to the rarer medium. However, since it is impossible for us on the basis of the previously-described experiment to gauge the refraction produced when the ray passes from a denser to a rarer liquid, we have taken it upon ourselves to analyse the relationship of angles as follows.
[14] [EXPERIMENT V.2] Let a clear glass semicylinder, represented by arc TKL [in figure] be made in accordance with the semicircular section of the round plaque, but let its diameter be smaller than the diameter of the aforementioned bronze plaque. Then let its base be attached to the plaque so that both are completely joined. Let its center be E, and let its diameter TL coincide with [the plaque’s] diameter BD, and let AE be perpendicular to the flat surface of the glass. Therefore, every line drawn from point E to arc BGD and to arc TKL will be normal [to those arcs].

[15] Accordingly, if we set up this experiment as we did before, and if we make a small mark on the midpoint of the semicylinder’s surface where it [s axis] touches point E, and if we look with either eye along line AE toward the edge of the glass and move a marker on the arc [BGD] opposite this arc [BAD] until it appears in front of it, it will be found to lie on G itself. For line AEG is normal to both TEL and TKL. And if we move our eye until it lies directly in line with this position, and if we look along line GE so that the marker that has been moved along the arc lies directly in line with GE, then that marker will be situated on line EA. For the same reason, moreover, there will be no refraction in the passing of the ray [orthogonally] from glass into air.

[16] But if we take some given arc AZ from point A and draw line ZE, coloring it black, and then if we sight along this line until the marker, which is moved behind the glass, appears to fall in line with it, and if we mark the place—e.g., point H—where we found it so that the black color coincides with EH, then we will also find in this case that angle AEZ > angle GEH. We will also find that the angular difference is greater than the angular difference in water, where the arc [measuring incidence] was the same.

[17] If, moreover, we station our eye at point H, which is opposite point E, and sight from point H along HE, [both of the points] E and Z will appear to coincide on one and the same line of sight. And since there appears to be a refraction of the ray in this situation, it is necessary that, whether the ray passes from air into glass, as represented by ZE, and is refracted along EH, or whether it passes from glass to air, as represented by HE, and is refracted along ZE, the refraction takes place toward T. And since the normals dropped from E to TKL are the same, rays [that pass along them] are not refracted, whether they pass from E to K or from K to E.

[18] In addition, if we now analyze the amount of refraction for each of [the previous angular] positions, we will find that, when the eye is placed at the same angular distances as before and when the angle measured from point E (i.e., the angle [of incidence] formed by normal AE and ray EZ) is 10 degrees of the 90
ascribed to the circle’s quadrant, then the [resulting] angle [of refraction] GEH will measure nearly 7 degrees. When the first angle is 20 degrees, the second will be 13.5. When the first is 30 [degrees], the second will be 19.5. When the first is 40 [degrees], the second will be 25. When the first is 50 [degrees], the second will be 30. When the first is 60 [degrees], the second will be 34.5. When the first is 70 [degrees], the second will be 38.5. And when the first is 80 [degrees], the second will be 42 [see the table below for a synopsis].

<table>
<thead>
<tr>
<th>incidence</th>
<th>refraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>7.0</td>
</tr>
<tr>
<td>20</td>
<td>13.5</td>
</tr>
<tr>
<td>30</td>
<td>19.5</td>
</tr>
<tr>
<td>40</td>
<td>25.0</td>
</tr>
<tr>
<td>50</td>
<td>30.0</td>
</tr>
<tr>
<td>60</td>
<td>34.5</td>
</tr>
<tr>
<td>70</td>
<td>38.5</td>
</tr>
<tr>
<td>80</td>
<td>42.0</td>
</tr>
</tbody>
</table>

[31] It is, however, possible to formulate in such a way a general claim about refraction on the basis of previously established points. We put it thus: the amount of refraction is the same whichever the direction of passage; the difference is one of kind [rather than of degree]. For in passing from the rarer to the denser medium, the ray inclines toward the normal, whereas in passing from the denser to the rarer medium, it inclines away from the normal.

- **Hint to Problem 8** The first three paragraphs list a number of common principles, and also express the assumption that, similarly to reflection, a quantitative relationship may be established between angles of incidence and refraction.

- **Hint to Problem 9** The relationship between the angles of incidence and refraction follow a simple pattern. They give similar results than the well-known sine-law of Descartes and Snell—where the ratio of the sine of the angle of incidence and the sine of the angle of refraction give a constant value for any two media (\(\sin i/\sin r = \text{constant}\)). Only the values for 80 degrees are significantly different.

But the pattern behind Ptolemy’s numbers have a source quite different. They result from adding 0.5 degrees less to every ten degree increase in the angle of incidence than the previous increment was. In the case of air to water (the third column shows the value added to previous value to get value for the angle of refraction):
Similarly for the table on air to glass the differences decrease from 7 degrees by 0.5 degree increments. This method is reminiscent of astronomical calculations already present in Babylon and their employment here is probably due to Ptolemy’s background in astronomy (Smith 1999). The so called “first differences” (i.e. the values of the increments) decrease, but the amount of this decrease is determined by the so called “second differences”, which in this case stay constantly at 0.5 degrees.

To find these laws generally applicable for media of a certain type (See Text 5 para. 12), however, is problematic.

7 FURTHER PROBLEMS

• Problem 10 Collect the “knowledge of optics” the Greeks had. What questions did they address? How do these relate to the present-day questions of optics? And of other related disciplines dealing with light, colour, and vision? Can you explain the significant differences?

• Problem 11 Compare one of the followings a) Ptolemy’s treatment of illusions with that of Lucretius; b) the use of mathematics in Euclid and Ptolemy; c) colour as discussed by Aristotle, Alexander of Aphrodisias, and Ptolemy.

• Problem 12 Summarise how Ptolemy, Aristotle (On the Soul II.7), and other Greek writers treated the role of the transparent in optics, and what they hypothesized about the speed of light. Reconstruct the arguments.

• Problem 13 The question hinted at in Problem 4 might deserve further work—try to find traces of earlier theories in Ptolemy’s text. Use the available works from the suggested readings at the end of the earlier chapters as well.
8 Suggested Readings

Fortunately in recent years the interest has increased in the optical works of ancient authors and a number of new authors can be consulted.

As a general introduction (Smith 1999) is a useful starting point. The first chapter of (Lindberg 1976) is a classic overview. For more on the mathematical knowledge of the time see (Fowler 1987, Boyer 1957).

For Euclid (Tobin 1990) and on the linear perspective (Knorr 1991) can be consulted.

For Ptolemy, apart from the translation and detailed introduction in (Smith 1996), the non-mathematical aspects are stressed in (Smith 1988) and (Smith 1998). Binocularity is treated in (Howard & Wade 1996).

Several works on illusions can be consulted, see also the bibliography to Chapter X.
Part B

Theories & Problems
Chapter VII
THE RAINBOW

1 INTRODUCTION

Rainbows are enigmatic natural phenomena. They display a great deal of regularity which has incited attempts at explaining their causes since Antiquity, but, at the same time, they show surprising variability. Usually one, but often two bows are seen. The sequence of colours are reverse—the red, outer band of the primary, lower bow faces the red, inner band of the outer, secondary bow. The colours themselves are also subject to change. Occasionally little or no blue or, at times, red is seen, while at twilight a rainbow might appear all red. Also, in case of fine mist, the bow can be all white. The raindrop size affects not only the colours, but the spacing of the arcs, the thickness of the coloured bands, and the radius of the bows: while normally the angle between the incoming rays of the sun and the rays producing the rainbow is $42^\circ$ in special cases it can shrink to about $36^\circ$.

One of the first explanations of the rainbow by Anaxagoras was included in Chapter II, Text 4, but the first extant and detailed explanation is that of Aristotle. This text has been included in Chapter IV, Text 1, but we will here return to it and discuss it in more detail.

With some exaggeration it could be stated that the Aristotelian explanation dominated for two millenia—although the works of Theodoric of Freiberg (see Figure VII.1) and Kamal al-Din al-Farisi at the beginning of the fourteenth century pointed in many senses towards the seventeenth century theory propounded by Descartes and later Newton, they became relatively little known.

After investigating Aristotle’s theory, this chapter, considers one of the medieval theories by Robert Grosseteste, Bishop of Lincoln, and finally that of Descartes. The story of the rainbow, or rather the story of the explanations of the rainbow, however, does not end with the Cartesian theory and its modification by Newton, based on the principles of differential refrangibility introduced in Chapter IX, and hinted at in proposition 10 of The New Theory on page 283.
In the course of the eighteenth century, while little new theoretical work has been done, the existence of anomalous rainbow-phenomena became generally accepted: the two rainbows are at times intercepted by “intruder arcs”, which have centers that do not lie on the line connecting the sun and the observer’s eye. While this phenomenon could be explained by showing that light reflected from water-surfaces can also contribute to the appearance of rainbows, the regular observation of supernumerary arcs could not be incorporated into the Cartesian-Newtonian theory. These arcs appear below the primary bow (at times also above the secondary bow), and consist of several bands of colours, but often contain only green and purple.

During the nineteenth century further theoretical developments took place to furnish a new theory. Thomas Young (1773-1829), and later G. B. Airy (1801-1892), W. H. Miller (1801-1880), and Richard Potter (1799-1886) worked on theories that were in agreement with the wave-theory of light and could account for phenomena like the supernumerary arcs. Already in the nineteenth century Félix Bilet (1808-1882), experimenting with vertical cylindrical threads of water (as a model of a water-droplet—an elongated
cross-section) and monochromatic light determined the positions of the rainbows of the first nineteen orders. But these experimental developments and even newer, twentieth century theoretical and experimental approaches employing higher mathematics are not included in this introductory book. The “Suggested Readings” section supplies some of the relevant literature on the topic.

2 The Aristotelian Rainbow

Aristotle’s Meteorology is the first extant comprehensive work on sublunar phenomena, including “meteors” of all kind: aerial (like wind), aqueous (like modes of precipitation), and luminous (halos, sundogs, northern lights, rainbow, igneous meteors).

The conditions for the appearance of the rainbow was thus first not a problem in optics, but one in meteorology. It has been, however, soon understood that the light of the sun and also clouds play important roles in producing the effect: hence KRS 501 cited on page 35, where Anaxagoras states that the rainbow is the result of the reflection of the sun’s rays from a cloud.

Our natural reaction today—to call such treatment naive and fallacious, as clearly the light is refracted first and foremost, and only additionally reflected—should be withheld for a moment. From Chapter VI we know that reflection was a well known and understood phenomenon, unlike refraction. The first could be characterised by the equality of the angles of the incoming and reflected ray. The latter, however, in the time of Aristotle was poorly understood, and no—similarly simple and appealing—law could be established. Also the terminology made no clear distinction between the two—the term anaklasis used for example in Aristotle’s account has usages that would today be translated as reflection or refraction, but the Latin terms refractio and reflectio have also been used interchangeably for several centuries, even by William Gilbert (1544-1603) in the sixteenth century. This conflation of terms often plagues our efforts to clearly understand (or even to translate) the Greek scientific terminology.

To further complicate the issue, Aristotle distinguishes between two types of reflection (anaklasis). One can be translated by the term “reflection”. But in the case of the other only colour is reflected, and the equality of angles does not necessarily hold: “a mirror of this kind renders the
colour of an object only, but not its shape” (from Chapter IV, Text 1, Part 4). This irregular type of reflection (or refraction?) is therefore the one important for the Aristotelian explanation. To illustrate this latter we could give instances of light scattering, where light is reflected from particles, and one can see colours (like the blue of the sky) but no image.

In the Aristotelian theory the physical agents needed for the appearance of the rainbow are a light source (sun), a dark rain-cloud and the eye of the observer. The dark cloud serves as a surface where irregular reflection can take place.

- **Problem 1** Is there a rainbow that is not seen by any observer? What is the shape and size of the bow? Explain Aristotle’s theory of the origin of the colours of the bow (you can also use Alexander’s account from his commentary of the *Meteorology* from page 64). Note that no hint is given to this problem.

Aristotle also provides a geometrical explanation of the rainbow. The drawings to this explanation are lost, and only the diagrams of medieval commentators are extant. Today the diagram generally reprinted in modern treatises is the following (Boyer 1959).

- **Problem 2** Label the reconstruction of the lost diagram of the last paragraph in Aristotle’s *Meteorology* Book III Part 4. How does the position of the sun determine the size of the rainbow? Does the diameter of the visible rainbow correlate to the size seen?

- **Hint to Problem 2** Aristotle makes use of the concept of what may be called a “meteorological sphere”. It is a sphere containing on its surface the sun and the part of the cloud causing the refraction. In Aristotle’s model the observer is equidistant from the cloud and the sun, and is at the centre of the sphere. Lengthening the line between the sun and the observer goes through the centre of the rainbow (see also figure VI.2 on page 118).
In the figure only a semicircle is drawn: it can be considered as the meteorological hemisphere.

In case when the sun is at the horizon (as is on the picture, on the left side) a full semicircular rainbow is seen. On the diameter of the semicircle \( S \) (sun) is at the left endpoint, \( O \) (observer) at the centre, and the \( R \) (where the rainbow appears) is on the semicircle, with \( A \) as the pole of the rainbow. If rotated around \( O \) clockwise, \( A \) will disappear under and \( R \) approaches the horizon. If the sun is above the horizon, only a smaller section of the rainbow is seen—much of it is now under the level of the horizon. If—like at a summer midday—the sun is at a higher angle than the maximum angle at which the rainbow is seen with the sun at the horizon, the bow disappears.

It would be trivial to conclude that the angular radius is not affected by the simple rotation of the figure, and that solar elevation only affects what portion of the full circle will be seen, and not the size of this circle. But probably due to hampered optical judgement Aristotle falsely maintained that the smaller proportion of the full circle seen (the smaller the arc of the rainbow) the larger its diameter.

By recognising that the sun, the eye of the observer and the centre of the rainbow are in one line (collinear), Aristotle could explain the behaviour of the bow in relation to the movement of the sun, but not why the bow appears where it does: the size of the bow. As the equality of angles did not seem to hold, he maintained that the rainbow is formed by the second type of reflection. The myriad of tiny mirrors reflect the visual ray—but only from a specific region of the sky.

**Problem 3** How can you construct the locus of such points \( R \) for which the ratio of the distances \( RS \) and \( RO \) is a constant \( K \) (and not equals 1).

**Hint to Problem 3** The solution is to construct the so called Apollonian semicircle. As it is described by (Boyer 1959): “The locus of points \( R \) such that \( RS:RO=K \) is a circle with center on the line \( SO \). Hence to find the point \( R \) on the meteorological great circle \( SA \) satisfying the Aristotelian requirement, one need only to find on the
line SA the two points B and F such that BS:BO=K and FS:FO=K and construct the Apollonian semicircle on the segment SA.”

This solution “saves the phenomena”: it gives a mathematical treatment, but has no physical plausibility.

The mathematical apparatus used in Aristotle’s explanation was highly sophisticated in the age. This explanation of the magnitude of the arc has not been rivaled before the sixteenth century. The meteorological sphere has been used well into the 16th century. (In Alhazen it plays a very special role: the source of radial reflection.)

- **Problem 4** Discuss the production of colours in the Aristotelian theory. What are the primary colours? Which other colours are formed and by what process? How are the colours of the secondary bow explained?

- **Hint to Problem 4** Colours for Aristotle are formed either as a) weakening of light as a result of reflection (or refraction), b) weakening caused by the medium through which light or the visual ray has to travel, and c) the result of the weakness of sight (as people with myopia or moist eyes can see bands of colours around light-sources).

The red colour of the rainbow is formed similarly to the red of the setting sun, the other primaries, green and blue are formed by even more weakened rays. If red is seen whitened by its juxtaposition with green, yellow is formed.

The first recorded account of the secondary rainbow is that of Aristotle’s. As no contemporary or medieval diagram exists for the production of this secondary bow, Aristotle’s theory is hard to reconstruct. The two bows are supposed to be formed in the same manner—but whether there is another set ratio of the distances RS and RO or a different explanation is valid, where the reversed order of colours is explained by maintaining that the rays travel more between the eye, the blue band, and the sun than between the eye, the red band, and the sun, remains unclear. It is quite probable that the second bow can only be explained using *ad hoc* assumptions.
Robert Grosseteste (c.1170-1253) was born probably before 1170 in Suffolk, England. Little is known about him until the mid-1220s (although it is suspected that he went to study in Paris for a few years during the suspension of studies at Oxford in 1209-14). He gave theology lectures at Oxford in 1225, and for a time was the chancellor of the university. In 1230 he changed his position to teach Franciscans in Oxford. From 1235 to 1253 he served as bishop of Lincoln, the largest diocese in England.

One main influence of his scientific works had been the new Aristotelianism spreading in Western Europe. The impact of Aristotle’s work on Physics, or On the Heavens, and not least the Posterior Analytics, a treatise on the nature of scientific knowledge can be traced in many of his writings. In the first decades of the thirteenth century he wrote works on comets, the generation of sounds, and produced a complete commentary on the Posterior Analytics. For years he lectured on Aristotle’s Physics, and in the 1230s wrote works on motion, on the influence of elements, on light, colour, and a series of other works connected to natural philosophy.

One central aspect in these works is Grosseteste’s account of the nature of light and its role in causality—both natural and divine. Here one can see the influence of Neoplatonic sources, Pseudo-Dionysius, and the ideas of Augustine.

He extended the Augustinian metaphor of illumination to issues in natural philosophy. Originally the metaphor was used to “illuminate” the means of understanding: if the light of an eternal idea is present to an observer, this idea (or truth) will be clear to him. As light is needed to see colours in the world yet our weak (created) eyes cannot apprehend the source of this light, the sun, so does the human understanding rely on the divine, but cannot directly apprehend it.

Sense-perception also acts as an arouser of reason—it is through the encounter with sensible things that the latter is awakened. Once awakened, it begins to draw distinctions and abstract until it arrives at the cognition first of universals, then of necessary truths and finally of the sorts of demonstrations that provide the strictest kind of knowledge.

Grosseteste was also fascinated by the light’s ability to multiply and spread itself instantaneously (as was generally believed even by Descartes) in all directions. He identified light with the primary corporeal form, first
dimensionless, spreading without finite multiplication. This account also enabled him to present an account of the generation of the physical universe.

In Text 1 Grosseteste’s treatise on the rainbow is reproduced. It does not only contain his explanation for the remarkable phenomenon, but also gives us insight into the medieval grouping of the sciences and includes a curious theory of refraction.

- **Problem 5** What are the disciplinary boundaries between physics and the study of perspective? What are the subdivisions of perspective? What is the connection between Grosseteste’s explanation and the Aristotelian account? Note that no hint is provided for this problem.

- **Problem 6** Which ancient theories surface in Grosseteste’s treatment of vision?

Probably not knowing the work of Ptolemy (he could not read Greek or Hebrew), or the later study of Alhazen (showing that the law of proportionality is false), Grosseteste attempts to explain refraction. He maintains that correctly understanding it would enable us “to read minute letters from incredible distances, or count . . . any minute objects”. This shows that even in the Middle Ages applicability of scientific findings was not wholly neglected.

- **Problem 7** How is refraction discussed by Grosseteste? What is his main argument to support his half-angle law? No further hint is provided, use the footnote material in Text 1.

After discussing the aim of his work, providing optics with a law of refraction, Grosseteste sets out to discuss the formation of the rainbow.

- **Problem 8** What are the possibilities for rainbow formation? How is the correct explanation picked?

The actual explanation of the rainbow is fairly obscure (see Further Problems), but the Aristotelian meteorological hemisphere is recognizable.
in the description. Instead of the circularity of the bow, however, Grosseteste seems to think of a conical surface. The interplay of the four transparent media through which rays travel seems highly implausible for us today. Here three refractions are suggested, but in other works, he maintains that the rainbow is formed by repercussion (i.e. reflection), like the echo.

For the formation of colours Grosseteste employs a modificationist framework (see problem on page 205). In the last paragraphs the impossibility of painting the rainbow-colours shows how alive still the distinction between apparent and real colours were.

Although a short and innovative treatise, Grosseteste’s work did not include a close study of the “innumerable small mirrors” which give rise to the colours of the bow. The next section focuses on one of the most significant theories studying the individual raindrops in the work of Descartes.

4 Descartes

René Descartes (1596-1650) is most often encountered in philosophy curricula today. There he is generally treated as the par excellence rationalist philosopher. As opposed to empiricism, which sees observation and experiments as essential to the enterprise of knowledge, rationalism holds that the source of knowledge is the intellect. From the numerous writings Descartes had published mostly the philosophical ones outlining his epistemological program are still fruitfully influencing today’s thought. In the seventeenth century, however, Descartes’ influence was much more far-reaching and stronger. His mathematical works were considered revolutionary, just as his natural philosophy. Descartes carried out numerous experiments, mostly in areas like optics and physiology.

A problematic point with Cartesian science and scientific methodology is that Descartes rarely connected the two. From the four volumes of the Discourse on method (see more in 4.2) only in one of the scientific works, in only one of its discourses is there a clear reference to the method.

This discourse being the one on the rainbow, it seems fruitful to investigate both the methodological and the scientific aspects.

4.1 Views on Scientific Theories and on Experimentation

Experiments and observation can be, in Descartes’s view, essential for the construction of a deductive science. Significantly, for Descartes, “deduction” does not mean that from certain premises the use of formal procedures
determines the outcome. No, deduction, like intuition, is a grasping of
the *truth* of certain propositions, and also of the inferential connections
between these propositions.

Scientific explanations are for Descartes *a priori*, but not in our (Kantian)
sense of the word, meaning “independent of experience or empirical
evidence”, but as providing the efficient cause of the phenomena (Clarke
1988).

The careful investigation of the problem leads the experimenter to-
wards the causes of phenomena. The experiments serve primarily to clarify
the questions that are to be answered. This can be seen as the reductive
part of a scientist’s work. The questions asked (and the experiments carried out)
lead the scientist from complex problems to simpler ones.

- **Problem 9** Trace the series of questions that are asked by Descartes
  in Text 2, and find their interconnections.

By finding the ever more basic, underlying questions, one is lead (of-
ten in several steps) to a starting point for the deductive scientist. At this
point, using certain (unquestioned) intuitions or knowledge that derives
from God, first the most basic questions, and from this the other questions
can be answered: this would constitute the constructive part of problem-
solving. The answers should not contradict the initial assumptions, should
not be derived by illogical methods, and should show that the implications
of the explanation are compatible with experience.

This way the result of the deduction should be able to account for the
experimental findings or observations that lead us to the formulation of
the questions. These results are supported by the data that were used in
the process to reduce the problem to a stage where they could be tackled
deductively. As he writes in the sixth part of the *Discourse*: “since experience
makes the majority of these effects very certain, the causes from which
I deduce them are used not so much to prove them as to explain them;
but, quite the contrary, it is the latter [the causes] which are proved by the
former” (AT VI 76, CSM I, 150\(^1\)).

- **Problem 10** Now summarise the series of answers found by Des-
cartes in his deduction in the second part of Text 2; once he solved

---

\(^{1}\)These abbreviations refer to standard editions of the Cartesian corpus: (Descartes
1964-76) and (Descartes 1985).
the problems concerning the nature of light, and how it passes through media with intuition.

For Descartes a scientific explanation specifies the efficient causes and describes mechanisms which account for causes. The description of an (unobservable) cause from which an observable effect follows is necessarily hypothetical, as in his later writings Descartes himself accepted:

But although the causes thus arrived at are hypothetical, to consider them “true” is not impossible. The success of the system can be considered a guarantee: “...if we use only those principles which seem to be most evident, if we deduce nothing from them unless by mathematical arguments, and if we find that those things we have thus deduced from them correspond accurately with all the phenomena of nature, then we would seem to insult God if we suspect that the causes which we have thereby discovered are false” (AT VIII A, 99, CSM I 255).

Descartes’s models are often of great originality. They explain phenomena with recourse to the movement and interaction of small particles. The micro-level particles have the same properties as those at the macro-level, and Descartes is consciously using transduction to find the properties of invisibly small entities: “I only compare some movements with other, or some shapes with others, etc.; that is to say, I compare those things which because of their small size are not accessible to our senses with those which are, and which do not differ from the former more than a large circle differs from a small one” (AT II, 367-8, CSMK 122, see also page 279).

To argue for the plausibility of the microscopic model he shows the plausibility of the macroscopic model—usually not arguing for the validity of the comparison. This is highly characteristic of the mechanistic and corpuscular traditions of the seventeenth century.

4.2 THE THEORY OF THE RAINBOW

One of Descartes’s most important work is the Discourse on the Method of Rightly Conducting the Reason, and Seeking Truth in the Sciences. As it is normally reprinted today (in fact, surprisingly often ever since the first edition in 1637) it is a purely philosophical work in a single volume. But the original work contained The Essays on this Method: Dioptrics, discussing problems of optics and publishing the sine law of refraction (see figure VII.2 and also page 240), Meteors analysing meteorological questions, and Geometry, the first published treatise on analytic geometry.
The eighth discourse in the *Meteors* or *Meteorology* is the only explicitly stated example that demonstrates the fruitfulness of Descartes’s method outlined in the *Discourse*.

Figure VII.2: The drawing from Descartes’s *Dioptice* (1692, Frankfurt) shows the connection between incident and refracted rays. The ratio of the lengths of segments opposite the angles of incidence and refraction are constant (e.g. KM:NL = AH:GI). Photo Deutsches Museum, Munich.

- **Problem 11** Find elements in the Cartesian account that have been (and could have been) conceived out earlier as well. What is radically new in his approach?

Descartes’s calculations answered a thousand-year puzzle: why the bow appears where it appears. His explanation of the colours of the bow belonged to the modificationist tradition analysed in more detail in the next chapter. This modificationist account was based on his experiments with prisms.

- **Problem 12** Discuss why Descartes switches to studying the colours produced in the prism as opposed to further studying the spherical glass. No hint is provided for this problem.

While groundbreaking in many respect, the Cartesian theory of the rainbow was not too popular in the decades after its publication, and discussions based on the Aristotelian theory still dominated in the seventeenth century.
5 Texts

5.1 Text 1: Robert Grosseteste

- The text is reproduced from (Grant 1974). The original footnotes have been retained.

Investigation of the rainbow\(^2\) is the concern of both the student of perspective and the physicist. It is for the physicist to know the fact and for the student of perspective to know the explanation. For this reason Aristotle, in his book on Meteorology, has not revealed the explanation, which concerns the student of perspective; but he has condensed the facts of the rainbow, which are the concern of the physicist, into a short discourse. Therefore, in the present treatise we have undertaken to provide the explanation, which concerns the student of perspective, in proportion to our limited capability and the available time.

In the first place, then, we state that perspective is the science based on visual figures and that this is subordinate to the science based on figures containing radiant lines and surfaces, whether that radiation is emitted by the sun, the stars, or some other radiant body. Nor is it to be thought that the emission of visual rays [from the eye] is only imagined and without reality, as those think who consider the part and not the whole. But it should be understood that the visual species [issuing from the eye] is a substance, shining and radiating like the sun, the radiation of which, when coupled with radiation from the exterior shining body, entirely completes vision.

Wherefore natural philosophers, treating that which is natural to vision (and passive), assert that vision is produced by intromission. However, mathematicians and physicists, whose concern is with those things that are above nature, treating that which is above the nature of vision (and active), maintain that vision is produced by extramission. Aristotle clearly expresses this part of vision that occurs by extramission in the last book of De animalibus, saying: “A deep-set eye sees from a distance; for its motion is neither divided nor destroyed, but a visual power leaves it and goes directly to the objects seen.”\(^3\) Again in the same book [Aristotle writes]: “The three senses referred to, namely vision, hearing, and smell, issue from the organs [of perception] as water issues from pipes, and for this reason long noses have a strong power of smell”\(^4\). Therefore, true perspective is concerned with rays emitted [by the eye].

---

\(^2\)I have translated this treatise from the Latin text edited by Ludwig Baur, pp. 72-78. Portions of this treatise have also been published in translation by Crombie, chapter 5; The majority of the manuscripts collated by Baur contain the title *De iride*; however, the British Museum MS contains the title *De iride et specula*, and Baur gives both titles.

\(^3\)*De generatione animalium* V.l.781a.1-2. 23.

Perspective has three principal subdivisions, according to the triple mode in which rays are transmitted to the visible object. For either transmission of the ray to the visible object is in a straight line through a transparent medium of a single kind interposed between the observer and the object; or in a straight line to a body having a spiritual nature which is such as to make it a mirror, from which the ray is reflected to the object; or through several transparent media of different kinds, at the junctions of which the visual rays are refracted to form angles, so that the ray reaches the object not by direct approach but by a path of several straight lines joined at angles.

The first subdivision comprises the science called “De visu”, the second subdivision the science called “De speculis”. The third part has remained untouched and unknown among us until the present time. Nevertheless, we know that Aristotle completed this third part, which surpasses the other parts in its exceeding subtlety and the greatly admired profundity of its natures. This part of perspective, if perfectly understood, shows us how to make very distant objects appear close, how to make nearby objects appear very small, and how to make a small object placed at a distance appear as large as we wish, so that it would be possible to read minute letters from incredible distances or count sand, seeds, blades of grass, or any minute objects. The way in which these astonishing things occur is made manifest as follows. The visual ray penetrating through several transparent substances of diverse natures is refracted at their junctions, and its parts, in the different transparent media existing at those junctions, are joined at an angle. This is revealed by that experiment which is considered fundamental in the book De speculis. If an object is placed in a vessel and the observer stations himself at a position from which the object cannot be seen, the object will become visible when water is poured in. The same thing is also revealed by the fact that the subject of a continuous ray is a substance possessing a single [homogeneous] nature. Therefore, at the interface between two transparent media of different kinds, a visual ray must be interrupted. However, since the entire ray is generated by a single principle, and its continuity cannot be entirely destroyed unless its generation ceases, the discontinuity of the ray at the interface between the two transparent media must be incomplete. But the only mean between complete continuity and complete

---


6These views, long marveled at in Roger Bacon, were obviously borrowed by the latter from Grosseteste.

discontinuity is a point touching the two parts of one ray, [joined] not in a straight line but at an angle.

The amount of divergence from rectilinearity of rays joined at an angle can be represented as follows. Imagine a ray from the eye incident through air on a second transparent medium, extended continuously and rectilinearly [into the second medium], and a line perpendicular to the interface drawn into the depth of the medium from the point at which the ray is incident on the [second] transparent medium. I say, then, that the path of the ray in the second transparent medium is along the line bisecting the angle enclosed by the ray which we have imagined to be extended continuously and rectilinearly and the perpendicular line drawn into the depth of the second transparent medium from the point of incidence of the ray on its surface.

That the size of the angle of refraction of a ray may be thus determined is evident from an experiment similar to those by which we have learned that reflection of a ray by a mirror occurs at an angle equal to the angle of incidence. The same thing is revealed to us by the following principle of natural philosophy, namely that every operation of nature takes place in a manner as limited, well-ordered, brief, and good as possible.

An object seen through several transparent media does not appear as it really is but appears to be situated at the intersection of the ray emitted by the eye, extended in a continuous straight line, and the line drawn perpendicularly from the visible object to the surface of the second transparent medium (that is, the one nearer the eye). This is revealed to us by similar reasoning and the same experiment as that by which we know that objects seen in mirrors appear at the intersection of the visual ray, extended in a straight line, and the perpendicular drawn [from the visible object] to the surface of the mirror.

Thus far we have considered the size of the angle of refraction at the interface between two transparent media and the place of appearance of an object seen through several transparent media. To these we add the following principles, which the optical theorist appropriates from the natural philosopher, namely that the size, position, and order of the visible object are determined

---

8 That is, the angle of refraction equals half the angle of incidence. Diagrams illustrating this conception appear in Crombie, page 121. This half-angle law of refraction is remarkably primitive by contrast with Ptolemy’s elaborate attempts to find, by empirical means, the true mathematical law of refraction; see Albert Lejeune, *Recherches sur la catoptrique grecque* (Bruxelles: Palais des académies, 1957), pp. 152-166. It seems doubtful that Grosseteste would have formulated his half-angle law had he known Ptolemy’s *Optica*.

9 This paragraph reveals that Grosseteste’s half-angle law of refraction was determined not through measurement but rationally, on grounds of symmetry and brevity of action. See Bruce S. Eastwood, “Grosseteste’s ‘Quantitative’ Law of Refraction: A Chapter in the History of Non-Experimental Science,” *Journal of the History of Ideas*, XXVIII (1967), 403-414.
from the size of the angle and the position and order of the rays under which the object is seen and that it is not great distance that makes an object invisible, except accidentally, but rather the smallness of the angle under which it is observed. From these principles, it is perfectly evident by geometrical reasoning how an object of known distance, size, and position will appear with respect to location, size, and position. From the same principles it is evident how one must shape transparent substances so that they will receive the rays emitted by the eye, according to the desired angle formed in the eye, and will draw the rays back [toward convergence] in any desired manner onto visible objects, whether these objects are large or small, far or near. Thus all visible objects are made to appear to the observer in the desired position and of the desired size; and very large objects, if so desired, are caused to appear very small; and conversely, very small objects placed far away are caused to appear large and very easily perceptible by sight.

The science of the rainbow is subordinate to this third part of perspective [that is, to the science of refraction]. Now, a rainbow cannot be formed by means of solar rays falling in a straight line from the sun into the concavity of a cloud, for they would produce a continuous illumination of the cloud in the shape of the opening toward the sun through which the rays enter the concavity of the cloud rather than in the shape of a bow. Nor can a rainbow be formed by the reflection of solar rays from the convexity of the mist descending from a cloud as from a convex mirror, in such a way that the concavity of the cloud would receive the reflected rays and the rainbow would thus appear; for if that were so, the rainbow would not always be in the form of an arc, and as the sun rose the rainbow would become proportionately larger and higher, and as the sun set the rainbow would become smaller; the contrary, however, is evident to sense. Therefore, the rainbow must be formed by the reflection of solar rays in the mist of a convex cloud. I maintain that the outside of a cloud is convex and the inside concave, as is evident from the nature of light and heavy. And that which we see of a cloud must be less than a hemisphere, although it may look like a hemisphere; and since the mist descends from the concavity of the cloud, it must be pyramidally convex at the top, descending to the earth, and therefore more condensed near the earth than in the higher part.

---

10This is roughly the Aristotelian view, expressed in the Meteorologica III.2-5. On the history of the theory of the rainbow, see Carl B. Boyer, The Rainbow: From Myth to Mathematics (New York: Yoseloff, 1959), and Selection 63.2 from the De iride of Theodoric of Freiberg.

Therefore, there are, in all, four transparent media through which a ray of the sun penetrates: [first,] pure air containing the cloud; second, the cloud itself; third, the higher and rarer mist coming from the cloud; and fourth, the lower and denser part of the same mist. Therefore, in accordance with what was said before about the refraction of rays and the size of the angle of refraction at the interface between two transparent media, solar rays must be refracted first at the interface between the air and the cloud and then at the interface between the cloud and the mist. By these refractions the rays converge in the density of the mist and, being refracted there once more as from the vertex of a pyramid, spread out not into a found pyramid but into a figure like the curved surface of a found pyramid\textsuperscript{12} expanded opposite the sun. Therefore it assumes the shape of an arc, and in these [northern] regions the rainbow does not\textsuperscript{13} appear in the south. And since the vertex of the aforementioned figure is near the earth and it is expanded opposite the sun, half the figure or more must fall on the surface of the earth and the remaining half or less onto a cloud opposite the sun. Therefore, near sunrise or sunset the rainbow appears semicircular and is larger, and when the sun is in other positions the rainbow appears as part of a semicircle; and the higher the sun, the smaller the [visible] part of the rainbow. For this reason, in regions where the sun closely approaches the zenith, the rainbow never appears at noon. Aristotle’s claim that the variegated rainbow is of small measure at sunrise and sunset is not to be interpreted as smallness of size, but as smallness of luminosity, which occurs because the rays pass through a greater multitude of vapors at this time of day than at other times. Aristotle himself indicates this later, when he says that this [smallness] occurs because of the diminution of that which shines from the solar ray in the clouds.

However, since color is light mixed with a transparent medium—the transparent medium being diversified according to purity and impurity, while light is divided in a fourfold manner (according to brightness and darkness and according to multitude and paucity), and all colors being generated and diversified according to the combinations of these six differences—the variety of colors in different parts of one and the same rainbow occurs chiefly because of the multitude and paucity of solar rays. For where there is a greater multiplication of rays, the color appears clearer and more luminous, and where there is a smaller multiplication of rays, the color appears more bluish and obscure. And since the multiplication of light and the diminution determined by this multiplication result only from the splendor of the illumination on the mirror or from the transparent medium: which collects light in a certain place because of its shape [that is, the shape of the medium] and, after it has come together, diminishes it

\textsuperscript{12}That is, a hollowed-out cone.

\textsuperscript{13}Baur’s text does not include the negative, but Crombie (p. 126) argues for its inclusion and adduces manuscript support for his position.
by separation, and since this disposition for the reception of light is not fixed, it is manifest that painters do not have the ability to represent the rainbow; nevertheless, its representation according to a fixed disposition is possible.

Actually, the difference in color between one rainbow and another arises both from the purity and impurity of the recipient transparent medium and from the brightness and darkness of the light impressed on the medium. For if the transparent medium is pure and the light is bright, the color of the rainbow will be whitish and similar to light. But if the recipient transparent medium should contain a mixture of smoky vapors and the light is not very bright, as occurs near the rising or setting [of the sun], the color of the rainbow will be less brilliant and darker. Similarly, [the production of] all variations in color of the variegated bow is sufficiently apparent from other combinations of brightness and darkness of light and purity and impurity of the transparent medium.

Here ends the treatise on the rainbow by [the Bishop] of Lincoln.

- **Hint to Problem 6** Grosseteste is familiar with both intromissionist and extramissionist theories. He subscribes to the widespread medieval species-theory, owing a lot to atomist theories, like the eidola of Lucretius. But Grosseteste also maintains that the radiating image from the eye is coupled with the radiation from light-sources, thus reviving the mixed Platonic theory. As the behaviour of rays and images can be investigated, the geometrical optical tradition is also used. It is also employed when invisibility is discussed (compare it to Euclid’s proposition III. on page 123: “it is not great distance that makes an object invisible, …but rather the smallness of the angle under which it is observed.”)

- **Hint to Problem 8** The three possibilities correspond to the three branches of perspectiva. The propagation of rays in a straight line is ruled out as the shape of the bow cannot be explained this way. Reflection from the from the convexity of the mist descending from a cloud would result in small rainbows in case the sun is near the horizon and large bows when it is close to the zenith: this is against the observations. Ruling out these two possibilities leaves only one—and for Grosseteste this indirect proof suffices to validate his search for an explanation is a specific direction. His conviction, that the

---

14 According to Baur, the manuscripts collated for his text are unanimous in reading *non fixam*. However, omission of the negative seems to be required if sense is to be made of the sentence.
The rainbow is such a remarkable phenomenon of nature, and its cause has always been so carefully sought after by good minds, yet so little understood, that I could not choose anything better to show you how, by means of the method I am using, we can arrive at knowledge not possessed by any of those whose writings we have. First, taking into consideration that this arc can appear not only in the sky but also in the air near us whenever there are drops of water.
in it that are illuminated by the Sun—as we can observe in certain fountains—it was easy for me to judge that it came merely from the way in which rays of light act against those drops, and from there tend toward our eyes. Then, knowing that these drops are round, as we have demonstrated above, and seeing that their size does not affect the appearance of the arc, I decided to make a very large [drop] so as to be able to examine it better. For this purpose, I filled a perfectly round and transparent large flask with water, and I discovered that, for example, when the Sun came from the part of the sky marked AFZ [see Fig. VII.3], my eye being at point E, when I placed this globe at the spot BCD, its part D appeared to me completely red and incomparably more brilliant than the rest; and I found that whether I approached it or drew back from it, whether I placed it to the right or to the left, or even made it turn around my head, provided that the line DE always made an angle of approximately 42° with the line EM, which one must take to extend from the centre of the eye to the centre of the Sun, this part D always appeared equally red. But as soon as I made this angle DEM slightly larger, the red disappeared. And if I made it slightly smaller, it did not disappear immediately, but rather divided first into two less brilliant parts, in which yellow, blue, and other colours were to be seen.

Next, looking at the part of the globe marked K, I perceived that if I made the angle KEM around 52°, this part K would also appear red, but not as brilliantly as at D; and if I made it slightly larger, other fainter colours would appear; but I found that if I made it very slightly smaller still, or very much larger, none at all would appear. From this, I readily understood that if all the air around M were filled with such globes—or, in their place, drops of water—a strong red and very brilliant point must appear in each of those drops from which lines drawn toward the eye E make an angle of around 42° with EM, as I assume to be the case with those marked R. And if these points are viewed all together, without our noting anything about their position except the angle at which they are seen, they must appear as a continuous band of red. And in the same way there must be points in those drops marked s and T from which lines drawn to E make slightly more acute angles with EM, and these points make up bands of weaker colours. And in this consists the primary or main rainbow. And then again I found that if the angle MEX was 52°, a red band must appear in the drops marked X, and other bands of fainter colours in the bands marked Y, and that this is what the secondary or minor rainbow consists in. And finally, no colours at all can appear in all the other drops, marked V.

After I had done this, I examined in more detail what caused the part D of the globe BCD to appear red, finding that it was the rays of the Sun which, coming from A toward B, were bent on entering the water at point B, and went toward C, from where they were reflected toward D; and there, being bent again as they left the water, they tended toward E. For when I put an opaque or dark body in
some place on the lines AB, EC, CD, or DE, this red colour would disappear. And even if I covered the whole globe, except for the two points, and put dark bodies everywhere else, provided that nothing hindered the action of the rays ABCDE, the red colour appeared nevertheless. Then, looking also for the cause of the red that appeared at K, I found that it was the rays coming from F toward G, which were bent there toward H, and at H they were reflected toward I, at I it was reflected again toward K, and then finally bent at point K and tended toward E. So that the primary rainbow is caused by the rays reaching the eye after two refractions and one reflection, and the secondary by rays reaching it only after two refractions and two reflections; which is what prevents the second appearing as clearly as the first.

But the principal difficulty still remained: namely, to understand why, when there are many other rays there which, after two refractions and one or two reflections, can tend toward the eye when the globe is in a different position, it is nevertheless only those of which I have spoken that cause certain colours to appear. To resolve this difficulty, I looked to see if there was something else in which they appeared in the same way, so that by comparing these with each other I would be in a better position to gauge their cause. Then, remembering that a prism or triangle of crystal causes similar colours to be seen, I considered one of them such as MNP, which has two completely flat surfaces, MN and NP, inclined to one another at an angle of around 30° or 40°, so that if the rays of the Sun ABC cross MN at right angles, or almost at right angles, so that they do not undergo any noticeable refraction there, they must suffer a reasonably large refraction on leaving through NP. And when I covered one of these two surfaces with a dark body, in which there was a rather narrow opening DE, I observed that the rays, passing through this opening and from there making for the cloth or paper FGH, paint all of the colours of the rainbow on it; and that they always paint the colour red at F, and blue or violet at H. From this I learned, first, that the surfaces of the drops of water do not need to be curved in order to produce these colours, as those of this crystal are completely flat. Nor does the angle under which they appear need to be of any particular size, for this can be changed without any change in them, and although we can make the rays travelling toward F bend more or less than those travelling toward H, they nevertheless always colour it red, and those going toward H always colour it blue. Nor is reflection necessary, for there is no reflection here, nor finally do we need many refractions, for
there is only one refraction here. But I reasoned that there must be at least one refraction—and, in fact, one whose effect was not destroyed by another—for experience shows that, if the surfaces MN and NP are parallel, the rays, being straightened as much in one as they were able to be bent in the other, would not produce these colours. I did not doubt that light was also needed, for without it we see nothing. And moreover, I observed that shadow, or some limitation on this light, was necessary; for if we remove the dark body from NP, the colours FGH cease to appear; and if the opening DE is made large enough, the red, orange, and yellow at F reach no further because of that, any more than do the green, blue, and violet at H. Instead, all the extra space at G between these two remains white. After this, I tried to understand why these colours are different at H and at F, even though the refraction, shadow, and light combine there in the same way. And conceiving the nature of light to be such as I described it in the *Dioptrics*, namely as the action of motion of a certain very subtle matter, whose parts should be imagined as small balls rolling in the pores of terrestrious bodies, I understood that these balls can roll in different ways, depending on the causes that determine them; and in particular that all the refractions that occur on the same side cause them to turn in the same direction. But when they have no neighbouring balls that move significantly faster or slower than they, their rotation is approximately equal to their rectilinear motion, whereas when they have some on one side that move more slowly, and others on the other side that move as fast or faster, as happens when they are bounded by shadow and light, then when they encounter those which are moving more slowly on the side toward which they are rolling, as do those making up the ray EH, this causes them to rotate less quickly than if they were moving in a straight line. And the opposite happens when they encounter them on the other side, as do those of ray DF.

To understand this better, imagine the ball 1234 being propelled from V toward X, in such a way that it travels only in a straight line, and that its two sides 1 and 3 descend equally quickly toward the surface of the water YY, where the movement of the side marked 3, which encounters it first, is retarded, while that of side 1 still continues; this causes the whole ball to begin inexorably to rotate following the numbers 1234. Then, imagine it is surrounded by four others—Q, R, S, T—of which Q and T tend to move toward X with a greater force than does the ball, and the other two—S
and T—tend there with less force. It is clear from this that Q, which presses the part of the ball marked 1, and S, which retains that marked 3, increase its rotation; and that R and T do not hinder it, because R is disposed to move toward X faster than the ball follows it, and T is not disposed to follow the ball as quickly as it precedes it. This explains the action of the ray DF. And on the other hand, if Q and R tend more slowly than it toward X, and S and T tend there more rapidly, R hinders the motion of that part marked 1, and T that of part 3, without the two others—and S—doing anything. This explains the action of the ray EH. But it is worth noting that since this ball 1234 is quite round, it can easily happen that, when it is pressed hard by the two balls R and T, it is turned and rotates around the axis 42, rather than their causing its rotation to stop. And so, changing its position in an instant, it subsequently rotates following the numbers 321; for the two balls R and T, which caused it to begin to rotate, make it continue until it has completed a half-turn in this direction, and then they can increase its rotation instead of retarding it. This enabled me to resolve the major difficulty that I had in this matter. And it seems to me that it is very evident from all of this that the nature of the colours appearing at F consists just in the parts of the subtle matter which transmit the action of light having a much greater tendency to rotate than to travel in a straight line. As a consequence, those which have a much stronger tendency to rotate cause the colour red, and those which have only a slightly stronger tendency cause yellow. The nature of those that are visible at H, on the other hand, consists just in the fact that these small parts do not rotate as quickly as normal, when there is no particular cause hindering them; so that green appears when they rotate just a little more slowly, and blue when they rotate very much more slowly. And usually this blue is combined with a pinkish colour at its edges, which makes it vivacious and sparkling, and changes it into violet or purple. The cause of this is without doubt the same as that which usually stows down the rotation of the parts of the subtle matter when it has enough strength to change the position of some of them and increase their rotation, while slowing that of others. And the explanation agrees so well with observation in all of this that I do not believe it possible, after one has attended carefully to both, to doubt that things are such as I have explained. For if it is true that the sensation we have of tight is caused by the movement or inclination to movement of some matter touching our eyes, as is indicated by many other things, it is certain that different movements of this matter must cause different sensations in us. And as these movements cannot differ other than in the way I have mentioned, we observe no difference in the sensations we have of them other than a difference in colour. And we can find nothing at all in the crystal MNP that can produce colours except the way in which it sends the tiny bits of subtle matter toward the line FGH, and from there toward our eyes. From this, it seems to me obvious enough that we should not look for anything else in the colours that other objects make appear; for ordinary
observation shows that tight or white, and shadow or black, together with the
colours of the rainbow that have been explained here, are enough to make
up all the others. And I cannot accept the distinction the Philosophers make
between true colours and others which are only false or apparent. For because
the entire true nature of colours consists only in their appearance, it seems to
me to be a contradiction to say both that they are false and that they appear. But
I acknowledge that shadow and refraction are not always necessary to produce
them, and that instead of these, the size, shape, situation, and movement of the
parts of the bodies that one terms ‘coloured’ can combine in various ways with
light to increase or diminish the rotation of the parts of the subtle matter. So,
even in the rainbow, I initially doubted whether the colours there were produced
in the same way as in the crystal MNP; for I did not notice any shadow cutting
off the light, nor did I yet understand why they appeared at different angles,
until, having taken my pen and calculated in detail all the rays that fall on the
various points on a drop of water, in order the find out at what angles they would
come to our eyes after two refractions and one or two reflections, I found that
after one reflection and two refractions, far more of them can be seen at the
angle of 41° to 42° than at any smaller one; and that none of them can be seen
at a larger angle. Next I also found that after two reflections and two refractions,
far more of them come to the eye at an angle of 51° to 52° than at any greater
one; and no such rays come at a smaller one. So that there is a shadow on both
sides, cutting off the light which, having passed through innumerable raindrops
illuminated by the Sun, comes toward the eye at an angle of 49° or slightly
less, thus causing the primary, main rainbow. And there is also one cutting off
the light at an angle of 52° or slightly more, which causes the outer rainbow;
for failing to receive rays of light in your eyes, or receiving very much fewer
of them from one object than from another which is near it, is to see a shadow.
This is a clear demonstration that the colours of these arcs are produced by
the same cause as those that appear with the aid of the crystal MNP, and that
the radius of the inner arc must not be greater than 42°, nor that of the outer
one less than 51°, and finally that the outside surface of the primary rainbow
must be much more restricted than the inside one; and the opposite in the case
of the secondary one, as observation shows us. But so that those who have a
knowledge of mathematics can understand whether the calculation I have made
of these rays is sufficiently exact, I should explain it here.
Let AFD be a drop of water whose radius CD or AB I divide into as many equal parts as I wish to calculate rays, so as to attribute an equal amount of light to them all. Then I consider one of these rays in detail: EF, for example, instead of passing directly through G, is deflected toward K, is reflected from K toward N, from where it goes toward the eye P; or alternatively, it is reflected once more from N to Q, and from there is turned toward the eye R. And having drawn CI at right angles on FK, I know from what was said in the Dioptrics that the ratio between AE (or HF) and CI is that by which the refraction of water is measured. So that if HF contains 8,000 parts—taking AB to contain 10,000—CI will contain around 5,984 because the refraction of water is slightly greater than 3/4. On the most exact measurement I have been able to make, it is 187/250. Having thus the two lines HF and CI, I could easily find the size of the two arcs, FG, which was 73° 44’, and FK, which was 106° 30’. Then, taking double the arc FK from the arc FG added to 180°, I obtain 40° 44’ for the size of the angle ONP, on the assumption that ON is parallel to EF. And taking this 40° 44’ from FK, I have 65° 46’ for the angle SQR, assuming that SQ is parallel to EP. And doing the same calculation for all the other rays parallel to EF which pass through the divisions of the diameter AB in the same way, I compile a table [below].

<table>
<thead>
<tr>
<th>Line</th>
<th>Line</th>
<th>Arc</th>
<th>Arc</th>
<th>Angle</th>
<th>Angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF</td>
<td>CI</td>
<td>FG</td>
<td>FK</td>
<td>ONP</td>
<td>SQR</td>
</tr>
<tr>
<td>1000</td>
<td>748</td>
<td>168°30’</td>
<td>171°25’</td>
<td>5°40’</td>
<td>165°45’</td>
</tr>
<tr>
<td>2000</td>
<td>1496</td>
<td>156°55’</td>
<td>162°48’</td>
<td>11°19’</td>
<td>151°29’</td>
</tr>
<tr>
<td>3000</td>
<td>2244</td>
<td>145°4’</td>
<td>154°4’</td>
<td>17°56’</td>
<td>136°8’</td>
</tr>
<tr>
<td>4000</td>
<td>2992</td>
<td>132°5’</td>
<td>145°10’</td>
<td>22°30’</td>
<td>122°4’</td>
</tr>
<tr>
<td>5000</td>
<td>3740</td>
<td>120°</td>
<td>136°4’</td>
<td>27°52’</td>
<td>108°12’</td>
</tr>
<tr>
<td>6000</td>
<td>4488</td>
<td>106°16’</td>
<td>126°40’</td>
<td>32°56’</td>
<td>93°44’</td>
</tr>
<tr>
<td>7000</td>
<td>5236</td>
<td>91°8’</td>
<td>116°51’</td>
<td>37°26’</td>
<td>79°25’</td>
</tr>
<tr>
<td>8000</td>
<td>5984</td>
<td>73°44’</td>
<td>106°30’</td>
<td>40°44’</td>
<td>65°46’</td>
</tr>
<tr>
<td>9000</td>
<td>6732</td>
<td>51°41’</td>
<td>95°22’</td>
<td>40°57’</td>
<td>54°25’</td>
</tr>
<tr>
<td>10000</td>
<td>7480</td>
<td>0°</td>
<td>83°10’</td>
<td>13°40’</td>
<td>69°30’</td>
</tr>
</tbody>
</table>

It can be readily seen from table [I] that there are many more rays making
the angle ONP around 40° than there are those making it less; and also more of them that make SQR around 54° than make it larger. Then, so as to make it still more precise, we have [table II].

And I see here that the largest angle, ONP, can be 41° 30′, and the smallest, SQR, 51° 54′; when I add or subtract around 17′ for the radius of the Sun, I have 41° 47′ for the largest radius of the inner rainbow; and 51° 37′ for the smallest radius of the outer one.

<table>
<thead>
<tr>
<th>Line</th>
<th>Line</th>
<th>Arc</th>
<th>Arc</th>
<th>Angle</th>
<th>Angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF</td>
<td>Cl</td>
<td>FG</td>
<td>FK</td>
<td>ONP</td>
<td>SQR</td>
</tr>
<tr>
<td>8000</td>
<td>5984</td>
<td>73° 44′</td>
<td>106° 30′</td>
<td>40° 44′</td>
<td>65° 46′</td>
</tr>
<tr>
<td>8100</td>
<td>6058</td>
<td>71° 48′</td>
<td>105° 25′</td>
<td>40° 58′</td>
<td>64° 37′</td>
</tr>
<tr>
<td>8200</td>
<td>6133</td>
<td>69° 50′</td>
<td>104° 20′</td>
<td>41° 10′</td>
<td>63° 10′</td>
</tr>
<tr>
<td>8300</td>
<td>6208</td>
<td>67° 48′</td>
<td>103° 14′</td>
<td>41° 20′</td>
<td>62° 54′</td>
</tr>
<tr>
<td>8400</td>
<td>6283</td>
<td>65° 44′</td>
<td>102° 09′</td>
<td>41° 26′</td>
<td>61° 43′</td>
</tr>
<tr>
<td>8500</td>
<td>6358</td>
<td>63° 34′</td>
<td>101° 2</td>
<td>41° 30′</td>
<td>60° 32′</td>
</tr>
<tr>
<td>8600</td>
<td>6432</td>
<td>61° 22′</td>
<td>99° 56′</td>
<td>41° 30′</td>
<td>58° 26′</td>
</tr>
<tr>
<td>8700</td>
<td>6507</td>
<td>59° 4′</td>
<td>98° 48′</td>
<td>41° 28′</td>
<td>57° 20′</td>
</tr>
<tr>
<td>8800</td>
<td>6582</td>
<td>56° 42′</td>
<td>97° 40′</td>
<td>41° 22′</td>
<td>56° 18′</td>
</tr>
<tr>
<td>8900</td>
<td>6657</td>
<td>54° 16′</td>
<td>96° 32′</td>
<td>41° 12′</td>
<td>55° 22′</td>
</tr>
<tr>
<td>9000</td>
<td>6732</td>
<td>51° 41′</td>
<td>95° 22′</td>
<td>40° 45′</td>
<td>54° 20′</td>
</tr>
<tr>
<td>9100</td>
<td>6806</td>
<td>49° 0′</td>
<td>94° 12′</td>
<td>40° 36′</td>
<td>53° 36′</td>
</tr>
<tr>
<td>9200</td>
<td>6881</td>
<td>46° 8′</td>
<td>93° 2′</td>
<td>40° 4′</td>
<td>52° 38′</td>
</tr>
<tr>
<td>9300</td>
<td>6956</td>
<td>43° 8′</td>
<td>91° 51′</td>
<td>39° 26′</td>
<td>52° 52′</td>
</tr>
<tr>
<td>9400</td>
<td>7031</td>
<td>39° 45′</td>
<td>90° 38′</td>
<td>38° 38′</td>
<td>52° 0′</td>
</tr>
<tr>
<td>9500</td>
<td>7106</td>
<td>36° 24′</td>
<td>89° 26′</td>
<td>37° 32′</td>
<td>51° 54′</td>
</tr>
<tr>
<td>9600</td>
<td>7180</td>
<td>32° 30′</td>
<td>88° 12′</td>
<td>36° 6′</td>
<td>52° 6′</td>
</tr>
<tr>
<td>9700</td>
<td>7255</td>
<td>28° 8′</td>
<td>86° 58′</td>
<td>34° 12</td>
<td>52° 46′</td>
</tr>
<tr>
<td>9800</td>
<td>7330</td>
<td>22° 57′</td>
<td>85° 43′</td>
<td>31° 31′</td>
<td>54° 12′</td>
</tr>
</tbody>
</table>

It is true that when the water is warm its refraction is slightly less than when it is cold, which can alter certain things in the calculation. Nevertheless, it will only increase the radius of the inner rainbow by one or two degrees at the most, in which case that of the outer rainbow will be nearly twice that of the smaller. This is worth noting because by these means we can demonstrate that the refraction of water can be hardly any more or less than I have supposed. For if it were slightly larger, it would make the radius of the inner rainbow less than 41°, whereas the common belief is that it is 45°; and if we assume it to be small enough to make it exactly 45°, we will find that the radius of the outer arc is also hardly more than 45°, whereas it appears much larger to the eye.
than the inner one. And Macrolius, who I believe is the first to have determined
the figure of $45^\circ$, determines the other to be around $56^\circ$, which shows how
little faith we should have in observations which are not accompanied by true
rationale. For the rest, I have had no difficulty in understanding why red is on
the outside of the inner arc, nor why it is on the inside on the outer one; for the
same thing that causes it to be near $F$ rather than $H$ when it appears through the
crystal $MNP$, also causes us to see the red toward its thicker part $MP$, and blue
toward $N$, when we look at this crystal with our eye at the white screen $FGH$,
because the ray coloured red which goes toward $F$ comes from $C$, the part of
the Sun that is closest to $MP$. And this same cause brings it about that, when
the centre of the drops of water—and as a result their broadest part—are on the
outside of the coloured points forming the interior rainbow, the red must appear
there on the outside; and that when they are on the inside of those that make up
the outer rainbow, the red must correspondingly appear on the inside.

Thus I believe that there remains
no difficulty in this matter, unless
it perhaps concerns the irregularities
which one encounters here: for ex-
ample, when the arc is not exactly
found, or when its centre is not in a
straight line passing through the eye
and the Sun, which can occur if the
wind changes the shape of the rain-
drops. For losing the smallest part of
their roundness must make a significant difference in the angle at which the col-
ours must appear. I have been told that there has sometimes also been observed
a rainbow so reversed that its ends were turned upwards, as represented here
at $FF$. I can only account for this in terms of the reflection of the rays of the
Sun falling on the water of the sea or some take. Assume they come from the
part of the sky $SS$, fall on the water $DAE$, and from there are reflected toward
the rain $CF$; then the eye $B$ will see the arc $FF$, whose centre is at point $C$,
so that if $CB$ is projected right to $A$, and $AS$ passes through the centre of the
Sun, the angles $SAD$ and $BAE$ are equal, and so the angle $CBF$ is around $42^\circ$.
Nevertheless, for this effect to occur there must also be absolutely no wind to
disturb the surface of the water $E$, and there must also perhaps be a cloud
such as $G$ which prevents the Sun’s light, which travels in a straight line toward
the rain, from effacing the light that this water $E$ sends there. Consequently, this
is a rare occurrence. Besides, it is possible for the eye to be in such a position
with respect to the Sun and the rain that the lower part, where the band of the
rainbow is terminated, is seen, but not the upper part; and then it will be taken
to be an inverted arc, even though we do not see it near the sky, but near the
water or the earth. I have also been told that a third rainbow has sometimes been
seen above the two usual ones, but that it was much fainter, and approximately as distant from the secondary one as that is from the primary. I do not think this could have happened unless there had been numerous found and transparent grains of hail mixed in with the rain. Since the refraction in these is significantly greater than that in water, the outer rainbow must have been very much larger there, and so appears above the others. As for the inner one, which for the same reason would have to have been smaller than that of the fain, it possibly will not have been noticed, because of the great lustre of the outer one. Or alternatively, because their edges are joined, the two of them will be counted as one, but one whose colours are arranged differently than is usual.

[The final paragraph of Discourse 8 describes how to produce optical illusions with fountains.]

- **Hint to Problem 9** According to (Garber 1988) the Cartesian account asks the following questions (references and partial answers are given in square brackets):

  Q1: What causes the rainbow (two regions of colors)? [Rainbows appear only in the presence of water droplets; size is irrelevant to the phenomenon].

  Q2: What causes the two regions of colour in any spherical ball?

  Q2a: What causes the two regions?

  (The two regions result from two combinations of refraction.)

  Q2b: What causes the colour?

  (Colour is produced without a curved surface and without reflection; it requires a restricted stream of light, and a refraction.)

  Q3a: Why do the two combinations of reflection and refraction result in two discrete regions?

  Q3b: How does refraction cause colour under appropriate circumstances?

  Q4: How does light pass through media?

  Q5: What is light?
• **Hint to Problem 10** Also based on (Garber 1988):

*Intuition:* The nature of light, and how it passes through media [Cf. Q5, Q4]

D1a: Law of refraction

D1b: The only change in a restricted stream of light passing from one medium to another (refraction aside) is a differential tendency to rotation.

D2a: All parallel rays of light converge into two discrete streams after two refractions and one or two reflections, emerging from the drop (flask) in two discrete regions. [Cf. Q3a]

D2b: Colour can only be the differential tendency to rotation produced in passing from one medium to another in refraction. [Cf. Q3b]

D3: Parallel rays of light produce two discrete regions of colour on a spherical ball of water. [Cf. Q2]

D4: Sunlight (parallel rays of light) on a region of water droplets will produce two regions of colour. [Cf. Q1]

• **Hint to Problem 11** Descartes presents most of his work on the rainbow as original contribution. He claims to be the first to study the bow experimentally by analysing the rays of light in a single raindrop. Similar studies, however, had been carried out hundreds of years earlier by Theodoric from Freiberg and the Arab scientist Kamal al-Din. His values for the angle under which the primary and secondary bow appear had also been closely approximated in a number of works that Descartes does not mention.

One of his innovations was to carry out laborious studies of the paths of several single rays. He found that for the angles between the rays of the Sun and the rays leaving the drops suffering two refractions and one reflection there is a maximum value around 42°. For two refractions and two reflections the minimum angle is around 52°. This explains why the upper edge of the primary bow and the lower edge of the secondary bow appears sharp. This also explains why the dark band of Alexander appears between the bows.
A natural supposition would be that for these calculations the correct law of refraction had to be discovered. However, even the Ptolemaic table of refractions would yield a value around 42° for the primary bow (though for the secondary bow the value is less precise).

6 Further Problems

- **Problem 13** Compare the possible interpretations of the scheme given by Grosseteste. Evaluate the arguments of the different historians. Use the texts mentioned in footnote 11 and (Boyer 1959). You can extend the work to include other sources as well.

- **Problem 14** Carl Boyer in an influential monograph on the rainbow writes about Descartes (Boyer 1959):

  “One of his weaknesses was the wholesale postulation of microcosmic particles with complicated qualitative properties of the macrocosm. But how false his emphasis on matter and motion here played him! There was a characteristic of the rainbow which Descartes largely disregarded—its width—and it was in this narrow band, rather than in rotary motion, that the secret of color lay unnoticed. Had he measured what was measurable, instead of speculating on a motion which could be neither seen nor measured, he might have discovered the secret.”

Discuss this way of writing history, Boyer’s critique, and the plausibility of his suggestion.

7 Suggested Readings

For the cultural importance of the rainbow see volumes like (Lee & Fraser 2001). The best general introductory book to medieval science is most probably (Lindberg 1992). As a source book (Grant 1974) very useful.

In the recent years more and more of the original medieval treatises have become available for the English reader. For translations of original works see (Pecham 1970, Alhazen 1989, Bacon 1996).

Reprinted selections of essays include (Eastwood 1989) (containing one of the articles referred to in Problem 13), as well as (Grant 1981), or (Crombie 1990). Some individual studies are in (Lindberg 1966).
For the seventeenth century (Sabra 1967) and (Lindberg 1976) are good starting points. For Descartes see (Garber 1988) and (Clarke 1988) in (Cottingham 1998). Also the articles by Armogathe and Schuster in (Sutton, Gaukroger & Schuster 2000). New monographs include (Gaukroger 2002), but a thorough listing of the “Descartes-industry” seems impossible, as, fortunately, the long neglected natural philosophy of Descartes has received increased attention in the last years.
Chapter VIII

THE MODIFICATIONIST TRADITION

1 INTRODUCTION

The scientific theories of colour before the age of Newton are relatively little studied. In general, it can be stated that a tradition from the time of Aristotle until well into the eighteenth century dominated the field. This modificationist tradition of colour held that colours are modifications of the white and homogeneous light. There was never scientific consensus as to what this modification precisely is, or whether it takes place on surfaces, in media, or at the edges of light and shadow. Several rival accounts existed, but this tradition characterised the treatises that attempted to give a causal account of colour phenomena. Unlike optics proper, these theories of colour were not parts of the mixed mathematical sciences, but often appeared appended to scientific theories of vision and optics.

The chapter will not provide a detailed study of the tradition, but only analyse two examples after a short historical introduction. The first one is Robert Hooke’s seventeenth century modificationist theory, one of the most detailed theories of the age. It appeared a few years before Newton’s radically new anti-modificationist theory—the subject of the next chapter. Hooke builds on the newly discovered law of refraction and the Cartesian theory of colours, discussed in the previous chapter (from page 171).

In the second example excerpts are given from Johann Wolfgang von Goethe’s monumental work on colour—a rich source of observations on colours, but also one of the last and most detailed modificationist theories. It is pronouncedly anti-Newtonian in its approach and consciously utilizes earlier modificationist theories.

As will be seen from the two examples, both are closely connected to the Newtonian theory: the one was replaced by it, the other was a—by and large—failed attempt to replace it.

---

1 As will be seen, the previous chapter also included modificationist theories. However, as different aspects of these writings were stressed there, questions connected to modificationism will be discussed here.
2. **The Beginnings of the Tradition**

As we have already seen in Chapter IV and VII, Aristotle in the third book of his *Meteorologica* discusses the formation of the rainbow, halos and other atmospheric phenomena. In Book III, Part 4 he writes that “the rainbow is a reflection of sight to the sun”. To explain the colours, he lays down the basic facts (374b9-14): “first, that white colour on a black surface or seen through a black medium gives red; second, that sight when strained to a distance becomes weaker and less; third, that black is in a sort the negation of sight; so everything at a distance looks black, because sight does not reach it”.

In this account we clearly have a modification (not of light, but of visual rays). The sources of modificationism, and with it the luminosity scale of colours thus reaches back at least to Aristotle. White is light, black the negation of it. The intermediary stages in luminance correspond to colours of different luminance. The decrease in luminance and thus the appearing colour in case of the rainbow is the result of the interaction of a “black medium”. Such accounts can be termed *medium-modificationist* theories.

3. **The Historical Development of the Theories**

Dennis Sepper gives a broad definition of modificationism: “the term characterises all theories that attribute color to some kind of modification or change of simple white light” (Sepper 1988, p. 108).

It is important to realise that while modificationism was widespread, it was hardly an elaborated theory. Writers usually restricted their claims about colours to a few passages. The reason for the lack of clearly articulated and detailed theories is manifold. First of all, the geometrical science of optics did not incorporate the complicated colour-appearances, as quantification seemed nearly impossible for centuries. Most great works like Euclid’s *Optics* had precious little to say about colour. Other treatises, like the first book of Ptolemy’s *Optics*, with a large section on colours, were lost (as we have learnt from Chapter VI).

Secondly, even though modificationism was widespread, no great synthesis had ever appeared. There were certain optical appearances which gave rise to so called *emphatical* or *apparent* colours. As opposed to these colours, which could not be considered as properties of surfaces or objects, the so called *real* colours were localisable on the outer surfaces of objects.
As Seneca defined the former sort of colours in his *Naturales questiones* 1.7.2 (Seneca 1971):

Obviously, a rainbow is caused by the sun. Obviously there must be something smooth and like a mirror which may reflect the sun. Finally, it is obvious that no actual colour is formed but only the appearance of a counterfeit colour, the sort that the neck of a dove alternately takes on or puts aside whenever it changes position, as I have already said. This also is the case in a mirror, which assumes no colour but only a kind of copy of the colour of something else.

Modificationist theories—with differing success—accounted for the colours of halos, coronae, prismatic colours, but most importantly of the rainbow between Ancient times and the seventeenth century.

For example, some time after 1304 Theodoric of Freiberg successfully explained the primary and secondary bows of the rainbow. His lost explanation—clearly a modificationist theory—was rediscovered only in 1814. Roger Bacon, who measured that the primary bow is seen at 42°, Robert Grosseteste, who stated in 1255 that the Sun’s rays are refracted as opposed to being reflected, or Albertus Magnus, who held that single droplets of rain are responsible for the rainbow, all subscribed to some sort of modificationism, yet no unified and generally accepted explanation existed.

- **Problem 1** Find elements of the modificationist tradition in Grosseteste’s explanation of the rainbow (see page 183).

- **Hint to Problem 1** For Grosseteste “color is light mixed with a transparent medium”. Three pairs of differences give the colours: the medium can be pure or impure, light can be bright or dark, much (multitude) or little (paucity). The upper band of the rainbow has more rays—it is more luminous (the red end), while the lower end has “smaller multiplication of rays”, and is “bluish and obscure”.

  Here modificationism is used to explain only the apparent colours with recurse to the properties of the medium and of the light rays.

Thirdly, the multitude and variability of colour phenomena simply defied any simplistic and general handling, as the history of colour and vision
science in the last centuries has clearly shown. For a successful theory, it was a precondition that certain classes of colour phenomena be disregarded, as Newton's theory in the next chapter beautifully illustrates this. Here the theory 'defined' what is to be seen, what legitimate colour phenomena are, and monstrosities, unexpected appearances of colour were relegated as illusions, malfunctioning of the sense of sight, or simply the result of incompetent experimentation.

![Colour Triangle](image.png)

Figure VIII.1: An early colour triangle from 1702 in the *Oculus artificialis stabilimentum*, 2nd ed. by Johannes Zahn, printed in Nurnberg.

Modificationism, however, offered a convenient way of ordering colours. Usually light intermixed with darkness yielded the other colours, and a number of basic schemata had been developed to arrange colours. Linear ordering according to luminance was common (with white and black at the extremities)\(^2\), white and black could also be at the two poles or apices of

---

\(^2\) The first colour-circles appeared by connecting the two ends of the colour-system, as can be seen in the early printed book of Robert Fludd, *Medicina Catholica*, Frankfurt: C.
the table and the colours arranged in the middle. As there is no obvious and unproblematic way of arranging colours according to their luminance only, often two-dimensional charts have been proposed. One example can be seen in Figure VIII.1. Originally appearing in 1685, the model includes the colours yellow, red, and blue at the base (the white-black axis). A number of other colours are present, like pale yellow, green, blue-black, and grey (Zahn 1658). The base of the triangle corresponds to a luminance scale, but points above each other are isoluminous, like grey, green and red, all being equidistant from white and black, i.e. light and lack of light.

These colour schemes might look utterly unscientific or unfounded, if one thinks of white and black as pigments. How could mixing these pigments yield different colours? As Newton expressed it in one of his early notebooks: “No color will arise out of the mixture of pure black and white, for then pictures, drawn with ink would be colored, or printed would seem colored at a distance, and the verges of shadows would be colored, and lamp-black and Spanish whiting would produce colours.” (McGuire & Tamny 1983, p. 389). But this is misunderstanding the gist of modificationism which treats white and black more as presence and absence of light and not as colour pigments.

4 Seventeenth Century Modificationism

Thomas Kuhn, in his introduction to Newton’s optical papers remarked that coloured fringes seen in the emergence of the spectrum seem to strengthen the “ancient theory of the nature of the rainbow’s colors, a theory which held that a succession of modifications of sunlight by the droplets of a rain cloud produced the colors of the bow” (Kuhn 1958, p. 30). He added that this theory, first employed to explain the colours of the rainbow, was present in all theories of colour in the 150 years preceding Newton’s work.

Seventeenth century modificationism was combined with the revival of corpuscular and mechanistic ideas as a result of the success of Epicurean atomism (rediscovered in the Renaissance, mostly through the writings of Lucretius), and the influential work of Gassendi and other atomists. Colours

Rötelli, printed in 1629 and 1631. Here white and black are next to each other in the circle.

3 Zahn, as has also been noted by (Gage 1993, p. 289), stated that most opposite colours are liveliest when juxtaposed, but did not further specify them. The colour-terms are: cinereus; aqueus; fuscus; incarnatus; viridis; subrubeus; subalbus; anxeus; porpurus; subcaeruleus; albus; flavus; rubus; caerules; niger.
were more and more often considered as properties of light and not of objects, but, again, several partially successful and only partly elaborated accounts existed.

In this period the so called *boundary-modificationist* accounts as opposed to earlier *medium-modificationism* were widespread. These explained the modification as the interaction of light with either surfaces or the boundary between the bundle of light and darkness. This type of corpuscular modificationism appeared in Gassendi’s work, and also in Descartes’s.

- **Problem 2** Summarize the modificationist elements in the Cartesian explanation of colour (use the text on page 189).

Walter Charleton—an early and important influence for Newton—followed Gassendi, claiming that intermediate colours, like red, blue, and green “are but the off-spring of the Extreme, arising from the intermission of light and shadow, in various proportions; or, more plainly, that the sense of them is caused in the Retina Tunica according to the variety or Reflections and refractions, that the incident Light suffers from the superficial particles of objects” (Charleton 1654 (1966), p. 191).

Robert Boyle, one of the leading experimental scientists was more cautious to take a stance in the subject. His extensive work *Touching Colours* was more a collection of observations than a theory. While accepting modificationism, he wrote: “But whether I think this Modification of the Light to be perform’d by Mixing it with Shades, or by Varying the Proportion of the Progress and Rotation of the Cartesian Glabuli Coelestes, or by some other way which I am not now to mention, I pretend not here to Declare.” (Boyle 1664, p. 90).

Isaac Barrow, Lucasian professor of Mathematics in Cambridge (and also Newton’s mentor, see page 348 for the history of this chair) wrote at the end of a section of his book on geometrical optics (Barrow 1987): As he writes:

    since the matter of colours has cropped up, suppose I hazard a few guesses about them—though uncharacteristically and out of place…

    Colour too is, it seems, practically nothing else but light impinging on rather larger bodies that it meets, retaining to some extent the stable position of their parts, and, according to the differing shape,
disposition or texture of the particles of which they consist, diverted or bouncing off in some way or other; with the result of course that the light that had fallen on these bodies comes out such as it does, whether in its motion, or its power of action, or simply in its quantity (I mean in regard to its rarity or density and the copiousness or scantiness of its rays), and according to the distinction of its type produces different appearances, which we denote by the various colour-names.

This chapter picks only one of these contemporary theories, the one proposed by Robert Hooke in his *Micrographia*, first published in 1665.

4.1 HOOKE’S THEORY

In the “scientific” literature (i.e. not in the work of artisans, painters, etc.) Hooke’s theory was one of the most complex and worked-out example of modificationism—and even this comprises only a part of observation IX in his *Micrographia* (Hooke 1665). It starts with a description of Moscovy-glass or mica. The mostly transparent substance exhibits varied colours when separated into thin sheets.

- **Problem 3** What characteristics of mica are discussed?

- **Problem 4** What colours are perceived in the thin sheets of mica? Are they the same as the ones in the rainbow?

After listing the observable phenomena, Hooke sets out to “examine the causes and reasons of them, and to consider, whether from these causes demonstratively evidenced, may not be deduced the true causes of the production of all kind of Colours.” (page 49 of the *Micrographia*, see page 218.)

- **Problem 5** Why does Hooke believe that finding the causes of this phenomenon may yield the solution (i.e. the true causes) for all colour appearances?

Starting from this—possibly chance—observation, Hooke sets out to find the material cause of the phenomena. He attempts to reproduce and control the experimental situation and achieve consistent results. In the artificial reproduction two transparent surfaces with a very thin layer in between (filled with air or other transparent media) gives rise to similar
colours, just as very thin films of several substances (soap bubbles, oily substances floating on water).

Hooke, however, as he later admits, fails to quantify the data—a few years later the young Newton with an ingenious setting finally measures the width of the extremely thin layer of air necessary for the production of colour.

An interlude discussing colour changes caused by the heating and tempering of metals shows how closely bound Hooke’s scientific interests were to technological problems.

- **Problem 6** What are the general conditions under which the colour-phenomena appear according to Hooke (from page 53 and 54 of the *Micrographia*)? How does Hooke reconstruct Descartes’s theory and how does he criticise it? Discuss his use of the term *experimentum crucis*.

After showing weaknesses of the Cartesian theory, Hooke sets out to develop his own theory of light. He uses the term *zetetically* to refer to a name adopted by some Pyrrhonists, and meaning a process of seeking or proceeding by inquiry. In mathematics the term is used for finding the value of unknown quantities by direct search. He postulates that “there is no luminous Body but has the parts of it in motion more or less” (p. 54) and the properties of this motion.

- **Problem 7** Summarize what Hooke has to say about the “motion of light”. Use pp. 55-57 in the *Micrographia*. Note that no hint is given for this problem.

After the propagation of light in an homogeneous medium has been cleared, Hooke discusses how “this pulse or motion will be propagated through differingly transparent *mediums*”. Although often propagated as a forerunner of the wave theory of light—made famous by the work of Young and Fresnel in the nineteenth century—, when closely examined Hooke’s theory is very different.

- **Problem 8** Discuss Hooke’s treatment of refraction. How does Hooke express the sine law? What are the basic tenets of his model of refraction?
Hooke, after describing a mathematical ray introduces the physical ray and states that: “as for the pulses themselves they will by the refraction acquire another property” crucial for the production of colour.

- **Problem 9** Reconstruct how Hooke explains the production of colour after refraction. What are the principal colours? How do other colours arise?

Hooke, after giving a phenomenological description of colour production in a vessel after refraction of incident light, deriving all the appearing colours from “principal” colours, and giving a micro-level explanation, uses the data acquired to turn Descartes’ own example against him. He objects that when closely investigated not even in this case are the Cartesian necessary conditions justified. In the glass sphere the two refractions should cancel each other (page 228), and even the Cartesian model of the propagation of light is misleading (page 229). It also fails to explain colour-production in thin plates: there are no non-parallel refractions and no termination of light and shadow. A new explanation of colour is called for: “There must be therefore some other propriety [property] of refraction that causes colour” (page 229).

- **Problem 10** Discuss Hooke’s explanation of the colours of thin films. What are the necessary conditions and the “principal cause”? Evaluate the similarities and differences between this account and the previous explanation about colours appearing after refraction. Use the last pages of Text 1. No hint is provided for this problem.

Hooke finishes the observation by stating that he his ideas “are capable of explicating all the Phenomena of colours, not onely of those appearing in the Prisme, Water-drop, or rainbow, and in laminated or plated bodies, but of all that are in the world, whether they be fluid or solid bodies, . . .” Apparent or emphatical colours thus become the basis for understanding all colour phenomena.

5 **Anti-Newtonian Modificationism**

With the death of Hooke, Newton’s most significant rival in England, and with the publication of Newton’s *Opticks* in 1704 and its successive editions
as well as its Latin and French translation, the question of light and colours seemed settled for many.

The modificationist tradition did not disappear with the emergence of the Newtonian theory, which remained heatedly debated, and writers working in a modificationist framework went on to challenge. When dealing with the problem of the elongation of the spectrum “these theories could separate the geometric from the chromatic problem” (Sepper 1988), while Newton’s theory connected the geometric and chromatic problems (the size/shape and colour of the iris), and proposed a solution where the law of sines, a major discovery of the seventeenth century could be saved in a modified form.

The eighteenth century offered many alternatives to the Newtonian system—but these mostly represented an undercurrent. Many disagreed with Newton on the question of the number of basic colours. Charles François de Cisternay DuFay (1698-1739), Edmé Gilles Guyot (1706-1786), and others, like Jacob Christoph Le Blon (1667-1741) the developer of the three-colour printing postulated the existence of three primaries. Lazarus Nuguet’s is a two-colour, boundary modificationist account, operating with light and darkness. Louis Bertrand Castel (1688-1757), developer of the “ocular harpsichord” rejected the account of refraction according to Newton and adopted a position very close to J. W. von Goethe’s, probably the most famous anti-newtonian modificationist.

5.1 Goethe’s Work on Colours

Johann Wolfgang von Goethe (1749-1832) worked on colours from the early 1790s to nearly his death in 1832. His gigantic undertaking occupied much time, and his knowledge in the history of the subject and in certain types of experiments was hardly rivalled in the period. A curious feature of his work is that, unlike many before or after him, it did not simply follow an individual line of research, but was mercilessly polemical in its approach.

The reception of the work has not always been positive, but claiming that scientists always treated his work with scorn is also false. In general, Goethe’s reception can be seen as complementary to that of the reception of Newton’s work (Schneider 1988). For a very long time, however, the general negative criticism had been taken for granted. In fact, Goethe’s work had received many praises together with the better known scolding. Just one example: Leopold von Henning taught the Goethean theory of colour
at the University of Berlin in the 1820s. Governmental support enabled him to buy optical instruments and was promoted to professorship (Jungnickel & McCormmach 1986).

One reason for the diverse reception is that Goethe’s contribution bears upon several traditions of thought and academic disciplines. Although his use of language is connected to the terminology of the physical optics of his age, he had little schooling in geometrical optics and had hardly any influence upon this area. His contributions to physiological optics is clearly recognised, but this does not imply that Goethe belonged to this tradition. Nevertheless his work on colour-blindness, on successive and simultaneous contrasts, sustained and mixed contrast, retinal afterimages, coloured shadows, etc. clearly secure him a position in the history of the subject. His work on the sensory-moral effect of colour is often mentioned in psychology textbooks. Among painters his influence was considerable, but the acceptance of his ideas was by no means unconditional. Few painters have followed closely his instruction. Among the contemporaries his impact was significant on Philip Otto Runge and J. M. W. Turner (Gage 1993).

Goethe himself lists several disciplines where he “endeavoured to serve”. The philosopher, the physician, the investigator of nature, the chemist, the mathematician, and the practical man, the dyer should all benefit from his work (Goethe 1970). In spite of this multitude of interests and methods employed, his work is very closely connected to the modificationist tradition. It is the mainly modificationist segment of his work that will be investigated here, a segment of paramount importance for understanding his thoughts on colours.

Rupprecht Matthei grouped Goethe’s work on colours in four major periods (Matthaei 1971). From these generally accepted periods, this chapter will only consider the second, the writing of the Farbenlehre, 1795-1810. I will notanalyse the writing and reception of the Beyträge zur Optik, 1791-5, his work on the entoptic phenomena, from 1810-1820 and his engagement in Purkinje’s work, 1820-1832, even though in these periods some significant changes took place in Goethe’s thinking.

5.2 The Theory of Colours

Goethe’s Theory of Colours (Farbenlehre) has three major parts: a didactic one, describing and grouping phenomena, a historical one, on the development
of theories of colour, and a polemical part attacking a number of Newton’s propositions in his *Opticks*.

In the historical part of his *Farbenlehre* Goethe tried to collect all those who purported the partisan views—in some books this collection serves as the basic study of the area even today (Crone 1999). Though Goethe started his fight as a solitary fighter—he soon found many who had shared his views.

The polemical part of the book—a bitter and often merciless critique of the first book of Newton’s *Opticks*—is one of the least studied works of Goethe. The attack is manifold: not only the methodological, epistemological aspects are criticised, but also some of the experimental results, the language used for transmitting the results, as well as the theory, and the structure.

The didactic part itself has three sections that deal with the physiological, the physical, and the chemical colours. The sequence is reminiscent of the *scala naturae*, from the most transient colours belonging to the eye itself (physiological colours), followed by increasingly less transient ones; catoptric, paroptic (diffraction), dioptric, and epoptic (interference) colours, ending in the fixed chemical colours. Surprisingly, Goethe does not start with light to understand what colour is. He finds it futile and irrelevant for his approach to explain what light is. The systematic study of effects (a purely phenomenological aim) suffices. Colours are “acts of light”, they are nature conforming to its own laws in relation to the sense of sight, and they are elemental natural phenomena. For their emergence light and the negation of light (darkness, shadow) is necessary.

The part about physiological colours establishes different polar opposites: white-black, red-green, blue-yellow, and studies the effect of light and darkness, of black and of white, and of the different colours on the eye. It discusses simultaneous contrasts (coloured shadows), and successive contrasts (afterimages), with a short study of subjective halos and an early account of the colour perception of colour-blinds.

5.3 The Dioptric Colours

Text 2 reproduces a segment of the discussion about physical colours, the so called dioptric colours. Goethe further subdivides this group into first and second class dioptric colours.
• **Problem 11** What distinguishes dioptric colours of the first and of the second class? List typical phenomena.

The text is not written for the specialist and the ample observational material testifies to Goethe’s emphasis on a purely phenomenal approach, but closer study shows that behind the seemingly endless list of phenomena there is a specific and peculiar philosophy of science, and a determined modificationist approach. There are oblique references to Newton (e.g. §176) and the covert polemics structures Goethe’s treatment of refraction.

• **Problem 12** What is the method Goethe uses to tackle a scientific problem in this text? What is the hierarchical structure of phenomena for Goethe? Describe the archetypal phenomenon.

When discussing dioptric colours of the second class, Goethe correctly points to a number of phenomena unexplained in the Newtonian theory, like why are the colours appearing at the borders symmetrical, why one-one or two-two colours are seen respectively, why the “thicknesses” of these coloured bands seem identical, why at some distance the number of colours decreases and not increases, etc. Today these phenomena have physiological (and often exceedingly complex) explanations—Goethe was correct in finding faults with the Newtonian theory (not all his findings were valuable, see (Lampert 2000)), but when investigated scrupulously his own theory does not fare better.

• **Problem 13** How is light modified in the case of primary and secondary dioptric colours? What is the connection of the two modes of modification?

Interestingly, even though staying in a modificationist framework, Goethe’s attitude shows a clear turn toward the physiological approach to colour which characterised the most fruitful approach from the second quarter of the nineteenth century in colour research. His earlier *Contributions to Optics* do not focus on the perceiver, his late work with Purkinje definitely focuses on the subject of experience.

• **Problem 14** What is the role of misperceptions and illusions for Goethe?
The evaluation of the Goethean scientific enterprise resembles a minefield even today, from utter rejection to the highest praise. The small section reproduced in Text 2 shows that it can be analysed as belonging to the little appreciated modificationist theories (neglecting this fact is one of the many reasons for the polarised reception). The text testifies to Goethe's commitment—and partial failure—to approach nature from a phenomenalistic point of view, by amassing experimental results, or, as a recent critic termed it, to carry out “exploratory experimentation”, see §228, and (Steinle 2001).

6 Texts

6.1 Text 1: Hooke’s *Micrographia*, Observation IX

- The rather long section is the reproduction of Observation IX in (Hooke 1665). Figures described in the observation appear on page 230.

**Observ. IX. Of the Colours observable in Muscovy Glass, and other thin Bodies.**

Muscovy-glass, or *Lapis specularis*, is a Body that seems to have as many Curiosities in its Fabrick as any common Mineral I have met with: for first, It is transparent to a great thickness: Next, it is compounded of an infinite number of thin flakes joyned or generated one upon another so close & smooth, as with many hundreds of them to make one smooth and thin Plate of a transparent flexible substance, which with care and diligence may be slit into pieces so exceedingly thin as to be hardly perceptible by the eye, and yet even those, which I have thought the thinnest, I have with a good *Microscope* found to be made up of many other Plates, yet thinner; and it is probable, that, were our *Microscopes* much
much better, we might much further discover its divisibility. Nor are these flakes only regular as to the smoothness of their Surfaces; but thirdly, In many Plates they may be perceived to be terminated naturally with edges of the figure of a Rhomboeid. This Figure is much more conspicuous in our English talk, much whereof is found in the Lead Mines, and is commonly called spar, and Kauck, which is of the same kind of substance with the Selenite, but is seldom found in so large flakes as that is, nor is it altogether so tough, but is much more clear and transparent, and much more curiously shaped, and yet may be cleft and flaked like the other Selenite. But fourthly, this stone has a property, which in respect of the Microscope, is more notable, and that is, that it exhibits several appearances of Colours, both to the naked Eye, but much more conpiciously to the Microscope; for the exhibiting of which, I took a piece of Moscovy-glass, and splitting or cleaving it into thin Plates, I found that up and down in several parts of them I could plainly perceive several white specks or flaws, and others diversly coloured with all the Colours of the Rainbow; and with the Microscope I could perceive, that these Colours were ranged in rings that encompassed the white speck or flaw, and were round or irregular, according to the shape of the spot which they terminated; and the position of Colours, in respect of one another, was the very same as in the Rainbow. The consecution of those Colours from the middle of the spot outward being Blew, Purple, Scarlet, Yellow, Green, Blew, Purple, Scarlet, and so onwards, sometimes half a score times repeated, that is, there appeared six, seven, eight, nine or ten several coloured rings or lines, each inclosing the other, in the same manner as I have often seen a very vivid Rainbow to have four or five several Rings of Colours, that is, accounting all the Gradations between Red and Blew for one: But the order of the Colours in these Rings was quite contrary to the primary or innermost Rainbow, and the same with those of the secondary or outermost Rainbow; these coloured Lines or Fringes, as I may so call them, were some of them much brighter than others, and some of them also very much broader, they being some of them ten, twenty, nay, I believe, near a hundred times broader than others; and those usually were broadish which were nearest the center or middle of the flaw. And oftentimes I found, that these Colours reach to the very middle of the flaw, and then there appeared in the middle a very large spot, for the most part, all of one colour, which was very vivid, and all the other Colours incompassing it, gradually ascending, and growing narrower towards the edges, keeping the same order, as in the secondary Rainbow, that is, if the middle were Blew, the next incompassing it would be a Purple, the third a Red, the fourth a Yellow, &c. as above; if the middle were a Red, the next without it would be a Yellow, the third a Green, the fourth a Blew, and so onward. And this order it always kept whatsoever were the middle Colour.

There was further observable in several other parts of this Body, many Lines or Threads, each of them of some one peculiar Colour, and those so exceedingly bright and vivid, that it afforded a very pleasant object through
through the Microscope. Some of these threads I have observed also to
be pieced or made up of several short lengths of differently coloured
ends (as I may so call them) as a line appearing about two inches long
through the Microscope, has been compounded of about half an inch of
a Peach colour, $\frac{1}{4}$ of a lovely Grass-green, $\frac{1}{2}$ of an inch more of a bright
Scarlet, and the rest of the line of a Watchet blew. Others of them were
much otherwise coloured; the variety being almost infinite. Another
thing which is very observable, is, that if you find any place where the
colours are very broad and conspicuous to the naked eye, you may, by
pressing that place with your finger, make the colours change places, and
go from one part to another.

There is one Phenomenon more, which may, if care be used, exhibit
to the beholder, as it has divers times to me, an exceeding pleasant,
and not less instructive Spectacle; and that is, if curiosity and diligence
be used, you may so split this admirable Substance, that you may have
pretty large Plates (in comparison of those smaller ones which you may
observe in the Rings) that are perhaps an $\frac{1}{4}$ or a $\frac{1}{2}$ part of an inch over,
each of them appearing through the Microscope most curiously, entirely,
and uniformly adorned with some one vivid colour: this, if examined
with the Microscope, may be plainly perceived to be in all parts of it
equally thick. Two, three, or more of these lying one upon another, exhibit
oftentimes curious compounded colours, which produce such a
Composition, as one would scarce imagine should be the result of such ingredients: As perhaps a faint yellow and a blue may produce a very deep
purple. But when anon we come to the more strict examination of these
Phenomena, and to inquire into the causes and reasons of these productions, we shall, I hope, make it more conceivable how they are produced,
and shew them to be no other than the natural and necessary effects arising from the peculiar union of concurrent causes.

These Phenomena being so various, and so truly admirable, it will certainly be very well worth our inquiry, to examine the causes and reasons
of them, and to consider, whether from these causes demonstratively evidenced, may not be deduced the real causes of the production of all
kind of Colours. And I the rather now do it, instead of an Appendix or Digression to this History, then upon the occasion of examining the Colours in Peacocks, or other Feathers, because this Subject, as it
doeth more variety of particular Colours, so doth it afford much better ways of examining each circumstance. And this will be made
manifest to him that considers, first, that this laminated body is more simple and regular than the parts of Peacocks feathers, this consisting only
of an indefinite number of plain and smooth Plates, heaped up, or in
cumbent on each other. Next, that the parts of this body are much more manageable, to be divided or joined, than the parts of a Peacocks fea-
ther, or any other substance that I know. And thirdly, because that in this
we are able from a colours body to produce several coloured bodies,
affording all the variety of Colours imaginable: And several others,
which the subsequent Inquiry will make manifest.
Micrographia.

To begin therefore, it is manifest from several circumstances, that the material cause of the apparition of these several Colours, is some Lamina or Plate of a transparent or pellucid body of a thickness very determinate and proportioned according to the greater or less refractive power of the pellucid body. And that this is so, abundance of Instancies and particular Circumstances will make manifest.

As first, if you take any small piece of the Muscovy-glass, and with a Needle, or some other convenient Instrument, cleave it oftentimes into thinner and thinner Lamina, you shall find, that till you come to a determinate thinness of them, they shall all appear transparent and colourless, but if you continue to split and divide them further, you shall find at last, that each Plate, after it comes to such a determinate thickness, shall appear most lovely ting'd or imbued with a determinate colour. If further, by any means you so flaw a pretty thick piece, that one part does begin to cleave a little from the other, and between those two there be by any means gotten some pellucid medium, those laminated pellucid bodies that fill that space, shall exhibit several Rainbows or coloured Lines, the colours of which will be disposed and ranged according to the various thicknesses of the several parts of that Plate. That this is so, is yet further confirmed by this Experiment.

Take two small pieces of ground and polish'd Looking-glass-plate, each about the bigness of a shilling, take these two dry, and with your fore-fingers and thumbs press them very hard and close together, and you shall find, that when they approach each other very near, there will appear several Iris or coloured Lines, in the same manner almost as in the Muscovy-glass; and you may very easily change any of the Colours of any part of the interposed body, by pressing the Plates closer and harder together, or leaving them more lax; that is, a part which appeared coloured with a red, may be presently ting'd with a yellow, blew, green, purple, or the like, by altering the approximation of the terminating Plates. Now that air is not necessary to be the interposed body, but that any other transparent fluid will do much the same, may be tried by working those approximated Surfaces with Water, or any other transparent Liquor, and proceeding with it in the same manner as you did with the Air; and you will find much the like effect, only with this difference, that those compress'd bodies, which differ most, in their refractive quality, from the compressing bodies, exhibit the most strong and vivid tinctures. Nor is it necessary, that this laminated and ting'd body should be of a fluid substance, any other substance, provided it be thin enough and transparent, doing the same thing: this the Lamina of our Muscovy-glass hint; but it may be confirmd by multitudes of other Instancies.

And first, we shall find, that even Glass itself may, by the help of a Lamp, be blown thin enough to produce these Phenomena of Colours: which Phenomena accidentally happening, as I have been attempting to frame small Glasses with a Lamp, did not a little surprize me at first, having never heard or seen any thing of it before; though afterwards comparing it with the Phenomena, I had often observed
observed in those Bubbles which Children use to make with Soap-water, I did the less wonder; especially when upon Experiment I found, I was able to produce the same Phenomena in thin Bubbles made with any other transparent Substance. Thus have I produced them with Bubbles of Pitch, Rosin, Colophony, Turpentine, Solutions of several Gums, as Gum-Arabick in water; any glutinous Liquor, as Wort, Wine, Spirit of Wine, Oyl of Turpentine, Glare of Snails, &c.

It would be needless to enumerate the several Instances, these being enough to shew the generality or universality of this propriety. Only I must not omit, that we have Instances also of this kind even in metalline Bodies and animal; for those several Colours which are observed to follow each other upon the polisht surface of hardned Steel, when it is by a sufficient degree of heat gradually tempered or softened, are produced from nothing else but a certain thin Lamina of a vitrified part of the Metal, which by that degree of heat, and the concurring action of the ambient Air, is driven out and fixed on the surface of the Steel.

And this hints to me a very probable (at least, if not the true) cause of the hardning and tempering of Steel, which has not, I think, been yet given, nor, that I know of been so much as thought of by any. And that is this, that the hardness of it arises from a greater proportion of a vitrified Substance interpersed through the pores of the Steel. And that the tempering or softening of it arises from the proportionate or smaller parcels of it left within those pores. This will seem the more probable, if we consider these Particulars.

First, That the pure parts of Metals are of themselves very flexible and tough; that is, will indure bending and hammering, and yet retain their continuity.

Next, That the Parts of all vitrified Substances, as all kinds of Glass, the Scoria of Metals, &c. are very hard, and also very brittle, being neither flexible nor malleable, but may by hammering or beating be broken into small parts or powders.

Thirdly, That all Metals (excepting Gold and Silver, which do not so much with the bare fire, unless assisted by other saline Bodies) do more or less vitrify by the strength of fire, that is, are corrode by a saline Substance, which I elsewhere shew to be the true cause of fire, and are thereby, as by several other Monstrous, converted into Scoria. And this is called, calcining of them, by Chimists. Thus Iron and Copper by heating and quenching do turn all of them by degrees into Scoria, which are evidently vitrified Substances, and unite with Glass, and are easily fusible; and when cold, very hard, and very brittle.

Fourthly, That most kind of Vitrifications or Calcinations are made by Salts, uniting and incorporating with the metalline Particles. Nor do I know any one calcination wherein a saline body may not, with very great probability, be said to be an agent or coadjuver.

Fifthly, That Iron is converted into Steel by means of the incorporation of certain Salts, with which it is kept a certain time in the fire.
Micrographia.

Sixthly, That any Iron may, in a very little time, be case hardned, as the Trades-men call it, by caging the iron to be hardned with clay, and putting between the clay and iron a good quantity of a mixture of Urine, Soot, Sea-salt, and Horse hoofs (all which contain great quantities of Saline bodies) and then putting the cace into a good strong fire, and keeping it in a considerable degree of heat for a good while, and afterwards heating, and quenching or cooling it suddenly in cold water.

Seventhly, That all kind of vitrify'd substances, by being suddenly cool'd, become very hard and brittle. And thence arises the pretty Phenomena of the Glas Drops, which I have already further explained in its own place.

Eighthly, That those metals which are not so apt to vitrifie, do not acquire any hardness by quenching in water, as Silver, Gold, &c.

These considerations premis'd, will, I suppose, make way for the more ease recepception of this following Explication of the Phenomena of hardned and temper'd Steel. That Steel is a substance made out of Iron, by means of a certain proportionate Vitrification of several parts, which are so curiously and proportionately mixt with the more tough and unalter'd parts of the Iron, that when by the great heat of the fire this vitrify'd substance is melted, and consequently rarify'd, and thereby the pores of the iron are more open, if then by means of dipping it in cold water it be suddenly cold, and the parts hardned, that is, stay'd in that same degree of Expansion they were in when hot, the parts become very hard and brittle, and that upon the same account almost as small parcels of glass quenched in water grow brittle, which we have already explicated. If after this the piece of Steel be held in some convenient heat, till by degrees certain colours appear upon the surface of the brightened metal, the very hard and brittle tone of the metal, by degrees relaxes and becomes much more tough and soft; namely, the action of the heat does by degrees loosen the parts of the Steel that were before streched or set asl as it were, and stay'd open by each other, whereby they become relaxed and set at liberty, whence some of the more brittle interjacent parts are thrust out and melted into a thin skin on the surface of the Steel, which from no colour increaseth to a deep Purple, and so onward by these gradations or consecutions, White, Yellow, Orange, Minium, Scarlet, Purple, Blue Watchet, &c. and the parts within are more conveniently, and proportionately mixt; and so they gradually subside into a texture which is much better proportion'd and closer joyn'd, whence that rigindnesse of parts ceases, and the parts begin to acquire their former ductilnesse.

Now, that 'tis nothing but the vitrify'd metal that sticks upon the surface of the colour'd body, is evident from this, that if by any means it be scrap'd and rubb'd off, the metal underneath it is white and clear; and if it be kept longer in the fire, so as to increase to a considerable thickness, it may, by blows, be beaten off in flakes. This is further confirm'd by this observables, that that Iron or Steel will keep longer from rupting which is covered with this vitrify'd cace: Thus also Lead will, by degrees, be
all turn'd into a litharge; for that colour which covers the top being
sum'd or shov'd aside, appears to be nothing else but a litharge or
vitrify'd Lead.

This is observable also in some sort, on Brass, Copper, Silver, Gold,
Tin, but is most conspicuous in Lead: all those Colours that cover the
surface of the Metal being nothing else, but a very thin vitrify'd part
of the heated Metal.

The other Instance we have, is in Animal bodies, as in Pearls, Mother
of Pearl-shells, Oyster-shells, and almost all other kinds of stony shells
whateover. This have I also sometimes with pleasure observ'd even
in Muscles and Tendons. Further, if you take any glutinous substance
and run it exceedingly thin upon the surface of a smooth glass or a pol-
lift metaline body, you shall find the like effects produced: and in
general, wherefoever you meet with a transparent body thin enough,
that is terminated by reflecting bodies of differing refraductions from it,
there will be a production of these pleasing and lovely colours.

Nor is it necessary, that the two terminating Bodies should be both of
the same kind, as may appear by the vitrified Lamina on Steel, Lead, and
other Metals, one surface of which Lamina is contiguous to the surface of
the Metal, the other to that of the Air.

Nor is it necessary, that these colour'd Lamina should be of an even
thickness, that is, should have their edges and middles of equal thickness,
as in a Looking-glass-plate, which circumstance is only requisite to make
the Plate appear all of the same colour; but they may resemble a Lens,
that is, have their middles thicker than their edges, or else a double con-
centric, that is, be thinner in the middle then at the edges; in both which
cases there will be various coloured rings or lines, with differing conceptions
or orders of Colours; the order of the first from the middle out-
wards being Red, Yellow, Green, Blew, &c. And the latter quite con-
trary.

But further, it is altogether necessary, that the Plate, in the places
where the Colours appear, should be of a determinate thickness: First, It
must not be more than such a thickness, for when the Plate is increased to
such a thickness, the Colours cease; and besides, I have seen in a thin
piece of Mussory-glass, where the two ends of two Plates, which appearing
both single, exhibited two distinct and differing Colours; but
in that place where they were united, and constituted one double Plate
(as I may call it) they appeared transparent and colourless. Nor, Seco-
dly, may the Plates be thinner then such a determinate size; for we
always find, that the very outmost Rim of these flaws is terminated in
a white and colourless Ring.

Further, in this Production of Colours there is no need of a determina-
tive Light of such a bigness and no more, nor of a determinate position
of that Light, that it should be on this side, and not on that side; nor of a
terminating shadow, as in the Prisme, and Rainbow, or Water-ball: for
we find, that the Light in the open Air, either in or out of the Sun-beams,
and within a Room, either from one or many Windows, produces much
the
Micrographia.

...the same effect: only where the Light is brightest, there the Colours are most vivid. So does the light of a Candle, collected by a Glass-ball. And further, it is all one whatever side of the coloured Rings be towards the light; for the whole Ring keeps its proper Colours from the middle outwards in the same order as I before related, without varying at all, upon changing the position of the light.

But above all it is most observable, that here are all kind of Colours generated in a pellucid body, where there is properly no such refraction as Des Cartes supposes his Globules to acquire a verticity by: For in the plain and even Plates it is manifest, that the second refraction (according to Des Cartes his Principles in the fifth Section of the eighth Chapter of his Meteors) does regulate and restore the supposed turbinate Globules unto their former uniform motion. This Experiment therefore will prove such a one as our thrice excellent Verulam calls Experimentum Crucis, serving as a Guide or Land-mark, by which to direct our course in the search after the true cause of Colours. Affording us this particular negative Information, that for the production of Colours there is not necessary either a great refraction, as in the Prisme; nor Secondly, a determination of Light and Shadow, such as is both in the Prisme and Glass-ball. Now that we may see like wise what affirmative and positive Instruction it yields, it will be necessary, to examine it a little more particularly and strictly, which that we may the better do, it will be requisite to premise somewhat in general concerning the nature of Light and Refraction.

And first for Light, it seems very manifest, that there is no luminous Body but has the parts of it in motion more or less.

First, That all kind of fiery burning Bodies have their parts in motion, I think, will be very easily granted. That the spark strik'd from a Flint and Steel is in a rapid agitation, I have elsewhere made probable. And that the Parts of rotten Wood, rotten Fitch, and the like, are also in motion, I think, will as easily be conceded by those, who consider, that those parts never begin to shine till the Bodies be in a state of putrefaction; and that is now generally granted by all, to be caused by the motion of the parts of putrifying bodies. That the Bononian stone shines no longer then it is either warmed by the Sun-beams, or by the flame of a Fire or of a Candle, is the general report of those that write of it, and of others that have seen it. And that heat argues a motion of the internal parts, is (as I laid before) generally granted.

But there is one Inference more, which was first shewn to the Royal Society by Mr. Clayton a worthy Member thereof, which does make this Affection more evident then all the rest: And that is, That a Diamond being rubb'd, strik'd, or heated in the dark, shines for a pretty while after, so long as that motion, which is imparted by any of those Agents, remains (in the same manner as a Glass, rubb'd, strik't, or by a means which I shall elsewhere mention) heated, yields a sound which lasts as long as the vibrating motion of that sonorous body: Several Experiments made on which Stone, are since published in a Discourse of Colours, by the truly honou-
honourable Mr. Boyle. What may be said of those ignes fatui that appear in the night, I cannot so well affirm, having never had the opportunity to examine them myself, nor to be informed by any others that had observed them: And the relations of them in Authors are so imperfect, that nothing can be built on them. But I hope I shall be able in another place to make it at least very probable, that there is even in those also a Motion which causes this effect. That the shining of sea-water proceeds from the same cause, may be argued from this; That it shines not till either it be beaten against a Rock, or be some other waves broken or agitated by Storms, or Oars, or other percussing bodies. And that the Animal Energies or Spiritual agit parts are very active in Cats eyes when they shine, seems evident enough, because their eyes never shine but when they look very intently either to find their prey, or being hunted in a dark room, when they seek after their adversary, or to find a way to escape. And the like may be said of the shining Bellies of Gloworms, since tis evident they can at pleasure either increase or extinguish that Radiation.

It would be somewhat too long a work for this place Zetetically to examine, and positively to prove, what particular kind of motion it is that must be the efficient of Light; for though it be a motion, yet tis not every motion that produces it, since we find there are many bodies very violently mov'd, which yet afford not such an effect; and there are other bodies, which to our other senses, seem not mov'd so much, which yet shine. Thus Water and quick-silver, and most other liquors heated, shine not; and several hard bodies, as Iron, Silver, Brass, Copper, Wood, c. though very often struck with a hammer, shine not presently, though they will all of them grow exceeding hot; whereas rotten Wood, rotten Fish, Sea water, Gloworms, c. have nothing of tangible heat in them, and yet (where there is no stronger light to affect the Sensory) they shine some of them so vividly, that one may make a shift to read by them.

It would be too long, I say, here to insert the discursive progress by which I inquir'd after the properties of the motion of Light, and therefore I shall only add the result.

And, First, I found it ought to be exceeding quick, such as those motions of fermentation and putrefaction, whereby, certainly, the parts are exceeding nimbly and violently mov'd; and that, because we find those motions are able more minutely to shatter and divide the body, then the most violent heats or menstruums we yet know. And that fire is nothing else but such a dissolution of the Burning body, made by the most universal menstruum of all sulphurous bodies, namely, the Air, we shall in another place of this Tractate endeavour to make probable. And that, in all extremely hot shining bodies, there is a very quick motion that causes Light, as well as a more robust that causes Heat, may be argued from the celerity wherewith the bodies are dissolv'd.

Next, it must be a vibrative motion. And for this the newly mention'd Diamond affords us a good argument; since if the motion of the parts did not
not return the Diamond must after many rubbings decay and be wafted: but we have no rea[n] reason to suspe[n]ct the latter, especially if we consider the exceeding difficulty that is found in cutting or wearing away a Di-
амond. And a Circular motion of the parts is much more improbable, since, if that were granted, and they be supposed irregular and Angular parts, I see not how the parts of the Diamond should hold so firmly to-
gether, or remain in the same sensible dimensions, which yet they do. Next, if they be Globular, and mov’d only with a turbinated motion, I
know not any cause that can impress that motion upon the pellucid me-
dium, which yet is done. Thirdly, any other irregular motion of the
parts one amongst another, must necessarily make the body of a fluid consistence, from which it is far enough. It must therefore be a Vibrat-
ing motion.

And Thirdly, That it is a very short vibrating motion, I think the in-
stances drawn from the shining of Diamonds will also make probable.
For a Diamond being the hardest body we yet know in the World, and
consequently the least apt to yield or bend, must consequently also have
its vibrations exceeding short.

And these, I think, are the three principal proprieties of a motion, re-
quisite to produce the effect call’d Light in the Object.

The next thing we are to consider, is the way or manner of the traject-
ion of this motion through the interpos’d pellucid body to the eye:
And here it will be easily granted,

First, That it must be a body susceptible and impervious of this motion
that will deserve the name of a Transparent. And next, that the parts of
such a body must be Homogeneous, or of the same kind. Thirdly, that the
constitution and motion of the parts must be such, that the appulse of
the luminous body may be communicated or propagated through it to the
greatest imaginable distance in the least imaginable time; though I see
no reason to affirm, that it must be in an instant: For I know not any one
Experiment or observation that does prove it. And, whereas it may be
objected, That we see the Sun ri[n] en at the very instant when it is above
the sensible Horizon, and that we see a Star hidden by the body of the
Moon at the same instant, when the Star, the Moon, and our Eye are all
in the same line: and the like Observations, or rather suppositions, may
be urg’d. I have this to answer, That I can as easily deny as they affirm;
for I would fain know by what means any one can be assured any more
of the Affirmative, then I of the Negative. If indeed the propagation
were very slow, ’tis possible something might be discovered by Eclipses of the Moon; but though we should grant the progress of the light
from the Earth to the Moon, and from the Moon back to the Earth, in
again to be full two Minutes in performing, I know not any possible
means to discover it: nay, there may be some instances perhaps of Ho-
izontal Eclipses that may seem very much to favour this supposition of
the slower progression of Light then most imagine. And the like may
be said of the Eclipses of the Sun, &c. But of this only by the by.

Fourthly, That the motion is propagated every way through an Homoge-


cous
...medium by direct or straight lines extended every way like rays from the center of a sphere. Filthily, in an Homogeneous medium this motion is propagated every way with equal velocity, whence necessarily every pulse or vibration of the luminous body will generate a sphere, which will continually increase, and grow bigger, just after the same manner (though indefinitely fitter) as the waves or rings on the surface of the water do swell into bigger and bigger circles about a point of it, where, by the sinking of a stone the motion was begun, whence it necessarily follows, that all the parts of those spheres undulated through an Homogeneous medium cut the rays at right angles.

But because all transparent mediums are not Homogeneous to one another, therefore we will next examine how this pulse or motion will be propagated through differing transparent mediums. And here, according to the most acute and excellent Philosopher Des Cartes, I suppose the sign of the angle of inclination in the first medium to be to the sign of refraction in the second. As the density of the first, to the density of the second. By density, I mean not the density in respect of gravity (with which the refractions or transparency of mediums hold no proportion) but in respect only to the trajectory of the Ray of light, in which respect they only differ in this, that the one propagates the pulse more easily and weakly, the other more slowly, but more strongly. But as for the pulses themselves, they will by the refraction acquire another propriety, which we shall now endeavour to explicate.

We will suppose therefore in the first Figure A C F D to be a physical Ray, or A B C and D E F to be two Mathematical Rays, trajected from a very remote point of a luminous body through an Homogeneous transparent medium L L L, and D A, E B, F C, to be small portions of the orbicular impulses which must therefore cut the ray at right angles; these rays meeting with the plain surface N O of a medium that yields an easier trajectory to the propagation of light, and falling obliquely on it, they will in the medium M M M be refracted towards the perpendicular of the surface. And because this medium is more easily trajected then the former by a third, therefore the point C of the orbicular pulse F C will be mov'd to H four spaces in the same time that F the other end of it is mov'd to G three spaces, therefore the whole refracted pulse G H shall be oblique to the refracted rays C H K and G L; and the angle G H C shall be acute, and so much the more acute by how much the greater the refraction be, then which nothing is more evident, for the sign of the inclination is to be the sign of refraction as G F to T C the distance between the point C and the perpendicular from G on C K, which being as four to three, H C being longer then G F is longer also then T C, therefore the angle G H C is less than G T C. So that henceforth the parts of the pulses G H and I K are mov'd as in, or cut the rays at oblique angles.

It is not my business in this place to set down the reasons why this or that body should impede the rays more, others less: as why water should transmit the rays more easily, though more weakly than air. Onely thus much
much in general I shall hint, that I suppose the medium MMM to have less of the transparent undulating subtile matter, and that matter to be less implicated by it, whereas LLL I suppose to contain a greater quantity of the fluid undulating substance, and this to be more implicated with the particles of that medium.

But to proceed, the same kind of obliquity of the Pulses and Rays will happen also when the refraction is made out of a more eafe into a more difficult mediūm; as by the calculations of G Q & C S R which are refracted from the perpendicular. In both which calculations 'tis obvious to observe, that always that part of the Ray towards which the refraction is made has the end of the orbicular pulse precedent to that of the other side. And always, the oftener the refraction is made the same way; or the greater the single refraction is, the more is this unequal progress. So that having found this odd propriety to be an inseparable concomitant of a refracted Ray, not frightened by a contrary refraction, we will next examine the refractions of the Sun-beams, as they are suffer'd only to pass through a small passage, obliquely out of a more difficult into a more eafe medium.

Let us suppose therefore ABC in the second Figure to represent a large Chymical Glass-body about two foot long, filled with very fair Water as high as AB, and inclin'd in a convenient posture with B towards the Sun: Let us further suppose the top of it to be cover'd with an opaque body, all but the hole c b, through which the Sun-beams are suffer'd to pass into the Water, and are thereby refracted to e d e f, against which part, if a Paper be expanded on the outside, there will appear all the colours of the Rain-bow, that is, there will be generated the two principal colours, Scarlet and blue, and all the intermediate ones which arise from the composition and dilutions of these two, that is, c d shall exhibit a Scarlet, which toward d is diluted into a yellow; this is the refraction of the Ray, i k, which comes from the under side of the Sun; and the Ray e f shall appear of a deep blue, which is gradually towards e dilute into a pale Watchet-blue. Between d and e the two diluted colours, Blue and Yellow are mixt and compounded into a Green; and this I imagine to be the reason why Green is so acceptable a colour to the eye, and that either of the two extremes are, if intense, rather a little offensive, namely, the being plac'd in the middle between the two extremes, and compounded out of both thofe, diluted also, or somewhat qualitied, for the composition, arising from the mixture of the two extremes undiluted, makes a Purple, which though it be a lovely colour, and pretty acceptable to the eye, yet is nothing comparable to the ravishing pleasure with which a curious and well tempered Green affects the eye. If removing the Paper, the eye be plac'd against c d, it will perceive the lower side of the Sun (or a Candle at night which is much better, because it offends not the eye, and is more easily manageable) to be of a deep Red, and if against e f it will perceive the upper part of the luminous body to be of a deep Blue; and these colours will appear deeper and deeper, according as the Rays from the luminous body fall more obliquely on the surface of the Water, and thereby suffer a greater refraction, and the
Micrographia.

more distinct, the further c d e f is removed from the trajectory hole.

So that upon the whole, we shall find that the reason of the phenomena seems to depend upon the obliquity of the orbicular pulse, to the Lines of Radiation and in particular, that the Ray c d which constitutes the scarlet has its inner parts, namely those which are next to the middle of the luminous body, precedent to the outermost which are contiguous to the dark and unirradiating skie. And that the Ray e f which gives a blue, has its outward part, namely, that which is contiguous to the dark skie precedent to the pulse from the innermost, which borders on the bright area of the luminous body.

We may observe further, that the cause of the dilating of the colours towards the middle, proceeds partly from the wideness of the hole through which the Rays pass, whereby the Rays from several parts of the luminous body, fall upon many of the same parts between c and f as is more manifest by the Figure: And partly also from the nature of the refraction it self, for the vividness or strength of the two terminating colours, arising chiefly as we have seen, from the very great difference that is betwixt the outsides of those oblique undulations & the dark Rays circumambient, and that disparity betwixt the approximate Rays, decaying gradually: the further inward toward the middle of the luminous body they are remov'd, the more must the colour approach to a white or an undisturbed light.

Upon the calculation of the refraction and reflection from a Ball of Water or Glass, we have much the same phenomena, namely, an obliquity of the undulation in the same manner as we have found it here. Which, because it is very much to our present purpose, and affords such an instancia crucis, as no one that I know has hitherto taken notice of, I shall further examine. For it does very plainly and positively distinguish, and shew, which of the two Hypotheses, either the Carcian or this is to be followed, by affording a generation of all the colors in the Rainbow, where according to the Carcian Principia there should be none at all generated. And secondly, by affording an instance that does more closely confine the cause of these phenomena of colours to this present Hypothesis.

And first, for the Carcian, we have this to object against it. That whereas he says (Meteorum Cap.s. Sec. 5.), sed indicabam unicum refractioni sit selectis ad minimis requiris, & quidem talem ejus effectum aliud contrario (refractione) non deflratur: Nam experientia docet, si superficies NM & NP (nempe refringentes) parallela front, radios tantundem per alteram iterum crebris quantum per unam frangerentur, pullos colores depicuros; This Principle of his holds true indeed in a prime where the refracting surfaces are plain, but is contradicted by the Ball or Cylinder, whether of Water or Glass, where the refracting surfaces are Orbicular or Cylindrical. For if we examine the passage of any Globule or Ray of the primary Iris, we shall find it to pass out of the Ball or Cylinder again, with the same inclination and refraction that it entered in withall, and that last refraction by means of the intermediate reflection shall be the same as if without any reflection at all the Ray had been twice refracted by two Parallel surfaces.
And that this is true, not only in one, but in every Ray that goes to
the constitution of the Primary Iris; nay, in every Ray, that suffers only
two refractions, and one reflection, by the surface of the round body, we
shall presently see most evident, if we repeat the Cartesian Scheme, men-
tioned in the tenth Section of the eighth Chapter of his Meteors, where
EFKNP in the third Figure is one of the Rays of the Primary Iris,
twice refracted at F and N, and once reflected at K by the surface of the
Water-ball. For, first it is evident, that KF and KN are equal, because
KN being the reflected part of KF they have both the same inclination
on the surface K that is the angles FKT, and NVK made by the two
Rays and the Tangent of K are equal, which is evident by the Laws of re-
fraction; whence it will follow also, that KN has the same inclination on
the surface N, or the Tangent of it at N that the Ray KF has to the sur-
face F, or the Tangent of it at F, whence it must necessarily follow, that
the refractions at F and N are equal, that is, KFE and KNP are equal.
Now, that the surface N is by the reflection at K made parallel to the sur-
fare at F, is evident from the principles of reflection; for reflection being
nothing but an inverting of the Rays, if we re-invert the Ray KNP, and
make the same inclinations below the line TKV that it has above, it will
be most evident, that KH the inverse of KN will be the continuation of
the line FK, and that LHI the inverse of OX is parallel to FY. And
HM the inverse of NP is Parallel to EF for the angle KHI is equal
to KNO which is equal to KFY, and the angle KH M is equal to KN P
which is equal to KFE which was to be proved.

So that according to the above mentioned Cartesian principles there
should be generated no colour at all in a Ball of Water or Glass by two
refractions and one reflection, which does hold most true indeed, if the
surfaces be plain, as may be experimented with any kind of prism where
the two refracting surfaces are equally inclin’d to the reflecting; but in
this the Phanomena are quite otherwise.

The cause therefore of the generation of colour must not be what Des
Cartes assigns, namely, a certain rotation of the Globuli atheraei, which are
the particles which he supposes to constitute the Pellucid medium. But
somewhat else, perhaps what we have lately supposed, and shall by and
by further prosecute and explain.

But, if I shall crave leave to propound some other difficulties of his,
notwithstanding exceedingly ingenious Hypothesis, which I plainly confess
to me seem such; and those are,

First, if that light be (as is affirmed, Dioptr. cap. 1. §. 8.) not so pro-
perly a motion, as an action or propulsion to motion, I cannot conceive
how the eye can come to be sensible of the verticity of a Globule, which
is generated in a drop of Rain, perhaps a mile off from it. For that Globule
is not carry’d to the eye according to his formerly recited Principles; and if
not so, I cannot conceive how it can communicate its rotation, or circular
motion to the line of the Globules between the drop and the eye. It cannot
be by means of every one turning the next before him; for if so, then
only all the Globules that are in the odd places must be turned the same
way
way with the first, namely, the 3. 5. 7. 9. 11. &c. but all the Globules interposed between them in the even places; namely, the 2. 4. 6. 8. 10. &c. must be the quite contrary; whence, according to the Cartesian Hypothesis, there must be no distinct colour generated, but a confusion. Next, since the Cartesian Globuli are suppos'd (Principiorum Philosoph. Part. 3. §. 86.) to be each of them continually in motion about their centers, I cannot conceive how the eye is able to distinguish this new generated motion from their former inherent one, if I may so call that other wherewith they are mov'd or turbinated, from some other cause than refraction. And thirdly, I cannot conceive how these motions should not happen sometimes to oppose each other, and then, in stead of a rotation, there would be nothing but a direct motion generated, and consequentely no colour. And fourthly, I cannot conceive how the Cartesian Hypothesis it is possible to give any plausible reason of the nature of the Colours generated in the thin laminae of these our Microscopical Observations; for in many of these, the refracting and reflecting surfaces are parallel to each other, and consequentely no rotation can be generated, nor is there any necessity of a shadow or termination of the bright Rays, such as is suppos'd (Chap. 8. §. 5. Et praeceper observavi umbrae quoque, aut limitationem luminis requiri: and Chap. 8. §. 9.) to be necessary to the generation of any distinct colours; Besides that, here is oftentimes one colour generated without any of the other appendant ones, which cannot be by the Cartesian Hypothesis.

There must be therefore some other propriety of refraction that causes colour. And upon the examination of the thing, I cannot conceive any one more general, inexpressible, and sufficient, than that which I have before affir'd. That we may therefore see how exactly our Hypothesis agrees also with the Phenomena of the refracting round body, whether Globe or Cylinder, we shall next subjoin our Calculation or Examen of it.

And to this end, we shall calculate any two Rays: as for instance; let E F be a Ray cutting the Radius C D (divided into 20. parts) in G 16. Fig. 3.

parts distant from C, and e f another Ray, which cuts the same Radius in g 17. parts distant, these will be refracted to K and k, and from thence reflected to N and n, and from thence refracted toward P and p, therefore the Arch F f will be 5. 4. 5'. The Arch F K 106. 4. 30', the Arch f k 101. 4. 2'. The line F G 6000. and f g 5267, therefore b f 733, therefore F e 980, almost. The line F K 16024. and f k 15436. therefore N d 196: and n o 147 almost, the line N n 1019 the Arch N n 5. 4. 51', therefore the Angle N n o is 34. 4. 43', therefore the Angle N o n is 139. 4. 56', which is almost 90. 4. more than a right Angle.

It is evident therefore by this Hypothesis, that at the same time that e f touches f, E F is arrived at c. And by that time e f k n is got to n, E F K N is got to d, and when it touches N, the pulle of the other Ray is got to a, and no farther, which is very short of the place it should have arriv'd to, to make the Ray n p to cut the orbicular pulse. N o at right Angles; therefore the Angle N o p is an acute Angle, but the quite contrary
M i c r o g r a p h i a.

trary of this will happen, if 17. and 18. be calculated in stead of 16. and 17. both which does most exactly agree with the Phenomena: For if the Sun, or a Candle (which is better) be placed about E e, and the eye about P p, the Rays E F e f. at 16. and 17. will paint the side of the luminous object toward n p Blue, and towards N P Red. But the quite contrary will happen when E F is 17. and e f 18. for then towards N P shall be a Blue, and towards n p a Red, exactly according to the calculation. And there appears the Blue of the Rainbow, where the two Blue sides of the two Images unite, and there the Red where the two Red sides unite, that is, where the two Images are just disappearing; which is, when the Rays E F and N P produc'd till they meet, make an Angle of about 41. and an half; the like union is there of the two Images in the Production of the Secondary Iris, and the same causes, as upon calculation may appear; only with this difference, that it is somewhat more faint, by reason of the duplicate reflection, which does always weaken the impulse the oftner it is repeated.

Now, though the second refraction made at N n be convenient, that is, do make the Rays glance the more, yet is it not altogether requisite; for it is plain from the calculation, that the pulse d n is sufficiently oblique to the Rays K N and k n, as well as the pulse f e is oblique to the Rays F K e f k. And therefore if a piece of very fine Paper be held close against N n and the eye look on it either through the Ball as from D, or from the other side, as from B. there shall appear a Rainbow, or colour'd line painted on it with the part toward X appearing Red, towards O, Blue; the same also shall happen, if the Paper be placed about K k, for towards T shall appear a Red, and towards V a Blue, which does exactly agree with this my Hypothesis, as upon the calculation of the progress of the pulse will most easily appear.

Nor do these two observations of the colours appearing to the eye about p differing from what they appear on the Paper at N contradict each other; but rather confirm and exactly agree with one another, as will be evident to him that examines the reasons set down by the ingenious Des Cartes in the 12. Sed. of the 8. Chapter of his Meteors, where he gives the true reason why the colours appear of a quite contrary order to the eye, to what they appear'd on the Paper if the eye be placed instead of the Paper: And as in the Prisme, so also in the Water, Drop, or Globe the Phenomena and reason are much the same.

Having therefore shewn that there is such a propriety in the prisme and water Globule whereby the pulse is made oblique to the progressive, and that so much the more, by how much greater the refraction is, I shall in the next place consider, how this conduces to the production of colours, and what kind of impression it makes upon the bottom of the eye; and to this end it will be requisite to examine this Hypothesis a little more particularly.

First therefore, if we consider the manner of the progress of the pulse, it will seem rational to conclude, that that part or end of the pulse which precedes the other, must necessarily be somewhat more obtunded, or impeded by
by the resistance of the transparent medium, than the other part or end of it which is subseuent, whose way is, as it were, prepared by the other; especially if the adjacent medium be not in the same manner enlightened or agitated. And therefore (in the fourth Figure of the sixth Iconium) the Ray A A A H B will have its side H H more deadned by the resistance of the dark or quiet medium P P P, Whence there will be a kind of deadness superinduced on the side H H H, which will continually increase from B, and strike deeper and deeper into the Ray by the line B R; Whence all the parts of the triangle, R B H O will be of a dead blue colour, and so much the deeper, by how much the nearer they lie to the line B H H, which is most deadned or impeded, and so much the more dilute, by how much the nearer it approaches the line B R. Next on the other side of the Ray A A N, the end A of the pulse A H will be promoted, or made stronger, having its passage already prepar'd as 'twere by the other parts preceding, and so its impression will be stronger; And because of its obliquity to the Ray, there will be propagated a kind of faint motion into Q Q, the adjacent dark or quiet medium, which faint motion will spread further and further into Q Q, as the Ray is propagated further and further from A, namely, as far as the line M A, whence all the triangle M A N will be ting'd with a red, and that red will be the deeper the nearer it approaches the line M A, and the paler or yellower the nearer it is the line N A. And if the Ray be continued, so that the lines A N and B R (which are the bounds of the red and blue dilute) do meet and cross each other, there will be beyond that intercession generated all kinds of greens.

Now, these being the proprieties of every single refracted Ray of light, it will be easy enough to consider what must be the result of very many such Rays collateral: As if we suppose infinite such Rays interjacint between A K S B and A N O B, which are the terminating: For in this case the Ray A K S B will have its red triangle entire, as lying next to the dark or quiet medium, but the other side of it B S will have no blue, because the medium adjacent to it S B O, is mov'd or enlightened, and consequently that light does destroy the colour. So likewise will the Ray A N O B lose its red, because the adjacent medium is mov'd or enlightened, but the other side of the Ray that is adjacent to the dark, namely, A H O will preserve its blue entire, and these Rays must be so far produc'd as till A N and B R cut each other, before there will be any green produc'd. From these Propositions well consider'd, may be deduc'd the reasons of all the Phenomena of the prism, and of the Globules or drops of Water which conduce to the production of the Rainbow.

Next for the impression they make on the Retina, we will further examine this Hypothesis: Suppose therefore A B C D E F, in the fifth Figure, to represent the Ball of the eye: on the Cornea of which A B C two Rays G A C H and K C A I (which are the terminating Rays of a luminous body) falling, are by the refraction thereof collected or converg'd into two points at the bottom of the eye. Now, because these terminating Rays, and all the intermediate ones which come from any part of the luminous body, are suppos'd by some sufficient refraction before they enter
Micrographia.

enter the eye, to have their pulses made oblique to their progression, and consequently each Ray to have potentially superinduced two proprieties, or colours, viz. a Red on the one side, and a Blue on the other, which notwithstanding are never actually manifest, but when this or that Ray has the one or the other side of it bordering on a dark or unmoved medium, therefore as soon as these Rays are entered into the eye, and so have one side of each of them bordering on a dark part of the humours of the eye, they will each of them actually exhibit some colour; therefore ADC the production GACH will exhibit a Blue, because the side CD is adjacent to the dark medium CQDC, but nothing of a Red, because its side AD is adjacent to the enlightened medium ADF A: And all the Rays that from the points of the luminous body are collected on the parts of the Retina between D and F shall have their Blue so much the more diluted by how much the farther these points of collection are distant from D towards F, and the Ray AFC the production of KCAI will exhibit a Red, because the side AF is adjacent to the dark or quiet medium of the eye AFA, but nothing of a Blue, because its side CF is adjacent to the enlightened medium CFDC, and all the Rays from the intermediate parts of the luminous body that are collected between F and D shall have their Red so much the more diluted, by how much the farther they are distant from F towards D.

Now, because by the refraction in the Cornea, and some other parts of the eye, the sides of each Ray, which before were almost parallel, are made to converge and meet in a point at the bottom of the eye, therefore that side of the pulse which preceded before these refractions, shall first touch the Retina, and the other side last. And therefore according as this or that side, or end of the pulse shall be impeded, accordingly will the impressions on the Retina be varied; therefore by the Ray GACH refracted by the Cornea to D there shall be on that point a stroke or impression confus'd, whose weakest end, namely, that by the line CD shall precede, and the stronger, namely, that by the line AD shall follow. And by the Ray KCAI refracted to F, there shall be on that part a confus'd stroke or impression, whose strongest part, namely, that by the line CF shall precede, and whose weakest or impeded, namely, that by the line AF shall follow, and all the intermediate points between F and D will receive impression from the converging Rays so much the more like the impressions on F and D by how much the nearer they approach that or this.

From the consideration of the proprieties of which impressions, we may collect these short definitions of Colours: That Blue is an impression on the Retina of an oblique and confus'd pulse of light, whose weakest part precedes, and whose strongest follows. And, that Red is an impression on the Retina of an oblique and confus'd pulse of light, whose strongest part precedes, and whose weakest follows.

Which proprieties, as they have been already manifested, in the Prisme and falling drops of Rain, to be the causes of the colours there generated, may be easily found to be the efficiencies also of the colours appearing in thin laminated transparent bodies; for the explication of which, all this has been premised.
And that this is so, a little closer examination of the phenomena and the figure of the body, by this hypothesis, will make evident.

For first (as we have already observed) the laminated body must be of a determinate thickness, that is, it must not be thinner than such a determinate quantity; for I have always observ’d, that near the edges of those which are exceeding thin, the colours disappear, and the part grows white; nor must it be thicker than another determinate quantity; for I have likewise observ’d, that beyond such a thickness, no colours appear’d, but the Plate looked white, between which two determinate thicknesses were all the colour’d Rings; of which in some substances I have found ten or twelve, in others not half so many, which I suppose depends much upon the transparency of the laminated body. Thus though the confections are the same in the scum or the skin on the top of metals; yet in those confections the same colour is not so often repeated as in the confections in thin Glass, or in Sopewater, or any other more transparent and glutinous liquor; for in these I have observ’d, Red, Yellow, Green, Blue, Purple; Red, Yellow, Green, Blue, Purple; Red, Yellow, Green, Blue, Purple; Red, Yellow, &c. to succeed each other ten or twelve times, but in the other more opaceous bodies the confections will not be half so many.

And therefore secondly, the laminated body must be transparent, and this I argue from this, that I have not been able to produce any colour at all with an opaceous body, though never so thin. And this I have often try’d, by pressing a small Globule of Mercury between two smooth Plates of Glass, whereby I have reduc’d that body to a much greater thinness, then was requisite to exhibit the colours with a transparent body.

Thirdly, there must be a considerable reflecting body adjacent to the under or further side of the lamina or plate: for this I always found, that the greater that reflection was, the more vivid were the appearing colours.

From which Observations, it is most evident, that the reflection from the under or further side of the body is the principal cause of the production of these colours; which, that it is so, and how it conduces to that effect, I shall further explain in the following Figure, which is here described of a very great thickness, as if it had been view’d through the Microscope; and this indeed much thicker than any Microscope (I have yet us’d) has been able to shew me those colour’d plates of Glass, or Muscovy-glass, which I have not without much trouble view’d with it; for though I have endeavoured to magnifie them as much as the Glasses were capable of, yet are they so exceeding thin, that I have not hitherto been able positively to determine their thickness. This Figure therefore I here represent, is wholly Hypothetical.

Let ABCDEFE in the sixth Figure be a frustum of Muscovy-glass, thinner toward the end AE, and thicker towards DF. Let us first suppose the Ray abbd coming from the Sun, or some remote luminous object to fall obliquely on the thinner plate BAEE, part therefore is reflected back, by egbd, the first superiory, whereby the perpendicular pulse.
pulse $ab$ is after reflection propagated by $cd, cd$, equally remote from each other with $ab, ab$, so that $ag + gc$, or $bb + bd$ are either of them equal to $aa$, as is also $ce$, but the body $BAE$ being transparent, a part of the light of this ray is refracted in the surface $AB$, and propagated by $gikb$ to the surface $EF$, whence it is reflected and refracted again by the surface $AB$. So that after two refractions and one reflection, there is propagated a kind of fainter ray $emnf$, whose pulse is not only weaker by reason of the two refractions in the surface $AB$, but by reason of the time spent in passing and repassing between the two surfaces $AB$ and $EF$, $ef$ which is this fainter or weaker pulse comes behind the pulse $cd$, so that hereby (the surfaces $AB$, and $EF$ being so near together, that the eye cannot discriminate them from one) this confus'd or duplicated pulse, whose strongest part precedes, and whose weakest follows, does produce on the Retina (or the optic nerve that covers the bottom of the eye) the sensation of a Yellow.

And secondly, this Yellow will appear no much the deeper, by how much the further back towards the middle between $cd$ and $cd$ the furious pulse $ef$ is remov'd, as in 2 where the surface $BC$ being further remov'd from $EF$, the weaker pulse $ef$ will be nearer to the middle, and will make an impression on the eye of a Red.

But thirdly, if the two reflecting surfaces be yet further remov'd aunder (as in 3 $CD$ and $EF$ are) then will the weaker pulse be so far behind, that it will be more than half the distance between $cd$ and $cd$.

And in this case it will rather seem to precede the following stronger pulse, then to follow the preceding one, and consequently a Blue will be generated. And when the weaker pulse is just in the middle between two strong ones, then is a deep and lovely Purple generated; but when the weaker pulse $ef$ is very near to $cd$, then is there generated a Green, which will be bluer, or yellower, according as the approximate weak pulse does precede or follow the stronger.

Now fourthly, if the thicker plate chance to be cleat into two thinner plates, as $CDFE$ is divided into two plates by the surface $GH$ then from the composition arising from the three reflections in the surfaces $CD$, $GH$, and $EF$, there will be generated several compounded or mixt colours, which will be very differing, according as the proportion between the thicknesses of those two divided plates $CD, GC$, and $GH, FE$ are varied.

And fifthly, if these surfaces $CD$ and $FE$ are further remov'd aunder, the weaker pulse will yet lagg behind much further, and not onely be coincident with the second, $cd$, but lagg behind that also, and that so much the more, by how much the thicker the plate be, so that by degrees it will be coincident with the third $cd$ backward also, and by degrees, as the plate grows thicker with a fourth, and so onward to a fifth, sixth, seventh, or eighth; so that if there be a thin transparent body, that from the greatest thinnes requisite to produce colours, does, in the manner of a Wedge, by degrees grow to the greatest thickness that a plate can be of, to exhibit a colour by the reflexion of Light from such a body, there shall
Micrographia.

shall be generated several consecutions of colours, whose order from the thin end towards the thick, shall be Yellow, Red, Purple, Blue, Green; Yellow, Red, Purple, Blue, Green; Yellow, &c., and these to often repeated, as the weaker pulse does lose pace with its primary, or first pulse, and is coincident with a second, third, fourth, fifth, sixth, &c., pulse behind the first. And this, as it is coincident, or follows from the first Hypothesis, I took of colours, so upon experiment have I found it in multitudes of instances that seem to prove it. One thing which seems of the greatest concern in this Hypothesis, is to determine the greatest or least thickness requisite for these effects, which, though I have not been wanting in attempting, yet so exceeding thin are these coloured Plates, and so imperfect our Microscope, that I have not been hitherto successful, though if my endeavours shall answer my expectations, I shall hope to gratify the curious Reader with some things more removed beyond our reach hitherto.

Thus have I, with as much brevity as I was able, endeavoured to explain (hypothetically at least) the causes of the phenomena I formerly recited, on the consideration of which I have been the more particular.

First, because I think these I have newly given are capable of explaining all the phenomena of colours, not only of those appearing in the Prism, Water-drop, or Rainbow, and in laminated or plated bodies, but of all that are in the world, whether they be fluid or solid bodies, whether in thick or thin, whether transparent, or seemingly opaque, as I shall in the next Observation further endeavour to shew. And secondly, because this being one of the two ornaments of all bodies discoverable by the sight, whether looked on with, or without a Microscope, it seem'd to deserve (somewhere in this Treatise, which contains a description of the Figure and Colour of some minute bodies) to be somewhat the more intimately enquired into.
• **Hint to Problem 3** Physical and structural properties of mica are discussed (composite, transparent substance that exhibits colours). The plates can be split into thinner plates, but even these thin plates that seem simple appear composite under the microscope. Optical instruments thus extend our senses, but the properties observable on the micro level are not fundamentally different from the properties in the macroscopic level: “were our Microscopes much better, we might much further discover its divisibility”. Such an attitude implies a certain homogeneity of nature on different scales, a view also shared by Newton (see page 279).

The discussion quickly focuses on the peculiar colours seen. This description was the first public account of these interference colours. Around the same time, this, and a number of other recent discoveries seemed as new instances of apparent colours (*colores emphatici*), of which the rainbow was the prime example. While earlier these apparent colours were rather the exceptions (like the in the tail of the peacock or the neck of the dove), in the seventeenth century most theories used them to find general explanations concerning all colour phenomena.

• **Hint to Problem 4** The colours seen in rainbows are present in interference patterns like the ones in the thin sheets of mica. “Purple”, however, an extraprismatic colour also appears. See also Goethe’s text §215. It is unclear what Hooke means by “a very vivid Rainbow, … four or five several Rings of Colours”. It is unlikely that this reference is an early observation of supernumerary arcs (see page 172). In case of the concentric colours the central bands are broader, while in the coloured ‘threads’ other extraprismatic colours are listed (‘Peach’, ‘Watchet’). The uniformly coloured plates, according to Hooke, are composite colours possibly from the mixture of yellow and blue, but the details are not discussed.

• **Hint to Problem 5** Hooke is looking for the “true causes” (*verae causae*) of colours. The colours of the thin plates are simple and regular, the parts manageable, and several coloured bodies are produced from colourless ones. Implicit is the assumption (already present in Descartes) that all colour appearances are explainable by finding the causes for the emergence of (some) apparent colours.
The decade of Hooke’s discovery had shown a number of new instances of apparent colours: Grimaldi’s first observation of diffraction in 1665 and Erasmus Bartholin’s 1669 discovery of birefringence or double refraction of Icelandic spar. The different phenomena gave rise to a number of theories: Hooke basing his on the colours of thin plates, Newton on dispersion, Huygens on double refraction, etc., all aiming to become universal theories.

- **Hint to Problem 6** A general and sufficient condition is “a transparent body thin enough, that is terminated by reflecting bodies of differing refractions from it”. The material is not relevant, but the thickness has to be within a given range: “it must not be more than such a thickness . . . nor, Secondly, may the Plates be thinner then such a determinate cize”. Various intensities and directions of the light suffice to give rise to the colours, and, contrary to the Cartesian theory no boundary of shadow and light is needed: “no need of . . . a terminating shadow, as in the Prisme, and Rainbow, or Water-ball”.

Hooke uses the newly discovered phenomenon to test the Cartesian theory. He mixes two Baconian terms, the experimentum lucifera and the instantia crucis and coins the pseudo-Baconian term experimentum crucis. This is used to provide negative information and disproves some of the conditions that Descartes thought to be necessary for the production of colour: “there is not necessary either a great refraction as in the Prisme; nor Secondly a determination of Light and Shadow.” The earlier described conditions therefore test (and disprove) an important theory of the time. Negative information (disproof) is achieved seemingly easily. For positive information, however, further study and a general theory of Light and Refraction is needed.

Interestingly, the possibility that the formation of the rainbow and the colours of the thin plates are the results of different processes and hence require different conditions are not investigated.

- **Hint to Problem 8** Hooke makes an interesting (but not uncommon) distinction between physical and mathematical rays. For the mathematical rays the rule of refraction holds true (today also known as the Snell-Descartes rule): “the sign of the angle of inclination in
the first medium to be to the sign of refraction in the second, As the
density of the first to the density of the second.” This seemingly run-
on phrasing set up the ratio of the sines as constant for any two given
media.

His example referring to figure I reproduced on page 230 states
the following relationship between angle of incidence (i) and of re-
fraction (r)—note that this is an assumption based on the Cartesian
theory:

$$\frac{\sin i}{\sin r} = \frac{v_r}{v_i} = n = \frac{4}{3}$$  \hspace{1cm} (VIII.1)

As Sabra notes (Sabra 1967): “Accordingly, with C as centre and a
radius equal to $3/4$ FG, he describes the smaller arc in the figure. He
then draws a tangent to this arc from G; the line joining C and the
point of tangency T gives the direction of refraction. For, FCG being
equal to the angle of incidence and CGT being equal to the angle of
refraction, we get:

$$\frac{FG}{CT} = \frac{\sin i}{\sin r} = \frac{4}{3} = \frac{v_r}{v_i}$$  \hspace{1cm} (VIII.2)

To find the position of the wave-front after refraction, CT is exten-
ded to meet the larger arc at H. GH therefore represents the wave-
front; it is oblique to the direction of propagation CH since, in the
triangle GTH, the angle GTH is a right angle.”

Here we saw the behaviour of the physical ray. Physical rays are tiny
fragments of “pulses”, and have characteristics irreducible to those
of mathematical rays. The changes of this pulse (by Sabra described
as “wave-front”) in different media is an important contribution by
Hooke. Note that in this conception the velocity of light is greater in
(optically) denser media—a view shared by Descartes, but later chal-
lenged by Huygens. One argument stated that the denser a medium,
the lower the proportion of subtile matter, and since light travels as
pulses in this subtile matter, the less subtile matter there is the faster
the pulses can propagate.
- **Hint to Problem 9** Colours arise near the edges of the physical ray of light after refraction. Generally “that part of the Ray towards which the refraction is made has the end of the orbicular pulse [spherical or circular pulse] precedent to that of the other side”. The second figure on page 230 shows how light is modified when it enters a more dense medium. At this stage the diagram only reflects the empirical findings.

Hooke, like most of his contemporaries holds that there are two “principal” colours, scarlet (red) and blue. The “intermediate” colours (yellow, watchet-blue, and from their mixing green) arise from “composition and dilutings of the two”.

After explaining the production colours based on the concept of principal colours, Hooke sets out to find the “reason of the Phenomena”—i.e. to give a micro-level account of the colours. This is summarized on 228.

The detailed diagram 4 on page 230 and the discussion on page 231 and 231 discuss how the forward end of the pulse is impeded by the resistance of the medium. It employs the concept of a medium: “that part or end of the pulse which precedes the other, must necessarily be somewhat more obtund, or impeded by the resistance of the medium”. This, however, happens at the edge of the pulse (page 230): “and therefore the ray AAAHB will have its side HH more deadned (sic) by the resistance of the dark or quiet medium PPP, Whence there will be a kind of deadness superinduced on the side HHH”.

6.2 **Text 2: Goethe’s Farbenlehre**

- The text is from the new translation of Goethe’s *Theory of Colours* reproduced with the kind permission of Suhrkamp Publishing House (Goethe 1988) ©Suhrkamp Publishers New York Inc., 1988. (Here the English translation of the most important other scientific works of Goethe cited in the chapter can also be found.) The references to earlier sections of the book have not been deleted, the numbers refer to the paragraphs in the original edition.
IX. DIOPTRIC COLORS

143. The term dioptric is applied to those colors which require a colorless medium for their creation, with the stipulation that light and dark must pass through the medium and affect either the eye or a surface standing opposite. It is therefore necessary that the medium be transparent or at least translucent to some degree.

144. In accordance with these requirements we will divide dioptric phenomena into two classes. We will include phenomena produced by translucent turbid media in the first, while the second will include those which appear when the medium is transparent in the highest possible degree.

X. DIOPTRIC COLORS OF THE FIRST CLASS

145. A space conceived of as empty would exhibit the property of transparency to an absolute degree. If this space is then filled by a substance imperceptible to the eye, a transparent medium will arise which is material in character and more or less dense. This medium may be aeriform or gaseous, liquid, or even solid.

146. Pure translucent turbidity is a property derived from transparency. Therefore it may also present itself to us in one of the three forms mentioned above.

147. Full turbidity is white. It is the brightest, most neutral form of space occupied by matter, and the first degree of opacity.

148. In empirical terms, transparency is the first degree of turbidity. The further stages of turbidity up to opaque white are infinite in number.

149. Brought into relation with light and darkness, any degree of turbidity short of opacity will yield phenomena which are simple and worthy of note.

150. The most energetic light is blinding and colorless (e.g., sunlight or phosphorus burning in oxygen). Similarly, the light of the fixed stars comes to us largely without color. When viewed through a medium which is the least bit turbid, however, this light will seem yellow. As the medium becomes more turbid or its thickness increases we will see the light gradually assume a yellow-red cast and ultimately intensify to ruby red.

151. On the other hand, darkness viewed through a turbid medium filled with light will create a blue color which grows lighter and paler as the medium becomes more turbid, but darker and deeper as it becomes more transparent. With the minimal degree of the most rarefied turbidity this color will appear to the eye as a beautiful violet.

152. The phenomenon described here occurs in the eye and may therefore be classed as subjective, but we can also find further corroboration for it in objective phenomena. A light subdued and beclouded in this fashion also casts a yellow, yellow-red, or purple illumination on objects, and although the effect
of darkness is not expressed so forcefully through a turbid medium, a blue sky will clearly be present along with other colored objects on the white paper in a camera obscura.

153. We can begin our discussion of the circumstances under which this important and fundamental phenomenon appears by considering atmospheric colors, most of which belong in this category.

154. Viewed through a certain degree of haze the sun appears as a yellowish disk. Often the center remains a blinding yellow even after the edge has turned red. The sun appears ruby red when a layer of fine dust is in the upper atmosphere (an event visible in northern Europe in 1794). When the atmospheric conditions peculiar to the sirocco prevail in southern regions the sun together with the clouds often surrounding it will seem an even deeper ruby red: these clouds will also radiate the color by reflection.

The red color at sunrise and sunset is produced in the same way. The sun is heralded by a red color because its rays come to us through a thicker layer of haze. The higher it climbs in the sky, the lighter and yellower its light becomes.

155. The darkness of infinite space viewed through atmospheric particles illuminated by sunlight will produce blue. On high mountains the heavens are seen by day as royal blue because only a few thin layers of haze float before the dark infinity of space. When we descend to the valleys this blue will become lighter until in certain regions, and with increasing haze, it finally changes completely to a whitish blue.

156. Mountains seem blue to us for the same reason: when so distant that the colors of their features are no longer visible and the light reflected from their surfaces no longer affects our eye, they will act as completely dark objects and look blue through the intervening haze.

157. When there is a fine haze in the air we also have the impression that the shadowed parts of nearby objects are blue.

158. Distant icebergs, however, still seem white with a tendency toward yellow. This is because even through the atmospheric haze they retain their effect of brightness on the eye.

159. The appearance of blue in the lower part of a candle flame also belongs in this category. If we hold the flame against a white background we will find no trace of blue, but this color will appear immediately when we hold the flame against a dark background. This phenomenon is most vivid when we light a spoonful of alcohol. Thus we may consider the lower part of the flame a haze; although extremely fine, this haze will become visible against the dark surface. It is so fine that we are able to read through it without difficulty. Conversely, the tip of the flame, through which no object can be seen, must be considered a body producing its own light.

160. Lastly, smoke must also be considered a turbid medium: it seems yellow or red against a light background, but blue against a dark one.
161. Turning now to the fluid media, we find that water of any sort will produce the same effect when made slightly turbid.

162. The infusion of nephritic wood (Guilandina Linnaei), an object of so much attention in the past, is merely a turbid liquid. It would necessarily look blue in a dark wooden cup but produce a yellow appearance when held to the sun in a transparent glass.

163. A few drops of perfume, spirit varnish, or certain metallic solutions are enough to create any degree of turbidity in the water to be used for such experiments. Tincture of soap is perhaps the most effective.

164. In bright sunlight divers see the ocean bottom as purple: here the water acts as a thick turbid medium. Under these circumstances they see shadows in green, the complementary color (§78).

165. The opal is foremost among solid media found in nature. At least in part, its colors may be attributed to the fact that it is really a turbid medium through which a mixture of light and dark substrata are visible.

166. The most satisfactory material for any of these experiments, however, is opal glass (vitrum astroides, girasole). It is manufactured by various methods and its turbidity is produced by metal oxides. Powdered and calcined bones may also be melted together with the glass to create turbidity and yield what is often called bone glass, but glass made by this method turns opaque far too quickly.

167. Glass for these experiments can be prepared in several ways. We may make it only slightly turbid, in which case we can progressively transform light from the palest yellow to the deepest purple by adding layers of glass. We may also use very turbid glass in sheets of varying thicknesses. Our experiments can be done by either method, but to see the deep blue color we must be especially careful to avoid glass which is too turbid or thick. Since darkness naturally produces a weak effect through turbidity, the medium will quickly turn white when too dense.

168. Clouded spots in window glass will cast a yellow light on objects: the same places in the glass look blue when we view a dark object through them.

169. Smoked glass also deserves mention here and must likewise be considered a turbid medium. It causes the sun to appear more or less ruby red, and although at first we might think that this color results from the brownish black of the soot, we can show that it is the consequence of a turbid medium by looking at a dark object through a lightly smoked glass illuminated frontally by the sun; we will then observe a bluish cast.

170. A striking experiment may be performed with sheets of parchment in a dark room. A single sheet of parchment over the aperture of a window shutter in the sunlight will seem whitish. Adding a second piece will produce a yellow color which will intensify and finally turn to red as we add further pieces of parchment.
171. We have already observed such an effect brought about in cataracts by a clouded lens (§132).

172. Although our discussion has now led us to the effect of a turbidity which allows very little light to pass, one strange case of temporary turbidity still deserves mention.

Some years ago the portrait of an eminent theologian was painted by an artist who was especially good at the practical use of color. His Reverence stood there in a glistening velvet coat which attracted as much attention as the man’s face and aroused great admiration. With the deposit of soot and dust the picture had gradually lost much of its original brilliance. It was therefore sent to an artist for cleaning and a new coat of varnish. The artist began to wipe the dirt carefully from the picture with a damp sponge. He had scarcely removed the worst of the dirt with a few strokes of the sponge when to his astonishment the black velvet of the coat suddenly changed to a light blue plush, giving the cleric a quite worldly if somewhat old-fashioned appearance.

The painter was too perplexed to continue with the cleaning: he could not understand how a light blue might serve as the ground for such a deep black, and still less how he could so easily have rubbed off a glaze heavy enough to transform the blue he now saw before him into black.

In short, he was aghast at having ruined the picture to this extent. Nothing of a religious nature remained apart from the richly curled, rounded wig, a feature which made the exchange of a fine new black velvet coat for one of faded plush altogether unsuitable. The damage seemed irreparable at the moment. Our good painter dejectedly turned the picture to the wall and retired full of care to his bed.

You may imagine his relief the following morning when he took up the picture once more and saw the black velvet coat restored to its full splendor. He could not resist the temptation to dampen one edge of the coat again, at which point the blue reappeared and then disappeared after a time.

When I heard of this phenomenon I immediately went to see the picture. While I was there a damp sponge was wiped across the picture and the transformation appeared in an instant. I saw a plush coat which was entirely pale blue, although somewhat faded: a few brown strokes on the arm of the coat indicated the folds.

My explanation for this is based on the theory of turbid media. The artist had probably applied a special varnish over an underlayer of black to give it depth. This varnish absorbed some moisture as it was washed, thereby becoming turbid and making the underlying black promptly appear as blue. Someone familiar with varnishes may discover by accident or deduction how to present this phenomenon experimentally to those interested in scientific research. Despite many attempts I have been unable to do so.
173. Our primary experiment with turbid media has helped us find an explanation for the most sublime atmospheric phenomena as well as effects more obscure but no less meaningful. These phenomena take many different forms in the world, and we are confident that observant friends of nature will continue to school themselves in the application of this approach to understanding and explaining them.

174. The principal phenomenon outlined in the above discussion might be called a fundamental or archetypal phenomenon. With the reader’s permission we will proceed at once to clarify what is meant by this.

175. In general, events we become aware of through experience are simply those we can categorize empirically after some observation. These empirical categories may be further subsumed under scientific categories leading to even higher levels. In the process we become familiar with certain requisite conditions for what is manifesting itself. From this point everything gradually falls into place under higher principles and laws revealed not to our reason through words and hypotheses, but to our intuitive perception through phenomena. We call these phenomena archetypal phenomena because nothing higher manifests itself in the world; such phenomena, on the other hand, make it possible for us to descend, just as we ascended, by going step by step from the archetypal phenomena to the most mundane occurrence in our daily experience. What we have been describing is an archetypal phenomenon of this kind. On the one hand we see light or a bright object, on the other, darkness or a dark object. Between them we place turbidity and through this mediation colors arise from the opposites; these colors, too, are opposites, although in their reciprocal relationship they lead directly back to a common unity.

176. In this sense we consider the error which has sprung up in scientific research on color to be a grievous one. A secondary phenomenon has been placed in a superior position and an archetypal phenomenon in an inferior one; moreover, the secondary phenomenon itself has been turned upside down by treating what is compound as simple and what is simple as compound. In this manner the most bizarre complications and confusions have come topsy-turvy into natural science, and science continues to suffer from them.

177. But even where we find such an archetypal phenomenon, a further problem arises when we refuse to recognize it as such, when we seek something more behind it and above it despite the fact that this is where we ought to acknowledge the limit of our perception. It is proper for the natural scientist to leave the archetypal phenomenon undisturbed in its eternal repose and grandeur, and for the philosopher to accept it into his realm. There he will discover that a material worthy of further thought and work has been given him, not in individual cases, general categories, opinions and hypotheses, but
in the basic and archetypal phenomenon.

XI. DIOPTRIC COLORS OF THE SECOND CLASS (REFRACTION)

178. After some observation we will soon find that the two classes of dioptric colors are closely related. Those of the first class appeared in the presence of turbid media; those of the second will now appear before us in transparent media. The close relationship between the two classes becomes evident when we consider that all transparent objects found in the empirical world may be viewed from the outset as turbid by nature, a fact demonstrated when we increase the mass of the medium we call transparent.

179. In dealing with transparent media, however, we will for now overlook the fact that they are inherently turbid to a degree, and concentrate our full attention on the phenomenon that arises here, a phenomenon known technically as refraction.

180. In our discussion of physiological colors we vindicated certain so-called optical illusions by showing them to be the activity of a healthy eye which is functioning properly (§2). Here we will again have an opportunity to say something on behalf of our senses which reaffirms their reliability.

181. Throughout the sensory world the relationship of one thing to another is of paramount importance, especially the relationship of the most significant thing on earth, man, to all the rest. Thus the world is divided into two parts and man as subject confronts the object. Here the practical person exhausts himself in experimentation, the thinker in speculation; they are required to enter a conflict never to be resolved peacefully nor concluded decisively.

182. But even here an accurate grasp of relationships is fundamental, and since our senses (insofar as they are healthy) most truly indicate outer relationships, we may conclude that wherever they appear to contradict reality they disclose the true situation all the more surely. Thus distant objects seem smaller to us, and through this very fact we become aware of their distance. Using colorless objects and colorless media we produced the phenomena of color, and thereby noted the degree of turbidity in such media.

183. Similarly, refraction reveals to our eye the varying degrees of thickness in transparent media, as well as other physical and chemical characteristics peculiar to them. This fact will lead us to undertake further tests to find the physical and chemical means of penetrating into these secrets fully, although from one standpoint they have already been laid open.

184. Objects seen through a medium having some degree of thickness do not appear where we would expect them to be according to the laws of perspective. Dioptric phenomena of the second class depend on this fact.

185. The laws of sight subject to mathematical formulation are based on the following: light travels in a straight line, and thus it should be possible to draw a straight line between the organ of sight and the object seen. Therefore
if we find light following a curved or broken line, if we see objects along a curved or broken line, we immediately recognize that the intervening medium has thickened, that it has somehow taken on a different character.

186. This deviation from the law of straight-line vision is generally called refraction. Although we assume that the reader is already familiar with refraction, we will describe it here briefly in its objective and subjective aspects.

187. We will let the sun shine diagonally into an empty cubical container so that the light falls only on the side opposite the sun not on the bottom. If we then fill the container with water, the position of the light will immediately change relative to the container. The light will draw back toward the side from which it came and illuminate part of the bottom. Where the light enters the thither medium it will deviate from its straight path and appear broken-hence the term “breaking” or “refraction”. Enough said of our objective experiment.

188. We may arrive at our subjective experiment as follows: let us locate our eye where the sun was, our line of sight also a diagonal over one side of the container so that the eye sees the entire inner surface of the side opposite, but not the bottom. When we pour water into the container the eye will also glimpse part of the bottom: this actually occurs in a way which leads us to believe we are still looking along a straight line, for the bottom appears to have been raised. This is why we will use the term “elevation” for this subjective phenomenon. Later we will discuss several other points of special interest in this regard.

189. To put this phenomenon into general terms we may repeat what was indicated above: the relationship of objects is altered or displaced.

190. In our present description, however, we intend to distinguish between objective and subjective effects, and so we will begin by describing the phenomenon subjectively with the statement that there has been a displacement of what we saw or were to see.

191. Something seen without boundaries may be displaced without our noting the effect. However when something seen as bounded is displaced we will have evidence of the displacement. Therefore if we wish to learn more about such a change in relationship we must limit ourselves largely to the displacement of what is bounded, the displacement of forms.

192. This effect as a whole may occur in media with parallel sides, since every such medium displaces the object by bringing it toward the eye along a perpendicular. However the displacement is more noticeable in media with nonparallel sides.

193. These can be completely spherical in shape or find application as convex or concave lenses; in our experiments we will also call upon such media. However, since they not only displace a form but also alter it in many ways, we prefer to use a medium with surfaces which are nonparallel yet flat; i.e., the prism, based on the triangle. We may think of it as part of a lens, but the prism
is especially useful here because it produces a strong displacement of the form without any significant distortion in shape.

194. To conduct our experiments with the greatest possible precision and avoid confusion, we will at first limit ourselves to subjective experiments; i.e., those in which the observer sees the object through a refractive medium. After we have dealt with these systematically, our objective experiments will follow in the same order.

XII. Refraction Without The Appearance Of Color

195. Refraction may occur without producing any appearance of color. Anything unbounded, a colorless or uniformly colored surface, will produce no color regardless of the degree to which refraction displaces it. This may be demonstrated in several ways.

196. If we place a glass cube on a homogeneous surface and view it perpendicularly or from an angle, the entire surface will be lifted toward the eye without any color appearing. When we look through a prism at a homogeneously gray or blue sky, or at a wall which is uniformly white or colored, the part of the surface we see will be displaced as a whole, but we will find no hint of color in it.

XIII. Conditions For The Appearance Of Color

197. In the above experiments and observations we found all homogeneous surfaces, large or small, to be without color. But color will appear at those boundaries where such a surface contrasts with a lighter or darker object.

198. Forms are created through a combination of boundary and surface. We will therefore state our basic observation as follows: a form must be displaced if colors are to appear.

199. Let us consider the simplest form, a light disk on a dark background (Plate VIII.2, fig. a). A displacement will occur in this form when we seemingly expand its borders away from the center by magnifying it. Any convex glass will accomplish this, and in this case we will observe a blue border (fig. b).

200. We can seemingly move the circumference of the same form inward toward the center by reducing the disk; in this case the borders will appear yellow (fig. c). A concave glass will do this, but the glass must be fairly thick, not ground thin as in ordinary eyeglasses. To observe this in a single experiment with the convex glass we may place a smaller black disk in the center of the light form lying on the black background. Magnification of a black disk on a white background has the same effect as reduction of the white disk: we move the black border toward the white one and therefore see the yellowish border color and the blue border color at the same time (fig. d).
201. These two colors, blue and yellow, appear across the white surface and at its edge. Where they extend across the black surface they take on a reddish appearance.

202. We have presented above the basic phenomena found in every appearance of color produced by refraction. These may, of course, be repeated, varied, enhanced, diminished, combined, complicated or confused in many ways, but in the end it is always possible to reduce them to their original, simple form.

203. Let us now consider what we have done. In the first case we seemingly moved the light border toward the dark surface, while in the second we moved the dark border toward the light surface: we replaced one with the other, thrust one across the other. We will now proceed step by step with the rest of our experiments.

204. When we displace the light disk as whole—prisms are particularly suited for use here—the disk will take on color in the direction of the apparent shift in accordance with the principles mentioned above. If we observe a disk in position a […] through a prism that it appears to be displaced in direction b, the outer edge will appear blue and blue-red (in accordance with the principle in figure B) while the inner edge will appear yellow and yellow-red (in accordance with the principle in figure C). This occurs because the light form appears to be shifted across the dark boundary in the first instance, while in the second the

Figure VIII.2: Goethe’s simple prismatic experiments. Light grey represents yellow and darker grey blue. Redrawn after the original plates by Goethe in (Goethe 1989a).
dark boundary is shifted across the light form. The same thing occurs when we seemingly shift the disk from a toward c, from a toward d, and so on through a full circle.

205. [The objects described in this paragraph are not depicted in Figure VIII.2.] Compound phenomena behave in the same manner as simple ones. If we look through horizontal prism ab at white disk e located some distance from the prism, the disk will be shifted in direction f and colored in accordance with the above principles. If we remove the horizontal prism and use vertical prism cd to view the form, it will appear at h with its colors conforming to the same principles. If we then place the two prisms across one another, the disk will appear to be displaced diagonally as required by natural law, and display the color produced by direction eg.

206. When we look closely at the colored borders lying opposite one another on the disk we will discover that they appear only in the direction of its apparent movement. Although a round form leaves us somewhat unclear about this relationship, a square form will provide unmistakable evidence of it.

207. Square form a displaced in direction ab or ad will display no color on the sides parallel to the direction of displacement. However, if the form is displaced in direction ac all four of its sides will appear colored, for it has moved diagonally.

208. Here we find confirmation for the assertion (§203 ff.) that the form must be displaced so that a light boundary appears to be shifted across a dark surface while a dark boundary is shifted across a light surface; the form appears to be shifted across the surface adjacent to it and the adjacent surface across the form. However, when the straight boundaries of a form are extended by refraction so that they run side by side without overlapping, no color appears: no color would appear even if the boundaries were extended to infinity.

XIV. CONDITIONS UNDER WHICH THE APPEARANCE OF COLOR INCREASES

209. We have seen above that the appearance of color in refraction is a result of displacing the boundary of a form across the form itself or across its background, of shifting the form, as it were, over itself or its background. In addition, with more pronounced displacement the appearance of color will increase. In subjective experiments (still the subject of our discussion) this occurs under the following conditions:

210. First, when the eye’s line of sight to a medium with parallel sides becomes more oblique.

Second, when the sides of the medium are no longer parallel and form an angle which is to some degree acute.
Third, through an increase in the mass of the medium: media with parallel sides may be enlarged in volume or the acute angle may be increased, provided it does not become a right angle.

Fourth, when the eye together with its medium of refraction are moved further from the form to be displaced.

Fifth, through a chemical property either added to the glass or intensified in it.

211. The maximum displacement of a form without appreciable distortion in shape is produced by using a prism, and thus the appearance of color may become quite powerful through a piece of glass fashioned in this way. However, in working with prisms we must try not to let these shining phenomena bedazzle us, but seek instead to keep the simple fundamentals established above firmly in mind.

212. The color which takes the lead when a form is displaced is always the broader one, and we will use the term “fringe” for it; the color remaining at the boundary is the narrower one, and we will use the term “border” for it.

213. When we shift a dark boundary across a light surface, a broad yellow fringe leads the way and a narrower yellow-red border follows at the boundary. When we displace a light boundary across a dark surface, a broad violet fringe takes the lead and a narrower blue border follows.

214. In a large form the central portion will remain without color: this center must be considered an unbounded surface which is displaced without being altered. But the center will be completely covered by color when the form is so narrow that (under the four conditions mentioned above) the yellow fringe can reach across to the blue border. For this experiment we can use a white stripe against a black background; the two extremes will easily merge across the white stripe and produce green. We will then see the following sequence of colors:

Yellow-red
Yellow
Green
Blue
Blue-red

215. If we place a black stripe on a white piece of paper, the violet fringe will extend across it to the yellow-red border. Here the black area in the middle will be eliminated, as was the white previously, and in its place a magnificent pure red will appear, a color we have frequently designated as purple. Now the sequence of colors will be as follows:

Blue
Blue-red
Purple
Yellow-red
216. The yellow and blue in the first instance (§214) may reach across one another far enough to merge the two colors as green, and the colored form will appear as follows:

Yellow-red
Green
Blue-red

Under similar circumstances in the second instance (§215) we will see only:

Blue
Purple
Yellow

This latter phenomenon appears beautifully when we look at bars of a window against a gray sky.

217. In all that we have observed we should always remember that this phenomenon must not be thought of as fixed or complete, but rather as evolving, growing, and open in many ways to modification. This is why a reversal of the five conditions listed above (§210) will lead to a gradual decrease in the phenomenon and ultimately to its complete disappearance.

XV. SOURCE OF THE FOREGOING PHENOMENA

218. Before continuing we will make use of the above to find a source or, if you prefer, an explanation for the fairly simple phenomena presented at the beginning of our investigation. This may give the friend of nature a clear insight into the more complex phenomena to follow.

219. Above all, we must remember that we are in the realm of forms. In general, our sense of sight is most attracted to what is seen as bounded. In our present discussion concerning the appearance of colors caused by refraction, we will consider nothing but what is seen as bounded, nothing but the form.

220. However, for the purposes of our chromatic observations we may divide forms in general into primary and secondary forms. The terms themselves indicate what is meant; the following will further clarify our meaning.

221. First, we may consider primary forms as original forms engendered in our eye by the object before us; such forms attest to the object’s real existence. In contrast, we may consider secondary forms as derived forms which remain in the eye when the object is removed, the afterimages and counterimages discussed at length in the section on physiological colors.

222. Secondly, we may also consider primary forms as direct forms which, like original forms, come from the object to our eye without mediation. In contrast, we may consider secondary forms as indirect forms merely passed
on to us secondhand by a reflecting surface. The latter are catoptric images which may also become double images in certain cases.

223. In other words, when the reflecting body is transparent and has two parallel surfaces lying one behind the other, an image from each surface may strike the eye. Double images will then be formed insofar as the upper image does not entirely cover the lower one. This can occur in various ways.

Let us hold a playing card up to a mirror. We will immediately observe the appearance of the card’s sharp, vivid image, but the border of the whole card and of each discrete form on it will be edged with a fringe which forms the beginning of a second image. The effect will vary from mirror to mirror, depending on differences in the thickness of the glass and random inconsistencies in polishing. This fringe shows up strongly in many mirrors when we stand in front of them wearing a white vest over dark clothing; here we may also clearly observe the double image of metal buttons on dark cloth.

224. Those familiar with the experiments described earlier (§80) will more easily follow the present discussion. Window bars reflected in a pane of glass will appear double, and their images may be separated entirely where the glass is thicker and the angle of reflection greater. Similarly, a container of water with a flat reflective bottom will show objects held before it as double and separated to a degree dictated by the circumstances. Here we will note that where the two images coincide a perfectly sharp image is actually created, but where this image separates and becomes double, forms appear which are weak, translucent, and ghostly.

225. We may use tinted media if we wish to discover which image is on the bottom and which on top, for a white form reflected from the bottom surface will have the color of the medium while its reflection from the top surface will have the complementary color. The reverse is true of dark forms; thus black and white squares are also useful here. This will provide another striking demonstration of how easily double images take on color or call it forth.

226. Thirdly, we may also consider primary forms as principal forms and, in effect, append secondary forms to them as auxiliary forms. Such an auxiliary form is a type of double image; however it cannot be divorced from the principal form although it constantly shows a tendency to become separate. It is these forms which will concern us when we come to prismatic phenomena.

227. An unbounded surface viewed by refraction produces no color (§195). The object viewed must be bounded, and therefore a form is required. The form is displaced by refraction, but not completely, not absolutely, not sharply: it is displaced incompletely so that an auxiliary form is created.

228. We must not come to a standstill when confronted by individual phenomena in nature, especially those which are significant or striking: we must not dwell on them, cling to them, or view them as existing in isolation. Instead, we should look about in the whole of nature to find where there is something
similar, something related. For only when related elements are drawn together will a whole gradually emerge which speaks for itself and requires no further explanation.

229. Hence it is appropriate to recall here that in certain cases refraction undeniably produces double images; e.g., in the case of what is called Iceland spar. Similar double images are also produced by refraction through large quartz crystals and in other ways—phenomena which have yet to receive the attention they deserve.

230. However, since the case under consideration (§227) involves auxiliary forms rather than double ones, we will return to a phenomenon mentioned earlier but not yet fully explored. We may call to mind our earlier observation (§16) that even with regard to the retina there is conflict of sorts in a light form on a dark background or a dark form on a light background. In this instance the light form seems larger and the dark one smaller.

231. On closer examination we will note that the forms are not sharply delineated against their backgrounds, but seem to exhibit a kind of gray, slightly colored border, an auxiliary form. If forms can produce these effects on the naked eye alone, what might they not do when a thick medium intervenes? The things we consider alive in the highest sense of the word are not the only ones to create effects and endure them; indeed, all things with any degree of interrelationship produce effects on one another, and often quite profound ones.

232. Thus when refraction affects a form, an auxiliary form is created close to the principal one. The true form seems to lag somewhat behind as if resisting the displacement, but an auxiliary form goes on ahead for a distance as described above (§§212-216), and in a direction determined by the movement of the refracted form across itself and its background.

233. We have also noted (§224) that double images appear as forms split in two, transparent phantoms of a sort, much in the way that double shadows always appear as half-shadows. The latter take on color easily and produce it readily (§69); the former do so as well (§80). This is also true of an auxiliary form: instead of separating from the principal form it projects from it as a form split in two—thus the rapidity, ease, and energy with which it takes on color.

234. There is more than one way to show that prismatic color is an auxiliary form. It appears in the exact shape of the principal form. Whether the principal form is straight or curved, serrated or undulating, the auxiliary form will always have the exact contours of the principal form.

235. The auxiliary form will share not only the shape of the actual form, but other features as well. Where the principal form contrasts sharply with the background (e.g., white on black), the colored auxiliary form will also appear quite forcefully: it will be vivid, clear, and powerful. It is most powerful, however, where a luminous form appears against a dark background. There are several ways to produce this effect.
236. But where the contrast between the principal form and its background is weak (e.g., gray forms on black and white, or even on one another), the auxiliary form will also be weak: it may become almost indiscernible where the difference in tone is small.

237. We can also observe remarkable effects with colored forms on a light, dark, or colored background. Here the color of the auxiliary form merges with the actual color of the principal form. The result is a compound color either enhanced by harmony or degraded by ugliness.

238. In general, however, semitransparency is a distinguishing feature of double and auxiliary forms. Thus if we suppose that within a transparent medium—with its tendency to become only semitransparent or translucent (noted above, §147)—if we suppose that within this medium there is a semitransparent phantom form, we will immediately recognize it as a turbid form.

239. Thus it is a simple matter to find the source for refractive color in the principles of turbid media. For where the forward fringe of the turbid form shifts away from a dark area and across a bright one, yellow appears: alternately, where a light boundary extends across the adjacent dark area, blue appears (§§150, 151).

240. The forward color is always the broader one. Thus the yellow extends across light in a broad fringe, but at the boundary of the dark area a narrower yellow-red edge is formed in accord with the principles of intensification and darkening.

241. On the opposite side the compressed blue remains at the boundary while the forward fringe spreads over the black as a thin veil of turbidity, thus giving us violet. This conforms to the principles indicated earlier in the section on turbid media: later we will find the effect of these principles equally evident in several other cases.

242. Since an explanation such as this must actually prove itself under the eyes of the researcher, we would ask of each reader that he study the above description with full attention, not just in passing. Here we have not sought to replace the phenomena with arbitrary symbols, letters, or whatever might suit our pleasure; here we have not passed on clichés to be repeated endlessly without thinking or giving cause for thought. We speak, instead, of phenomena which must be present before our physical eyes, and those of the mind, if we are to provide ourselves and others with a clear explanation of how they arise.

XVI. DECREASE IN THE APPEARANCE OF COLOR

243. To understand or produce a decrease in the appearance of color we need only reverse the progression of the five conditions (§210) leading to its increase. We will provide just a brief description and review of what the eye sees as this takes place.
244. When the opposite edges overlap fully, the colors will appear as follows (§216):

Yellow-red  Blue
Green        Purple
Blue-red     Yellow

245. With less overlap the phenomenon appears as follows (§§214, 215):

Yellow-red  Blue
Yellow       Blue-red
Green        Purple
Blue         Yellow-red
Blue-red     Yellow

Here the forms still appear wholly in color, but these sequences should not be thought of as primary ones developing out of one another like a series of steps or scales. We can and must separate them into their elements if we wish to learn more about their nature and character.

246. These elements are (§§199, 200, 201):

Yellow-red  Blue
Yellow       Blue-red
White        Black
Blue         Yellow-red
Blue-red     Yellow

Until now the principal form has been covered over and seemingly lost. Here it reemerges in the middle of the image and asserts its presence. Thus we recognize clearly the secondary nature of the auxiliary forms appearing as borders and fringes.

247. We may make these borders and fringes as narrow as we wish and even reduce refraction to the point where no color appears at the boundary.

This color phenomenon has now been sufficiently described. We will not declare it a primary phenomenon, for we have traced it to a simpler, more basic one: in conjunction with the principle of secondary forms, it originates in the archetypal phenomenon of light and dark seen through a turbid medium. Thus prepared we will describe in detail the effects produced by gray or colored forms when displaced by refraction, and so bring the section on subjective phenomena to a close.

- **Hint to Problem 11** When Goethe tackles the question of the physical colours, he starts with dioptric colours, colours that appear when light, darkness, and colourless transparent or translucent media interact. Dioptric colours of the first class (§§145-177) include phenomena produced by translucent turbid media (air, smoke, col-
loidal solutions, or the solution of *lignum nephriticum* also mentioned in Chapter IX).

The second class of dioptric colours, on the other hand, include colours “which appear when the medium is transparent in the highest possible sense”. However, we later learn that “all transparent objects found in the empirical world may be viewed from the outset as turbid by nature”. Spectral colours and coloured fringes seen through lenses are typical examples.

- **Hint to Problem 12** Goethe’s method in discussing the dioptric colours of the first class is to start with theory-directed definitions (defining turbidity as state of medium between pure translucence and white/opaque full turbidity, §§145-147) and important elements of the theory, like the “intensification” of colour (Steigerung, §150). The general rule is described in §151. Following this he investigates a large number of cases where the general rule is manifest more or less explicitly. The concatenation of phenomena leads up to the “archetypal phenomenon” (§175).

In a much earlier letter to his friend, Friedrich Schiller on January 17, 1798, Goethe depicts a structure of science that is very similar to the hierarchy discussed in §175.

> “The object of our work [i.e. investigating nature] would then be to demonstrate: (1) the empirical phenomenon, of which every individual is conscious in Nature and which later is elevated to (2) a scientific phenomenon by experimentation ... and (3) the pure phenomenon now standing forth as the result of all experiences and experiments.”

His hierarchy of these phenomena also resemble Bacon’s ideas (paralleling his “Senses and Particulars”, “Middle Axioms”, and “General Axioms”). His aim is, however, to arrive at fundamental (or basic) *phenomena*.

For Goethe, the discovery of the connections is the real task of science, not their explanation. His aim is to investigate the empirical conditions for the emergence of colour, not accepting that colours etc. are “secondary” qualities in relation to “primary”, quantitative-physical ones, and that they are derived from the latter.
One of the reasons for his attempts to remain on the phenomenal level is the caution from theory-driven experimentation, which, in his view is all too likely to end up verifying the hypotheses in question. As he wrote earlier in his *The Experiment as Mediator Between Subject and Object*: “I venture to assert that one experiment, even several experiments combined, prove nothing; indeed, that nothing can be more dangerous than the attempt to confirm a theory by experiments; and that the greatest errors have arisen precisely because its dangers and its inadequacies were not realized.”

As he summarized his aim in the same essay: “My purpose is to collect all data in the field, to set up my own experiments and carry them out in the greatest diversity, by methods easily duplicable and within range of more individuals than heretofore; furthermore, to formulate the propositions by which data of the higher type can be expressed and to have the patience to learn whether these too may be subordinated under a higher law. If imagination and wit should nevertheless impatiently hurry ahead, the procedural method will itself indicate the point to which they must again return.”

The result of such work should be the “archetypal phenomenon” and discovering the network of causes that if modified can either strengthen or weaken the effect (§231).

- **Hint to Problem 13** When light is seen through the medium, it turns yellow, when darkness is seen through a medium that is lit, it appears blue (§175). This is clearly a modificationist explanation employing only the concepts of light, darkness, and the medium. There is, however, no mentioning of edges of the meeting of light and dark at boundaries or surfaces that light has to pass through to be modified. It is thus not a boundary-modificationist account, but a medium-modificationist one.

The dioptric colours of the second class also appear due to the modification of light following a displacement of the image (§191, §§203-208). For their explanation Goethe introduces the concept of the auxiliary and double images [*Doppel und Nebenbildern*]. A distinction is made between the primary (real, principal) image and the secondary (indirect, auxiliary) image (§221). This model states that modification of the refracted image starts from the edges. Therefore colour
production is at the displaced edges of forms—modification is at the boundaries.

But the colours seen in refraction, principally treated by Goethe as boundary-modifications, can be derived from a medium-modificationist framework. This is achieved by treating the auxiliary forms as media through which the principal form is seen (§238). This medium gives rise to colours—just the way dioptric colours of the first class are seen. There are a number of ad hoc elements in the explanation, and this becomes even more obvious when e.g. epoptic (interference) colours are explained with recourse to the same theory. Goethe’s archetypal phenomenon belongs to medium-modificationism, and he sees in this a unified explanation for all colour-appearances. This is a curious case, where for the sake of unification he sacrifices the phenomenological approach (§247).

**Hint to Problem 14** The study of illusions have always formed an integral part of some of the approaches to the study of vision (see the works of Ptolemy and Lucretius in the earlier chapters). In the seventeenth and eighteenth century, however, these were mainly treated as failures of perception and little systematic study was carried out.

In the early nineteenth century this attitude fundamentally changed. Goethe states that the study of perceptions even if they appear to contradict reality (§182) is fruitful in understanding man’s relation to the world. His preoccupation with the occurrence and lawfulness of after-images testifies to the same interest. In his *Theory of Colours* he also discusses the recently discovered colour-blindness (often referred to as Daltonism). Even his early work in plant and animal morphology in the late 1780’s and early 1790’s showed this interest and aberrant plant organs etc. are used to test his morphological theory.

7 Epilogue

We have investigated two modificationist theories in this chapter, leaving out many. This gargantuan and heterogeneous group of theories has been little studied. When investigating them more closely or classifying them, a number of questions have to be asked.
Does darkness play any role in the formation of colours? If yes then on the edge of light or inside the medium, or at the surface of bodies? Does the modification give rise to a series of colours, or first a number of primaries is created (if yes, are these immutable?) and the other colours are mixed from these? Does the modificationist account tackle questions of perception or questions of physical optics? And in evaluating the theories what is the most important: economy of thought, plausibility, explanatory or predictive power, or other factors? How do these accounts relate to one another?

But as there is no detailed, comprehensive work in this field, these questions are for the moment left unanswered. We cannot even reply to the trivial (and anachronistic) question: is modificationism a success or a failure? While these two theories have been surpassed by the Newtonian one, some critics, like (Sabra 1967, Sepper 1988) claim that the debate was not restricted to Newton’s cohorts and Goethe. In the nineteenth century, a very similar debate took place (Henri Poincaré as opposed to Lord Rayleigh, M. Gouy, and Arthur Schuster), this time more favouring the modificationist position concerning the original compositeness of white light.

8 Further Problems

- **Problem 15** Evaluate Hideto Nakajima’s claim that Aristotle’s conception can be summarised as follows (Nakajima 1984): “(a) White is the original colour of light and other colours are produced by the modification of it. (b) The modification is raised by the admixture of darkness to light, and colour of light is determined by the quantity of darkness it contains. Red is the nearest approach to white. It contains the least darkness. As the admixture of darkness increases and the strength of light declines, there appear first green and finally violet, etc.”

- **Problem 16** Alan Shapiro suggests that until the seventeenth century it was believed that “colors arise from a mixture of light with darkness, that is white and black” (Shapiro 1994). He also claims that a new theory was proposed in the seventeenth century by Descartes, and Hooke among others, “abandoning the idea that color arises from light and shadow and made it depend on the relative strength of successive light pulses or waves”. Shapiro writes: “This fundamental
shift away from white and black, or light and shadow, I am confident, reflects the artist’ discovery that white and black are not true colors and cannot by their mixture generate chromatic colors”. Discuss this last statement.

- **Problem 17** Reconstruct and discuss Hooke’s geometrical proof on page 229. Evaluate Hooke’s attack on the Cartesian theory. Compare Hooke’s terminology with that of Descartes’s and Newton’s.

9 **Suggested Readings**

There are only few works concentrating on the modificationist tradition itself. As a good starting point see (Guerlac 1986), for Hooke’s position (Nakajima 1984), for Goethe’s attack (Steinle 1993b). See also the references in this chapter and my PhD thesis (Zemplén 2001) under http://hps.elte.hu/~zemlen/PhDthesis.pdf.

On the seventeenth century theories (Sabra 1967) is still a classic, on the wave-theory (Hakfoort 1995) is a valuable source.

The two-volume (Amrine 1996) gives around 10,000 titles in a massive bibliography on the evaluation of Goethe’s scientific work. The contemporary setting is analysed in (Burwick 1986), or (Cunningham 1990), the acceptance of the Newtonian theory in Germany in (Clark 1997).

Fortunately more and more of the material is available online, like http://www.colors-system.com/, a rich source of historical colour theories. Apart from paying services, like http://goethe.chadwyck.com/, Goethe’s *Theory of Colours* can be read in the original under i.e. http://www.steinerschule-bern.ch/goethe/farbenlehre.htm, as well as on other public servers.
Chapter IX

NEWTON’S 1672 “NEW THEORY”

1 INTRODUCTION

The following section, unlike the other chapters of the book consists of a detailed analysis and investigation of a single scientific journal article. Newton’s first publication is based on a letter sent by Newton to Henry Oldenburg, Secretary of the Royal Society in 1672.

The letter appeared in the *Philosophical Transactions* Volume 6, 3075-87, with emendations. The published article and the argumentation of the *experimentum crucis* is analysed in the “sister-volume” of the present book (Lampert 2000). In this chapter the main stress is on understanding the structure and aim of the article, on making the tacit elements of the arguments transparent.

As modern science is intimately bound to quantification and calculation, this aspect of the article will also be thoroughly investigated.

2 THE CLOSE READING OF THE LETTER

Contrary to the other chapters in the book, the text will be broken down to small units, all preceded by an introduction to concepts and smaller questions and problems. The introductions contain additional information that is useful for reading the text, and to direct the attention of the reader. They may be skipped and read only when difficulties surface in reading the text. At the end of each unit some hints to the problems and figures are presented.

While in other chapters a less thorough reading was dominant, it seems beneficial in one case at least to pay close attention to as many details as possible and present some of the surprising difficulties that arise with this kind of work. Newton’s letter (also his first published article) seems ideal for this task: it is the concise presentation of a significant new theory and properly understanding it was difficult even for the contemporaries.
2.1 Preliminary Observations

The Opening Lines

The opening paragraphs clearly point to the problem field of the letter. It was known since Antiquity that spherical surfaces do not collect the rays of the sun into a single point (see the parabolic and spherical mirrors on figure IX.1).

Figure IX.1: Burning lenses from the second edition of Athanasius Kircher’s *Ars magna lucis et umbrae* printed in 1671 in Amsterdam. The small diagrams show the differences in the focal characteristics of different mirrors when parallel rays are incident.

Newton knew Descartes’s work, in which he shows that spherical lenses also have aberration: they do not collect all parallel rays into a single focus point. Newton attempts to create non-spherical lenses, to overcome this problem, which greatly plagued the instrument-makers of the time. We know from his manuscripts that he did not only try to find the curvature
of such lenses by calculation, but also designed polishing apparatuses for grinding aspherical lenses.

- **Problem 1** Explain why, after mentioning the grinding of lenses Newton switches to talk about a glass prism.

- **Problem 2** How does Newton set up the experiment and what necessitates the setup? How does he describe the image formed?

_Sir,_

_To perform my late promise to you, I shall without further ceremony acquaint you, that in the beginning of the Year 1666 (at which time I applied my self to the grinding of Optick glasses of other figures than Spherical,) I procured me a Triangular glass-Prisme, to try therewith the celebrated Phaenomena of Colours. And in order thereto having darkened my chamber, and made a small hole in my window-shuts, to let in a convenient quantity of the Suns light, I placed my Prisme at its entrance, that it might be thereby refracted to the opposite wall. It was at first a very pleasing divertisement, to view the vivid and intense colours produced thereby; but after a while applying my self to consider them more circumspectly, I became surprised to see them in an oblong form; which, according to the received laws of Refraction, I expected should have been circular._

_They were terminated at the sides with streight lines, but at the ends, the decay of light was so gradual, that it was difficult to determine justly, what was their figure; yet they seemed semicircular._

_Comparing the length of this coloured Spectrum with its breadth, I found it about five times greater; a disproportion so extravagant, that it excited me to a more then ordinary curiosity of examining, from whence it might proceed._

- **Hint to Problem 1** Already in Antiquity methods have been developed for solving complicated problems of approximation in mathematics. The method of _exhaustio_ employed a series of calculations that approximated a value. For example, to calculate the area of a circle the area of polygons inscribed into the circle or circumscribed around the circle can be calculated. The area of the circle will always be greater than the area of the inscribed polygon and smaller than the circumscribed one.

Similarly here, instead of investigating the properties of a lens, Newton simplifies the object of study. The curvature of the lens can be
approximated with flat surfaces. A prism functions as a part of a lens, maintaining its main property: it contains two glass surfaces that are not parallel. Two prisms base to base resemble a convex lens, apex to apex a concave one. Using prisms Newton has a simpler tool for investigating certain properties of lenses—his original question.

**Hint to Problem 2** The highly specific setup of the experiment suggests that it is more than a chance observation. Newton’s notebooks suggest (McGuire & Tamny 1983) that he only arrived at this setting after several years of experimenting. His aim is to conform to the customs of the *Royal Society*, laying much emphasis on unbiased observation.

Interestingly, though Newton set up the experiment in order to observe the *colours*, his description of the image lacks any reference to them. The figure given is idealized, as the intensity levels across the spectrum and and the eye’s sensitivity is not constant: the appearing spectrum is in reality pear- or egg-shaped.

His surprise at the *non-circular* image is only explicable if the prism is held in a way that the *incident* and the *refracted* rays enclose the same angle with the surface of the prism. This position of *minimum deviation* is symmetrical, and requires very precise setting of the prism.

**The First Modificationist Theories**

Newton’s interest turns to the *cause* of the elongated image. He tries to rule out some of the possible (modificationist) explanations.

**Problem 3** Discuss how Newton eliminates the modificationist alternatives. Do you consider the elimination successful?

*I could scarce think, that the various Thickness of the glass, or the termination with shadow or darkness, could have any Influence on light to produce such an effect; yet I thought it not amiss to examine first these circumstances, and so tried, what would happen by transmitting light through parts of the glass of divers thicknesses, or through holes in the window of divers bignesses, or by setting the Prisme without, so that the light might pass through it, and be refracted before it was terminated by the hole: But I found none of those circumstances material. The fashion of the colours was in all these cases the same.*
Then I suspected, whether by any unevenness in the glass, or other contingent irregularity, these colours might be thus dilated. And to try this, I took another Prisme like the former, and so placed it, that the light, passing through them both, might be refracted contrary ways, and so by the latter returned into that course, from which the former had diverted it. For, by this means I thought, the regular effects of the first Prisme would be destroyed by the second Prisme, but the irregular ones more augmented, by the multiplicity of refractions. The event was, that the light, which by the first Prisme was diffused into an oblong form, was by the second reduced into an orbicular one with as much regularity, as when it did not at all pass through them. So that, what ever was the cause of that length, 'twas not any contingent irregularity.

- **Hint to Problem 3** Newton shows that different glass-thickness has no effect on the spectrum\(^1\). Similarly, the unevenness of the prism and other contingent factors are eliminated using more than one prism. Newton also rules out the possibility that shadow or darkness can be the cause of the elongated image, even though in all mentioned cases the light bundle is bounded by shadow or darkness. A light-shadow boundary can thus still be a necessary condition of the production of colour as in the theory of Descartes (see Chapter VII).

**The First Measurements**

Newton now investigates the phenomenon in more detail. He gives very precise measurements, but does not detail the measuring procedures. The very concise formulation does not distinguish between the data Newton calculated and those that he measured. The measuring procedures are not included. The setting is as de-

\(^1\)More precisely that it has no significant effect. The prisms of the time were of rather poor quality: the sides were not even, they contained bubbles, streaks etc.
scribed at the beginning of the letter, the prism is held with the refractive angle pointing to the floor.

The *incident* rays of the sun, after passing through the hole are refracted twice by the sides of the prism. At the angle of minimum deviation the two refractions are symmetrical. The incident light of the sun in the first refraction encloses the same angle with a perpendicular drawn to the first face of the prism as the refracted (outgoing) light of the second prism with a line drawn perpendicularly to the face of the second prism. Similarly, the light within the prism encloses the same angle with the perpendiculars drawn to both faces.

The room is 22 feet long. The rays after the two refractions fall perpendicularly on the wall, but if the prism is taken away, then they illuminate the floor of the room. The angle between these rays is 44° 56’.

In the 17th and 18th centuries the standard metaphor for seeing was *camera obscura*, a dark box or chamber with a small hole at the front (with or without a lens) producing a small, inverted image (see Figure above). Newton’s whole experimental setup works like a *camera obscura*. Inside the room an inverted image of the sun (and the whole sky) is produced. In a *camera obscura* objects that are seen under a certain angle from the opening produce an image that subtends the same angle (see more in Chapter X). The sun subtends an angle of about 30’, so the image of the sun should also subtend 30’. But the size of the opening in the shutter is relatively large: one point of the image is illuminated by several points of the sun and/or the sky, and so the spectrum appears enlarged. When Newton subtracts the size of the hole from the image, he wants to investigate the image, as if the hole itself was infinitesimally small.

*Problem 4* Draw a diagram of the experiment, and mark the measured values. Which values are measured and which only calculated? The measurements are surprisingly accurate – yet the sun and the image are moving. Calculate the speed of the movement of the image on the wall – is it large enough to obstruct the measurements? There is a simplification in the paragraph which is left unexplained. Which is this?

---

2 A note on measurements. A foot is around 30.5 cm, and is divided into 12 inches. A whole circle is divided into 360° (degrees), and each degree consists of 60’ (minutes).
I then proceeded to examine more critically, what might be effected by the difference of the incidence of Rays coming from divers parts of the Sun; and to that end, measured the several lines and angles, belonging to the Image. Its distance from the hole or Prisme was 22 foot; its utmost length 13 1/4 inches; its breadth 2 5/8 inches; the diameter of the hole 1/4 of an inch; the angle, which the Rays, tending towards the middle of the image, made with those lines, in which they would have proceeded without refraction, 44 deg. 56’. And the vertical Angle of the Prisme, 63 deg. 12’. Also the Refractions on both sides the Prisme, that is, of the Incident, and Emergent Rays, were as near, as I could make them, equal, and consequently about 54 deg. 4’. And the Rays fell perpendicularly upon the wall. Now subducting the diameter of the hole from the length and breadth of the Image, there remains 13 Inches the length, and 2 3/8 the breadth, comprehended by those Rays, which passed through the center of the said hole, and consequently the angle at the hole, which that breadth subtended, was about 31’, answerable to the Suns Diameter; but the angle, which its length subtended, was more then five such diameters, namely 2 deg. 49’.

- **Hint to Problem 4** Some of the measurements are probably only inferred. To find the angle of minimum deviation Newton probably moved the prism until .... Thus the 54°4’ value is most likely due to calculation: 63°12’ + 44°56’ = 108°8’. This, divided by two gives the value above.

The Earth makes one revolution (360°) every 24 hours. Thus every minute the Sun is seen to move 360 ° 24 ÷ 60 = 15’ minutes. The movement of the image is characterised by the same displacement in angles. If the image is 22 feet away, then the displacement is more than 1.15 inch/minute, i.e. nearly 3 cm/minute.

Subtracting the diameter of the opening from the length and breadth of the image to acquire the proper angular divergence of the sun’s rays is a simplification that yields only an approximate answer (but the error is not significant).

- **Problem 5** If the prism was in the angle of minimum deviation, then knowing the angle of incidence and the resulting deviation of the rays after two refractions the angle of refraction within the prism
can be calculated. Calculate this angle. What is the ratio of the sines of these angles? What is the refractive power of the glass?

- **Hint to Problem 5** The two refractions alter the direction of the rays by $44^\circ 56'$. The refractions being symmetrical this implies that the first angle of refraction is $54^\circ 4' - (44^\circ 56' \div 2) = 31^\circ 36'$. The ratio of the sines of the angle of incidence and that of refraction is $31^\circ 36' \div 54^\circ 4' = 1.545$, around 20 to 30.9.

**THE RECEIVED LAWS OF REFRACTION**

Newton arrives at the angular measurements of the image. The breadth of the spectrum is $31'$, corresponding to the diameter of the sun. The length is, however, far greater. He calculates whether this could be the result of the two refractions. Although not detailed here, Newton carried out the calculations a few years earlier in one of his lectures.

But the experimental data does not fit the theoretical value arrived at by using the law of sines, one of the most significant discoveries in seventeenth century optics. Newton first investigates his apparatus, as the calculated value should only be manifest at the angle of minimum deviation. But the contradictory phenomenon is very robust, movement of the prism hardly alters the image.

He also investigates another form of modificationism: may be the propagation of light is not rectilinear? But he refutes the theory that closely resembles Descartes’s explanation of refraction. The diagram here reproduces

---

3 The sine of an angle is the ratio of the longest side and the side opposite the angle in a triangle, where one of the angles is $90^\circ$.

4 As the professors at Cambridge had to deposit copies of the material of their lectures, the manuscript of Newton’s first lecture course, the *Lectiones opticae*, held in Latin in 1670-72 was preserved and later published.
duced is from Descartes, showing that in the Cartesian theory refraction results from the spinning of the corpuscles.

Having made these observations, I first computed from them the refractive power of that glass, and found it measured by the ratio of the sines, 20 to 31. And then, by that ratio, I computed the Refractions of two Rays flowing from opposite parts of the Sun’s discus, so as to differ 31’ in their obliquity of Incidence, and found, that the emergent Rays should have comprehended an angle of about 31’, as they did, before they were incident.

But because this computation was founded on the Hypothesis of the proportionality of the sines of Incidence, and Refraction, which though by my own /and others Experience I could not imagine to be so erroneous, as to make that Angle but 31’, which in reality was 2 deg. 49’; yet my curiosity caused me again to take my Prisme. And having placed it at my window, as before, I observed, that by turning it a little about its axis to and fro, so as to vary its obliquity to the light, more then by an angle of 4 or 5 degrees, the Colours were not thereby sensibly translated from their place on the wall, and consequently by that variation of Incidence, the quantity of Refraction was not sensibly varied. By this Experiment therefore, as well as by the former computation, it was evident, that the difference of the Incidence of Rays, flowing from divers parts of the Sun, could not make them after decussation diverge at a sensibly greater angle, than that at which they before converged; which being, at most, but about 31 or 32 minutes, there still remained some other cause to be found out, from whence it could be 2 deg. 49’.

Then I began to suspect, whether the Rays, after their trajectory through the Prisme, did not move in curve lines, and according to their more or less curvity tend to divers parts of the wall. And it increased my suspicion, when I remembred that I had often seen a Tennis-ball, struck with an oblique Racket, describe such a curve line. For, a circular as well as a progressive motion being communicated to it by that stroak, its parts on that side, where the motions conspire, must press and beat the contiguous Air more violently than on the other, and there excite a reluctancy and reaction of the Air proportionally greater. And for the same reason, if the Rays of light should possibly be globular bodies, and by their oblique passage out of one medium into another acquire a circulating motion, they ought to feel the greater resistance from the ambient Aether, on that side, where the motions conspire, and thence be continually bowed to the other. But notwithstanding this plausible ground of suspition, when I came to examine it, I could observe no such curvity in them. And besides (which was enough for my purpose) I observed, that the difference betwixt the length of the Image, and diameter of the hole, through which the light was
transmitted, was proportionable to their distance.

2.2 The experimentum crucis

In the first part of the letter Newton drew attention to the elongation of an image after refraction through a prism. The phenomenon could not be explained with the known laws of geometrical optics. At least one of these laws had to be modified to account for the newly discovered phenomenon.

One solution would be to drop the sine-law, and say that there is as yet no discovered regularity between the angle of incidence and of refraction. But saving the law of sines (even if in a modified form) seemed more desirable for Newton.

Newton sets up a new experiment, now operating with two prisms. He uses the term experimentum crucis, to signify that it is an experiment at “crossroads”: the outcome decides which way to go and which way to abandon (which theory should not be chosen). The experiment is analysed in detail in (Lampert 2000).

- **Problem 6** The term experimentum crucis clearly implies that the experiment is decisive, the experimenter has to chose between two (or more) hypothesis. Which are the possible options here?

The gradual removal of these suspitions at length led me to the Experimentum Crucis, which was this: I took two boards, and placed one of them close behind the Prisme at the window, so that the light might pass through a small hole, made in it for that purpose, and fall on the other board, which I placed at about 12 foot distance, having first made a small hole in it also, for some of that Incident light to pass through. Then I placed another Prisme behind this second board, so that the light, trajected through both the boards, might pass through that also, and be again refracted before it arrived at the wall. This done, I took the first Prisme in my hand, and turned it to and fro slowly about its Axis, so much as to make the several parts of the Image, cast on the second board, successively pass through the hole in it, that I might observe to what places on the wall the second Prisme would refract them. And I saw by the variation of those places, that the light, tending to that end of the Image, towards which the refraction of the first Prisme was made, did in the second Prisme suffer a Refraction considerably greater then the light tending to the other end. And so the true cause of the length of that Image was detected to be no other, then that Light consists of Rays differently refrangible, which,
without any respect to a difference in their incidence, were, according to their
degrees of refrangibility, transmitted towards divers parts of the wall.

- **Hint to Problem 6** The rival hypotheses aim to answer the ques-
tion: why is the spectrum elongated. As a result of the experiment
Newton concludes “that Light consists of Rays differently refran-
gible”. This statement implies that the criticised hypothesis is one
which maintains the law of sines in an unmodified form, and explains
the elongation by some sort of modification of the light.

2.3 **Glass-works**

**The Magnitude of Chromatic Aberration**

Newton’s conclusion has serious implications for his enterprise to grind
lenses that focus all light into one point. He calculates the magnitude of the
*chromatic aberration* for one of his prisms. He expresses the ratio of the sines
of incidence and refraction on the air-glass boundary as a ratio between
whole numbers. He also investigates the effect of the different refrangibility
of rays on the focusing properties of *refractors*, i.e. telescopes that employ
lenses.

- **Problem 7** How does Newton calculate the minimum size of the
  focus in relation to the size of the aperture?

* When I understood this, I left off my aforesaid Glass-works; for I saw,
  that the perfection of Telescopes was hitherto limited, not so much for want of
glasses truly figured according to the prescriptions of Optick Authors, (which
all men have hitherto imagined,) as because that Light it self is a Heterogeneous
mixture of differently refrangible Rays. So that, were a glass so exactly figured,
as to collect any one sort of rays into one point, it could not collect those also
into the same point, which having the same Incidence upon the same Medium
are apt to suffer a different refraction. Nay, I wondered, that seeing the differ-
ence of refrangibility was so great, as I found it, Telescopes should arrive to that
perfection they are now at. For, measuring the refractions in one of my Prisms,
I found, that supposing the common sine of Incidence upon one of its planes
was 44 parts, the sine of refraction of the utmost Rays on the red end of the
Colours, made out of the glass into the Air, would be 68 parts, and the sine of
refraction of the utmost rays on the other end, 69 parts: So that the difference is
about a 24th or 25th part of the whole refraction. And consequently, the object-
glass of any Telescope cannot collect all the rays, which come from one point of
an object so as to make them convene at its focus in less room then in a circular
space, whose diameter is the 50th part of the Diameter of its Aperture; which is
an irregularity, some hundreds of times greater, then a circularly figured Lens,
of so small a section as the Object glasses of long Telescopes are, would cause
by the unfitness of its figure, were Light uniform.

• **Hint to Problem 7** It is not trivial how Newton arrives at his 1/50 ratio. Newton in a letter to Oldenburg on 8 July 1672 returns to the
issue (Turnbull 1959).

Suppose DF be ye lens; CD & EF two line parallel to its
axis… And …let DH & FG be the most refracted and DG & FH the least refracted intersecting he former in G & H. …Now since by my Principles the difference of refraction of
the most difform rays is about the 24th or 25th part of their
whole refraction, the angle GDH will be about 25th part of the
angle MDH and consequently the subtense GDH will be about
25t part of the angle MDH and consequently the subtense GH
(wch is the diameter of the least space into wch the refracted
rays converge) will be about a 25t part of the subtens MH and
therefore about 49th part of ye whole line MN the diameter of
the lens; or in round numbers about a fiftieth part as I asserted.

Sepper asserts a formula, but does not prove it (Sepper 1994). Ac-
cording to this if I is the (common) sine of incidence, R is the sine of
refraction of the least refrangible rays, and S is the sine of refraction
of the most refrangible rays, then \((S - R) \div (S + R - 2I)\) times
the size of the aperture gives the size of the smallest possible area
on which the light coming from a single point can be focused. For
Newton’s values are:

\[
(69 - 68) \div (69 + 68 - 2 \times 44) = 1/49 \quad \text{IX.1}
\]

which is very close to 1/50.

**Reflecting Telescopes**

Not only lenses can collect light into a given point. Mirrors can also be
employed in optical instruments. As here only *spherical aberration* has to be
overcome, with the development of polishing techniques parabolic mirrors could be used in optical instruments.

- **Problem 8** What are the benefits of *catoptric* optical instruments over *dioptric* ones? What are the difficulties of production?

This made me take Reflections into consideration, and finding them regular, so that the Angle of Reflection of all sorts of Rays was equal to their Angle of Incidence; I understood, that by their mediation Optick instruments might be brought to any degree of perfection imaginable, provided a Reflecting substance could be found, which would polish as finely as Glass, and reflect as much light, as glass transmits, and the art of communicating to it a Parabolick figure be also attained. But these seemed very great difficulties, and I almost thought them insuperable, when I further considered, that every irregularity in a reflecting superficies makes the rays stray 5 or 6 times more out of their due course, than the like irregularities in a refracting one: So that a much greater curiosity would be here requisite, than in figuring glasses for Refraction.

Amidst these thoughts I was forced from Cambridge by the Intervening Plague, and it was more then two years, before I proceeded further. But then having thought on a tender way of polishing, proper for metall, whereby, as I imagined, the figure also would be corrected to the last; I began to try, what might be effected in this kind, and by degrees so far perfected an Instrument (in the essential parts of it like that I sent to London,) by which I could discern Jupiters 4 Concomitants, and shewed them divers times to two others of my acquaintance. I could also discern the Moon-like phase of Venus, but not very distinctly, nor without some niceness in disposing the Instrument.

From that time I was interrupted till this last Autumn, when I made the other. And as that was sensibly better then the first (especially for Day-Objects,) so I doubt not, but they will be still brought to a much greater perfection by their endeavours, who, as you inform me, are taking care about it at London.

I have sometimes thought to make a Microscope, which in like manner should have, instead of an Object-glass, a Reflecting piece of metall. And this I hope they will also take into consideration. For those Instruments seem as capable of improvement as Telescopes, and perhaps more, because but one
reflective piece of metal is requisite in them, as you may perceive by the annexed diagram, where A B representeth the object metall, CD the eye glass, F their common Focus, and 0 the other focus of the metall, in which the object is placed.

- **Hint to Problem 8** Apart from the benefits Newton lists, catoptric instruments are also smaller. The one he sent to the Royal Society was about fifteen times four centimeter, with a magnification of around thirty (Hecht 2001).

  Although Newton’s reflecting telescope caused sensation, within a matter of weeks the finely polished metal surface of the mirror became slightly corroded. Not only the choice of metal was crucial, it also had to be polished very precisely. To produce a spherical surface by polishing is relatively easy, but to find the suitable parabolic shape requires new methods. The heavy metal is also likely to be deformed due to its own weight, which further complicates the production of large telescopes (these let more light in and are more suited to astronomical investigations). Even catoptric instruments use lenses, so the chromatic aberration cannot be wholly eliminated.

2.4 **The Origin of Colours**

**The Mathematical Science of Colours**

Newton returns to his previous experiments and once again summarizes the conclusion of the *experimentum crucis*. This is followed by one of the most controversial points of the letter which was left out of the printed version (Newton 1671-72). Here Newton states that colour can be quantified, and the theory of colours can be incorporated into the mathematical science of optics.

  He also changes the style of the letter. The seemingly inductive, Baconian narration is taken over by the listing of propositions. In establishing differential refrangibility colour played no role. The second part of the letter, however, discusses a theory of colour intimately connected to the newly discovered property of light.

- **Problem 9** What supports Newton’s statement that his theory of colour is “not an Hypothesis but most rigid consequence”? Compare this conclusion of the *experimentum crucis* to the original one.
But to return from this digression, I told you, that Light is not similar, or homogeneal, but consists of difform Rays, some of which are more refrangible than others: So that of those, which are alike incident on the same medium, some shall be more refracted than others, and that not by any virtue of the glass, or other external cause, but from a predisposition, which every particular Ray hath to suffer a particular degree of Refraction.

I shall now proceed to acquaint you with another more notable difformity in its Rays, wherein the Origin of Colours is infolded. A naturalist would scarce expect to see ye science of those become mathematicall, & yet I dare affirm that there is as much certainty in it as in any other part of Opticks. For what I shall tell concerning them is not an Hypothesis but most rigid consequence, not conjectured by barely inferring ‘tis thus because not otherwise or because it satisfies all phenomena (the Philosophers universall Topick,) but evinced by ye mediation of experiments concluding directly & without any suspicion of doubt. To continue the historickall narration of these experiments would make a discourse too tedious & confused, & therefore I shall rather lay down the Doctrine first, and then, for its examination, give you an instance or two of the Experiments, as a specimen of the rest.

- **Hint to Problem 9** Newton grants very high epistemic status to his theory of colours. Claiming that it is a “rigid consequence” implies that it follows unquestionably from previously established knowledge. Newton therefore did not think of differential refrangibility as a hypothesis, but saw it as an established fact. It is seen as a newly discovered property of light rays. It belongs to a group of other properties, like extension, shape, etc., called “primary qualities” by Boyle (similar to “common sensibles” see Chapter IV). But to establish this fact Newton accepted the received laws of geometrical optics. These laws (hypothetical in nature, as all scientific laws are) were thus considered to have the same epistemic status as “brute facts”.

It is also problematic how the theory of colours follows from the differential refrangibility. Newton only claimed that “light, tending to that end of the Image, towards which the refraction of the first Prisme was made, did in the second Prisme suffer a Refraction considerably greater than the light tending to the other end”. All this description states is that refraction is conservative: if ray R1 has greater
refraction at the first prism than R2, then R1 will have greater refraction at the second prism as well. It does not even follow that the refraction of R1 is the same at the two prisms.

In the reformulation of the conclusion he asserts that “every particular Ray hath to suffer a particular degree of Refraction”. This is a much stronger claim than the original, but it is neither measured, nor argued for in the previous part of the letter. To turn the science of colours mathematical, however, an even stronger and more specific claim has to be established, namely that not only is refrangibility a stable property of light rays, but there is also a one to one correspondence between this property and the colours of the rays.

THE FIRST THREE PROPOSITIONS

Newton carried out a series of physical, psychological, and even physiological experiments studying the nature of light and colours. He only summarizes a small subset of his findings in this letter, which, without the experimental support, gave rise to objections and criticism. He outlined his theory of colours in a series of propositions.

In the first three propositions the basis of the doctrine is presented. Newton disqualifies all modificationist theories claiming that specific refrangibility and colour are both connate (innate) properties of the light rays. Both properties can be used to determine the other, as they are both immutable.

The Doctrine you will find comprehended and illustrated in the following propositions.

1. As the Rays of light differ in degrees of Refrangibility, so they also differ in their disposition to exhibit this or that particular colour. Colours are not Qualifications of Light, derived from Refractions, or Reflections of natural Bodies (as ‘tis generally believed,) but Original and connate properties, which in divers Rays are divers. Some Rays are disposed to exhibit a red colour and no other; some a yellow and no other, some a green and no other, and so of the rest. Nor are there only Rays proper and particular to the more eminent colours, but even to all their intermediate gradations.

2. To the same degree of Refrangibility ever belongs the same colour, and to the same colour ever belongs the same degree of refrangibility. The least Refrangible Rays are all disposed to exhibit a Red colour, and contrarily those Rays, which are disposed to exhibit a Red colour, are all the least refrangible: So the most refrangible Rays are all disposed to exhibit a deep Violet colour, and contrarily those which are apt to exhibit such a violet colour, are all the
most Refrangible. And so to all the intermediate colours in a continued series belong intermediate degrees of refrangibility. And this Analogy ‘twixt colours, and refrangibility, is very precise and strict; The Rays always either exactly agreeing in both, or proportionally disagreeing in both.

3. The species of colour, and degree of Refrangibility proper to any particular sort of Rays, is not mutable by Refraction, nor by Reflection from natural bodies, nor by any other cause, that I could yet observe. When any one sort of Rays hath been well parted from those of other kinds, it hath afterwards obstinately retained its colour, notwithstanding my utmost endeavours to change it. I have refracted it with Prisms and reflected it with Bodies, which in Day-light were of other colours; I have intercepted it with the coloured film of Air interceding two compressed plates of glass; transmitted it through coloured Mediums, and through Mediums irradiated with other sort of Rays, and diversly terminated it, and yet could never produce any new colour out of it. It would by contracting or dilating become more brisk, or faint, and by the loss of many Rays, in some cases very obscure and dark; but I could never see it changed in specie.

**Primary and Compound Colours**

The second proposition states that to one specific colour (a colour *species*) a specific refrangibility belongs. In discussing mixtures, however, Newton has to lax this condition (see Prop. 6). He gives a mechanistic model of colour mixing, similar to the one already seen in Chapter IV.4 from Aristotle. This model is corpuscular, and is based on *transduction*, a common seventeenth century method “by which the laws and properties of macroscopic bodies are extended to the imperceptible microscopic parts of bodies” (Shapiro 1993, ch. 3, esp. pp. 40-48).

- **Problem 10** How many primary colours does Newton distinguish?

4. Yet seeming transmutations of Colours may be made where there is any mixture of divers sorts of Rays. For in such mixtures, the component colours appear not, but, by their mutual allaying each other, constitute a midling colour. And therefore, if by refraction, or any other of the aforesaid causes, the difform Rays, latent in such a mixture, be separated, there shall emerge colours different from the colour of the composition. Which colours are not New generated, but only made Apparent by being parted; for if they be again intirely mix’t and blended together, they will again compose that colour, which they did before separation. And for the same reason, Transmutations made by the convening
of divers colours are not real; for when the difform Rays are again severed, they will exhibit the very same colours, which they did before they entered the composition; as you see, Blew and Yellow powders, when finely mixed, appear to the naked eye Green, and yet the colours of the Component corpuscles are not thereby really transmuted, but only blended. For, when viewed with a good Microscope, they still appear Blew and Yellow interspersedly.

5. There are therefore two sorts of colours. The one original and simple, the other compounded of these. The Original or primary colours are, Red, Yellow, Green, Blew, and a Violet-purple, together with Orange, Indico, and an indefinite variety of Intermediate gradations.

6. The same colours in Specie with these Primary ones may be also produced by composition: For, a mixture of Yellow and Blew makes Green; of Red and Yellow makes Orange; of Orange and Yellowish green makes yellow. And in general, if any two Colours be mixed, which in the series of those, generated by the Prisme, are not too far distant one from another, they by their mutual alloy compound that colour, which in the said series appeareth in the mid-way between them. But those, which are situated at too great a distance, do not so. Orange and Indico produce not the intermediate Green, nor Scarlet and Green the intermediate yellow.

- **Hint to Problem 10** Newton gives three answers. The five colours first mentioned are the most visible. He adds two, and in later writings argues that these seven colours are arranged according to the tones of the musical octave (Gouk 1988). In this he follows the Pythagorean tradition also witnessed in Aristotle’s writings. In a 1675 letter he published a diagram much like figure III.1 on page 50, but instead of writing ratios, he added the colour names and their musical equivalents. From AG to FM the dotted lines were labelled *Sol, La, Fa, Sol, La, Mi, Fa Sol*, while the regions between the dotted lines were bands of colour *Purple, Indigo, Blue, Green, Yellow, Orange, Red*. Finally, as his Prop. 2. necessitates, he states that the number of primary colours is indefinite.

Although to our ears this last answer sounds correct, it was heavily criticised by contemporaries. At the time most painters held that there are three basic colours (some added black and white), and all colours can be derived from these. To claim that there is an infinite number seemed to run counter to the practice of painting. Scientists, like Hooke and Huygens attacked Newton that this is against
the principle of economy—why multiply entities without necessity? Most contemporary modificationist theories postulated the existence of only two basic colours.

Even in his *Opticks* Newton maintained the strict analogy between musical tones and colours. This greatly supported the view that the number of basic colours is seven (Shapiro 1994). In the popularization of the Newtonian theory usually seven colours were depicted and discussed—a regularity even visible in eighteenth century English poetry (Nicolson 1963).

**Whiteness**

For the contemporary theories colours were *explananda*, the entities that had to be accounted for. White light, on the other hand, was the *explanans*, the principle used for the explanation. Newton’s theory radically inverted the roles. White light became the *explanandum* and it was explained with recourse to a number of more basic units, the coloured lights (note the plural) “hidden” in white light.

- **Problem 11** How does Newton support his claim that whiteness is the composition of all the primary colours?

  7. But the most surprising and wonderful composition was that of Whiteness. There is no one sort of Rays which alone can exhibit this. ‘Tis ever compounded, and to its composition are requisite all the aforesaid primary Colours, mixed in a due proportion. I have often with Admiration beheld, that all the Colours of the Prisme being made to converge, and thereby to be again mixed as they were in the light before it was Incident upon the Prisme, reproduced light, entirely and perfectly white, and not at all sensibly differing from the direct Light of the Sun, unless when the glasses, I used, were not sufficiently clear; for then they would a little incline it to their colour.

  8. Hence therefore it comes to pass, that Whiteness is the usual colour of Light; for, Light is a confused aggregate of Rays indued with all sorts of Colors, as they are promiscuously darted from the various parts of luminous bodies. And of such a confused aggregate, as I said, is generated Whiteness, if there be a due proportion of the Ingredients; but if any one predominate, the Light must incline to that colour; as it happens in the Blew flame of Brimstone; the yellow flame of a Candle; and the various colours of the Fixed stars.
• **Hint to Problem 11** If all colours are mixed, and one predominates, then the resulting light will incline to this colour (Prop. 8). To mix all colours is sufficient to create white light, but Newton nowhere supports the claim that it is necessary to use all colours. Huygens in a letter claimed that two colours, blue and yellow, mixed in the right proportions are also sufficient to produce white light. Although Newton rejected the possibility, in his later writings he restricted his claim to the white light of the Sun (Shapiro 1980).

**COLOUR PHENOMENA**

The last four propositions offer an explanation of both emphatical and real colours. Starting with the prismatic colours, Newton shows how his theory can explain the rainbow, the colour of liquids, and, finally, the colour of solid bodies. The very concise formulation is aimed at showing the fruitfulness of the “New Theory”.

A few substances appear in the text that are little known. *Lignum nephriticum*, or nephritic wood is a wood of which the infusion was formerly used as a remedy in diseases of the kidney. It probably refers to the horseradish tree (*Moringa pterygosperma*). It was noticed early on that the decoction changes colour according to the different lights it was viewed in.

*Minium*, a colour similar to vermilion, is the colour of red lead.

*Bise* is short for blue bice and indicates a shade of blue obtained from smalt. It is duller than ultramarine or azure.

• **Problem 12** To determine the colours of liquids, and solid bodies, Newton relies on a one to one correspondence between colour and a specific physical property. What is exactly this physical property? According to this conception “where are the colours”, i.e. what substances have the stable property to be coloured this way or that way?

• **Problem 13** From the Newtonian propositions which ones would Robert Hooke accept and which ones reject (see also the previous chapter)?

9. These things considered, the manner, how colours are produced by the Prisme, is evident. For, of the Rays, constituting the incident light, since those which differ in Colour proportionally differ in refrangibility, they by their unequal refractions must be severed and dispersed into an oblong form in an
orderly succession from the least refracted Scarlet to the most refracted Violet. And for the same reason it is, that objects, when looked upon through a Prisme, appear coloured. For, the difform Rays, by their unequal Refractions, are made to diverge towards several parts of the Retina, and there express the Images of things coloured, as in the former case they did the Suns Image upon a wall. And by this inequality of refractions they become not only coloured, but also very confused and indistinct.

10. Why the Colours of the Rainbow appear in falling drops of Rain, is also from hence evident. For, those drops, which refract the Rays, disposed to appear purple, in greatest quantity to the Spectators eye, refract the Rays of other sorts so much less, as to make them pass beside it; and such are the drops in the inside of the Primary Bow, and on the outside of the Second or Exteriour one. So those drops, which refract in greatest plenty the Rays, apt to appear red, toward the Spectators eye, refract those of other sorts so much more, as to make them pass beside it; and such are the drops on the exterior part of the Primary, and interior part of the Secondary Bow.

11. The odd Phenomena of an infusion of Lignum Nephriticum, Leaf gold, Fragments of coloured glass, and some other transparently coloured bodies, appearing in one position of one colour, and of another in another, are on these grounds no longer riddles. For, those are substances apt to reflect one sort of light and transmit another; as may be seen in a dark room, by illuminating them with similar or uncompounded light. For then they appear of that colour only, with which they are illuminated, but yet in one position more vivid and luminous than in another, accordingly as they are disposed more or less to reflect or transmit the incident colour.

12. From hence also is manifest the reason of an unexpected Experiment, which Mr. Hook somewhere in his Micrographia relates to have made with two wedg-like transparent vessels, fill’d the one with a red, the other with a blew liquor: namely, that though they were severally transparent enough, yet both together became opake; For, if one transmitted only red, and the other only blew, no rays could pass through both.

13. I might add more instances of this nature, but I shall conclude with this general one, that the Colours of all natural Bodies have no other origin than this, that they are variously qualified to reflect one sort of light in greater plenty then another. And this I have experimented in a dark Room by illuminating those bodies with uncompounded light of divers colours. For by that means any body may be made to appear of any colour. They have there no appropriate colour, but ever appear of the colour of the light cast upon them, but yet with this difference, that they are most brisk and vivid in the light of their own day-light colour. Minium appeareth there of any colour indifferently, with which
'tis illustrated, but yet most luminous in red, and so Bise appeareth indifferently of any colour with which 'tis illustrated, but yet most luminous in blew. And therefore Minium reflecteth Rays of any colour, but most copiously those indued with red; and consequently when illustrated with day-light, that is, with all sorts of Rays promiscuously blended, those qualified with red shall abound most in the reflected light, and by their prevalence cause it to appear of that colour. And for the same reason Bise, reflecting blew most copiously, shall appear blew by the excess of those Rays in its reflected light; and the like of other bodies. And that this is the intire and adequate cause of their colours, is manifest, because they have no power to change or alter the colour of any sort of Rays incident apart, but put on all colours indifferently, with which they are enlightened.

- **Hint to Problem 12** Refrangibility, a physical property of light rays is connected to the secondary quality colour. Colour is thus a property of light rays, and if the refrangibility of the rays is a stable property, then so is their colour. The colours of bodies are thus accidental: depending on the light source they reflect differently coloured rays in different proportions. This view challenges our very basic notion that the colours of objects are stable properties.

- **Hint to Problem 13** Hooke's first response to Newton's theory, written in a few hours, was read out at the Royal Society on 15 February, only a week after Newton's letter. He accepted proposition one that colours are connate properties of light, but only for blue and red rays, as this suited his theory. Similarly, propositions 2, 3, 6, and 9 to 11 were also taken for granted. His main objection to the fifth proposition was that positing an indefinite variety of “primary or originall colours” (Turnbull 1959) makes no sense as it is “wholly useless to multiply entities without necessity”, when his two colours are sufficient to explain the others. He similarly rejected proposition 7 (white is only one motion, colour is a second motion), and 8.

### 2.5 Conclusions

After the thirteen propositions, Newton draws more general philosophical conclusions. He seems to have trust in the truth of the propositions.
• **Problem 14** Reconstruct the Peripatetic position Newton is attacking from his diametrically opposite statements. Which of the statements do you think are trivially true or false, and which are strongly supported by the letter? (The Hint provided is only for the first part of the question.)

These things being so, it can be no longer disputed, whether there be colours in the dark, nor whether they be the qualities of the objects we see, nor perhaps, whether Light be a Body. For, since Colours are the qualities of Light, having its Rays for their intire and immediate subject, how can we think those Rays qualities also, unless one quality may be the subject of and sustain another; which in effect is to call it substance. We should not know Bodies for substances, were it not for their sensible qualities, and the Principal of those being now found due to something else, we have as good reason to believe that to be a substance also.

Besides, who ever thought any quality to be a heterogeneous aggregate, such as Light is discovered to be. But, to determine more absolutely, what Light is, after what manner refracted, and by what modes or actions it produceth in our minds the Phantasms of Colours, is not so easie. And I shall not mingle conjectures with certainties.

• **Hint to Problem 14** For Newton there are no colours in the dark, but that this is diametrically opposed to the Peripatetic doctrine is problematic. Aristotle in *De anima* 426a20-26 writes (trans. D. W. Hamlyn):

But the earlier philosophers of nature did not state the matter well, thinking that there is without sight nothing white nor black, nor flavour without taste. For in one way they were right but in another wrong; for since perception and the object of perception are so spoken of in two ways, as potential and actual, the statement holds of the latter, but it does not hold of the former. But they spoke undiscriminatingly concerning things which are so spoken of not undiscriminatingly.

Newton dismisses other (cavalierly reconstructed) elements of the Peripatetic teaching and in his argument uses a Peripatetic technique to prove his point.
2.6 THE LAST EXPERIMENT

Newton suggests that experiments be made to test his new theory of colours. As a conclusion he describes concluding experiments and draws attention to some of the experimental difficulties. This is in line with his earlier, Baconian approach, but is a rather dangerous move. He only gave the outline of a complex theory and simply doing experiments (the word experimentum has only recently separated in meaning from the word experimentia, meaning experience) can easily produce results seemingly counter to the theory.

And this is what happened. A number of people rejected Newton’s theory just because they carried out experiments, and the Newtonian theory did not obviously fit the results.

- **Problem 15** Discuss what colour concept is implicit in the paragraph on the importance of changing uncompound colour. (No hint is provided for this question.)

Reviewing what I have written, I see the discourse it self will lead to divers Experiments sufficient for its examination: And therefore I shall not trouble you further, than to describe one of those, which I have already insinuated.

In a darkened Room make a hole in the shut of a window, whose diameter may conveniently be about a third part of an inch, to admit a convenient quantity of the Sun’s light: And there place a clear and colourless Prisme, to refract the entring light towards the further part of the Room, which, as I said, will thereby be diffused into an oblong coloured Image. Then place a Lens of about three foot radius (suppose a broad Object-glass of a three foot Telescope,) at the distance of about four or five foot from thence, through which all those colours may at once be transmitted, and made by its Refraction to convene at a further distance of about ten or twelve foot. If at that distance you intercept this light with a sheet of white paper, you will see the colours converted into whiteness again by being mingled. But it is requisite, that the Prisme and Lens be placed steddy, and that the paper, on which the colours are cast, be moved to and fro; for, by such motion, you will not only find, at what distance the whiteness is most perfect, but also see, how the colours gradually convene, and vanish into whiteness, and afterwards having crossed one another in that place where they compound Whiteness, are again dissipated, and severed, and in an inverted order retain the same colours, which they had before they entered the composition. You may also see, that, if any of the Colours at the Lens be intercepted, the Whiteness will be changed into the other colours. And therefore,
that the composition of whiteness be perfect, care must be taken, that none of
the colours fall beside the Lens.

In the annexed design of this Experiment, A B C representeth the Prism set
dewise to sight, close by the hole F of the window E G. Its vertical Angle A
C B may conveniently be about 60 degrees: M N designes the Lens. Its breadth
2 1/2 or 3 inches. SF one of the streight lines, in which difform Rays may be
conceived to flow successively from the Sun. F P and F R two of those Rays
unequally refracted, which the Lens makes to converge towards Q, and after
decussation to diverge again. And HI the paper, at divers distances, on which
the colours are projected: which in Q constitute Whiteness, but are Red and
Yellow in R, r, and ρ, and Blew and Purple in P, p, and π.

If you proceed further to try the impossibility of changing any uncompoun-
ded colour (which I have asserted in the third and thirteenth Propositions,) 'tis
requisite that the Room be made very dark, least any scattering light, mixing
with the colour, disturb and allay it, and render it compound, contrary to the
design of the Experiment. 'Tis also requisite, that there be a perfter separation
of the Colours, than, after the manner above described, can be made by the
Refraction of one single Prisme, and how to make such further separations,
will scarce be difficult to them, that consider the discovered laws of Refractions.
But if tryal shall be made with colours not throughly separated, there must be
allowed changes proportionable to the mixture. Thus if compound Yellow light
fall upon Blew Bise, the Bise will not appear perfectly yellow, but rather green,
because there are in the yellow mixture many rays indued with green, and Green
being less remote from the usual blew colour of Bise than yellow, is the more
copiously reflected by it.

In like manner, if any one of the Prismatick colours, suppose Red, be inter-
cepted, on design to try the asserted impossibility of reproducing that Colour
out of the others, which are pretermitted: 'tis necessary, either that the colours
be very well parted before the red be intercepted, or that together with the red
the neighbouring colours, into which any red is secretly dispersed, (that is, the
yellow, and perhaps green too) be intercepted, or else, that allowance be made
for the emerging of so much red out of the yellow & green, as may possibly have
been diffused, and scatteringly blended in those colours. And if these things be
observed, the new Production of Red, or any intercepted colour will be found impossible.

This, I conceive, is enough for an Introduction to Experiments of this kind; which if any of the R. Society shall be so curious as to prosecute, I should be very glad to be informed with what success: That, if any thing seem to be defective, or to thwart this relation, I may have an opportunity of giving further direction about it, or of acknowledging my errors, if I have committed any.

Your humble Servt
ISAAC NEWTON

3 FURTHER PROBLEMS

- **Problem 16** In spite of the obvious differences, Hooke’s theory in his *Micrographia* shows a number of similarities to Newton’s 1672 letter. While attacking Hooke’s position, Newton also learnt a lot from his older rival. Pinpoint these similarities (not forgetting the differences) in argumentation, in the materials used and experiments carried out, in the treatment of physical and physiological considerations, in subjective and objective prismatic experiments, and in some cases even in the surprisingly similar phrases used.

- **Problem 17** Make a close reading of the correspondence between a) Hooke and Newton, and b) Anthony Lucas and Newton. The letters can be found in the first volume of Newton’s correspondence (Turnbull 1959). Use the suggested literature for help.

4 FURTHER READINGS

Only a short sample will be given of the immense literature. Most of Newton’s optical writings can now be read in print. For the earliest notebooks see (McGuire & Tamny 1983). For the mathematical notebooks with a number of geometrical optical problems (Whiteside 1966). His notes on reading Hooke’s *Micrographia*, an important source in his early development in optics is reprinted in (Hall & Hall 1978), a book with a self-contradictory title. His letters to the Royal Society and most of his debates are reprinted in the volumes of his correspondence (Turnbull 1959). Newton's optical lectures from Cambridge exist in two versions. Both are printed with introductions and translations by Alan Shapiro in (Newton 1984). The most accessible version of his *Opticks* is the reprint of the fourth edition in (Newton 1952).
For overviews of Newton’s scientific work consult recent collections like (Buchwald & Cohen 2001), an excellent handbook, (Gjertsen 1986), or an exhaustive biography, (Westfall 1980). The development of the mathematical sciences had been decisive (Dear 1995)—see also (Dear 2000)—just as the broader cultural setting (Dobbs & Jacob 1995) for Newton’s development.

The 1672 “New Theory” is analysed among others in (Sepper 1988) and the guided study of Newton’s writings in (Sepper 1994). For a classical study see (Sabra 1967). The experimentum crucis is analysed in (Fehér 1995) just as in (Lohne 1968b) and (Lohne 1968a). Other articles include (Laymon 1978), (Steinle 1993a), or (Bechler 1974), and more recent contributions, like (Steinle 2001) on Newton’s colour theory and perception or (Worrall 2000) on the Newtonian method of argumentation. See also the first part of (Zemplén 2001) or (Zemplén 2004).

For contemporary works, see the lectures of Newton’s predecessor (Barrow 1987), or the natural history of colour phenomena in (Boyle 1664), the extremely popular treatise on microscopic phenomena (Hooke 1665), or the best treatise on the contemporary wave-theory (Huygens 1790 (1962)).
Chapter X

THE MOON ILLUSION

1 INTRODUCTION

A very long standing and yet unsolved puzzle in the History of Science is the “moon illusion”. The written record about the enlargement of the horizon moon can be traced back to the seventh century B.C.—to a cuneiform script on a clay tablet from Niniveh (Plug & Ross 1989)—, and even today there is no agreement about the reason of the curious phenomenon.

The near-horizon enlargement is visible for all heavenly bodies, and even for the distances between the stars of a constellation. An early recognition of the phenomenon can be seen in the writings of Aristotle: “promontories in the sea ‘loom’ when there is a south-east wind, and everything seems bigger, and in a mist, too, things seem bigger: so, too, the sun and the stars seem bigger when rising and setting than on the meridian” (373b)\(^1\). Since the seventeenth century the phenomenon is normally referred to as the “moon illusion” as it is most noticeable and conspicuous for the moon. It is considered to be an illusion as it is generally accepted that the moon is not increased in actual size (distal size) when near the horizon.

Today the illusion is taken to have the above mentioned primary aspect: the size difference between the zenith and the horizon moon. The secondary, distance aspect of the illusion—that most people see the moon as closer to us when near the horizon—is also illusory, in fact it is the zenith moon that is somewhat nearer to the observer.

This problem was one of the standard puzzles that ancient and medieval theories attempted to solve—treating it either as a problem of physics, meteorology, optics, psychology, or physiology. The explanations often operated with the concept of the visual ray, and its weakening as it has to pass through media. The enlargement was mostly ascribed to refraction or reflection in atmospheric vapours. In these theories the moon really

\(^1\)This first known explanation is found in Aristotle’s *Meteorologica*, in Text 1 of Chapter 5.1, for the internet link see page 72. A recent bibliography collects 285 publications from antiquity (Plug 1989).
subtended a greater visual angle at the horizon i.e. its proximal size was greater, thus no psychological or physiological explanation was required.

An upsurge of interest in the seventeenth century was followed by relative neglect, until interest was revived in the second half of the nineteenth century. As more and more exact measurements lead to the rejection of the refraction theory in the seventeenth century, more and more attention was paid to psychological factors. Several curious observations were made, many contradictory.

In the twentieth century the experiments became more and more systematic, but as yet these were unable to settle the question. Since the influential work of Kaufmann and Rock (Kaufman & Rock 1962, Rock & Kaufman 1962) the received (or rather most popular) view is a ‘taking account of distance’ (TAD) theory, but several rival explanations were suggested. None of the current explanations are generally accepted, and the “several dozen perceptual theorists presently writing about the appearance of the moon can barely be categorized into groups larger than two persons each to represent common beliefs about the cause of the moon illusion” (Haber & Levin 1989).

The differing views point to conceptual problems within contemporary theories of size and distance perception. While no common ground is to be expected in the future, discussion on the illusion can highlight problematic and unfounded aspects of today’s theories.

2 Main Types of the Explanations of the Illusion

2.1 Classical Science and the Illusion: Ptolemy and the Aristotelian Vapour Theory

Ptolemy refers to the moon illusion three times in his extant works: in the Mathematical Composition (also referred to as the Almagest), in the Planetary Hypotheses, and in the Optics. As will be seen, there is a clear development of his understanding, from his earlier physical explanation to his mature psychological account.

\footnote{Castelli in 1639 observed that occluding the intervening far-away objects with the brim of his hat greatly reduces the size of the illusion (Ariotti 1973). Molyneux reported that the horizon moon is enlarged even if viewed through a cube (Molyneux 1687). Gauss in a letter to Bessel on 9th April 1830 reported that viewing the elevated full moon seen supine, in a backward reclining position seems bigger, while the horizon moon looks smaller with the head inclined forwards (Gauss 1880). See also (Wade 1998).}
Text 1 contains the refraction theory in Ptolemy’s influential book the *Almagest*. The refraction theory was one of the most popular explanations of the enlargement of the horizon moon and is still often accepted by laymen, even though astronomers have rejected it in the seventeenth century. As its tenability was questioned, the Aristotelian vapour theory was revived, as in the works of Thomas Hobbes (Hobbes 1839). Text 2 is a part of an article from 1762 by Samuel Dunn. It provides experimental support for Ptolemy’s theory. It also mentions the secondary aspect of the illusion: the horizon moon looks enlarged but also nearer to us than the zenith moon.

- **Problem 1** Reconstruct the argument of Text 1. Evaluate the analogy Ptolemy uses: the moon is enlarged just as objects placed in water are enlarged.

- **Problem 2** Text 2 provides empirical support for the thesis that the illusion is the result of atmospheric refraction. Point to similarities and differences of Text 1 and this account. How is the enlargement explained (§23)?

Text 3 is a physiological explanation of the illusion from Ptolemy’s *Optics*. Here, as opposed to the previous accounts, the size of the zenith moon is illusory.

2.2 Psychological Explanations in the Middle Ages: Alhazen

Ibn al-Haytham’s, or Alhazen’s book on optics was one of the major influences on the medieval optical tradition in the West, including the works of Roger Bacon (ca. 1263), John Pecham (ca. 1275), and Witelo (ca. 1273). In his book Alhazen suggested a radically new theory to explain the illusion. It came to be known as the ‘intervening objects’ theory. On this account the different apparent sizes of the moon are due to psychological mechanisms. Text 4 is from Book VII of Alhazen’s *Optics*.

- **Problem 3** Alhazen’s argument gives a permanent cause and an accidental cause of the horizon-enlargement of heavenly bodies. Summarize the two. How does Alhazen account for the secondary (distance) aspect of the illusion?
2.3 Natural Geometry or Contingent Cues

With the development of theories of perception the moon illusion played an ever-increasing role. Accepted as a psychological illusion most theories that attempted to account for size and distance perception attempted to incorporate the moon illusion. Although the enlargement of the moon is illusory, as it reflects no actual changes in size, from a psychological point of view it has to be the result of the lawful operation of our perceptual apparatus.

Text 5 is from Descartes’s *Dioptrics*, a short description of how the visual system calculates distance.

Text 6 comprises extracts from Berkeley’s *New Theory of Vision*.

- **Problem 4** Describe how Berkeley evades the necessity to include computational mechanisms for distance perception. How does he explain the moon illusion? Separate the distal cue (the account of the necessary viewing conditions for the illusion to occur) and the internal processes that are activated when the distal cues are present.
2.4 Further-Larger-Nearer Theories

A major reason why even today no accepted explanation of the moon illusion exists is that it seems to run counter to one of the generally accepted tenets of visual space perception. The ‘size-distance invariance hypothesis’ (SDIH) states that there is a geometrical relationship between the size of objects ($S$), their distance ($D$), and the angle they subtend in the field of vision (the visual angle, or $\alpha$):

$$\tan \alpha = \frac{S}{D} \quad \text{(X.1)}$$

The visual angle and the perceived distance ($d$) determines the perceived size ($s$) of an object, while the visual angle and the size determine the perceived distance:

$$s = \tan \alpha \times d \quad \text{(X.2)}$$

$$d = \frac{\tan \alpha}{s} \quad \text{(X.3)}$$

The existence of this connection between perceived size and distance is supported by experimental data. The perceived size of a stimulus of constant visual angle (e.g. a strong afterimage) changes with the perceived distance of the stimulus (whether we look at a nearby or faraway surface). This regularity is known as Emmert's law. The horizon moon is, however, not only reported by most observers as bigger, but also as nearer to us, thus providing a paradigm example of the so called size-distance paradox.

In the last four decades different researchers have given different answers as to how to modify the standard SDIH to incorporate the data from experiments studying the moon illusion. The other possibility is to reject the SDIH—but this runs counter to many of our most basic intuitions (and experiments) about size constancy.

One possibility is to claim that the reported distance of the moon is different from the perceived distance of the moon. Verbal reports of distance have been shown to differ from the “true” perceived distance obtained through psychophysical measurements (Gogel 1976). In this conception perceived distance (considered as “true” (Gogel & Mertz 1989)) is the input of the SDIH equation. The size-distance paradox is explained “as a consequence of cognitive contamination of distance report” (Hershenson 1989).

Other models either alter the input of the equation, add new variables, or propose radically different mechanisms for determining the size
and distance of objects. Some of these models are physiological and stress the importance of factors like accommodation convergence. Roscoe proposed that—because for the zenith moon there are no distance cues—accommodation approaches a neutral state (the near-sighted, *myopic* state of the dark focus). The visible texture near the horizon moon, however, results in outward accommodation. With this change of accommodation the retinal size of the two moons subtending the same angle changes (Roscoe 1989), the horizon moon being significantly larger.

Some researchers alter the conception of the visual angle (McCready 1986), or add new variables to equation X.2.

It is highly probable that more than one mechanism is responsible for the natural moon illusion (Reed 1989), or (McCready 1986). Some contemporary theories are distant relatives of the angle-of-regard theories seen in Texts 3 and 6. Enright, for example, states that by looking up the extraocular muscles automatically diverge the eyes (two of the six pairs of muscles that move the eyeball are responsible for torsion). To compensate for the outward movement, the eyes automatically converge (Enright 1989). Convergence is usually only needed for binocular viewing of close objects – thus it is associated with small distances, and small sizes (convergence micropsia). Thus vergence becomes one of the factors that can affect the apparent size of the moon. Other physiological factors might include the pupil size: with the opening of the pupil the depth of visual field decreases, and this results in outward accommodation from the myopic resting state of the lens.

Text 7 contains the two articles that even now shape the common opinion about the moon illusion. Irvin Rock was one of the most important proponents of a constructivist view of vision (Rock 1975, Rock 1983, Rock 1997), holding that perception is intelligent in that it is based on operations similar to those that characterize thought. Their theory is a TAD theory, in strong contrast with Berkeley’s ideas.

---

3That retinal image size changes when the eye misaccommodates (while viewing objects through an artificial pupil, a board some distance from the eye with a small hole on it) has been known for more than a century (Helmholtz 1867, Helmholtz 2000). This factor, however, cannot account for the whole of the illusion, as patients with artificial pupils after cataract operations (see the next chapter) also experience a moon illusion (Kaufman & Rock 1989).
• **Problem 5** A crucial element of the theory is that the SDIH can be maintained (the moon illusion can be considered as a special case of Emmert’s law) and the TAD theory provides the correct explanation “if it is true that the registered distance to the horizon moon is greater than that to the zenith moon” (p. 953). Where and how do the writers argue for this issue? Separate the arguments that support SDIH and those that argue for the existence of TAD mechanisms.

3 Texts

3.1 Text 1: Ptolemy’s *Almagest*

- The short text from Book I Chapter 3 of the *Almagest* is from Toomer’s translation (Toomer 1984, p. 39).

  ...

  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 variation occurs. For the apparent increase in their sizes at the horizons is caused, not by a decrease in their distances, but by the exhalation of moisture surrounding the earth being interposed between the place from which we observe and the heavenly bodies, just as objects placed in water appear bigger than they are, and the lower they sink, the bigger they appear.

- **Hint to Problem 1** Ptolemy explains the variation of the sizes of heavenly bodies that are not caused by the variation of their distances. The analogy with refraction in water has lead to many speculations, as it is incorrect in an unqualified form. When watching the enlargement of objects in water we are in the rarer medium, but when watching heavenly bodies, we are in a denser medium: the horizon moon should thus appear smaller.
Alhazen in his *Commentary to the Almagest* provides a detailed analysis of the problem showing that with the interposition of very dense vapour (exhalation of moisture) between the ether (rare medium) and the air (denser medium) the apparent size of celestial magnitude (stars, distance between stars) may be increased. Such qualifications of the analogy make it tenable, but it is more probable that at the time of composition Ptolemy only had limited knowledge of refraction, and thus the *Almagest* was probably written before the *Optics*. See more in (Sabra 1987).

### 3.2 TEXT 2: SAMUEL DUNN, 1762

- Samuel Dunn, when first observing the enlargement of the sun and the moon thought it is a result of “prejudice and an imposition” on his sight and judgement. But the astronomical measurements have shown that the sun is a spheroid near the horizon: the vertical diameter is smaller than the horizontal. Nearing the horizon the increase is quite sudden and many other features of the sun are changed—Dunn claims that the reason is the increasingly thick atmosphere between observer and object. He carries out experiments which are included here not because they were highly influential, but because they support Ptolemy’s theory (much outdated by the eighteenth century). Dunn is also one of the first theorists trying to account for the secondary, apparent distance aspect of the moon illusion. The article is from the *Philosophical Transactions*, 1762 (Dunn 1762).
LXXII. An Attempt to assign the Cause, why the Sun and Moon appear to the naked Eye larger when they are near the Horizon. With an Account of several natural Phenomena, relative to this Subject. By Mr. Samuel Dunn.

Chelsea, January 27, 1762.

Read Feb. 11, 1762.

The Sun and Moon, when they are in or near the horizon, appear to the naked eye of the generality of persons, so very large in comparison with their apparent magnitudes, when they are in the zenith, or somewhat elevated, that several learned men have been led to enquire into the cause of this phenomenon, and, after endeavouring to find certain reasons, founded on the principles of physics, they have at last pronounced this phenomenon as a mere optical illusion.

15. I took a cylindrical glass vessel, about two feet high, and having graduated its sides to inches, I placed it upright on a table, with a piece of paper under the bottom of the glass, on which paper were drawn parallel right lines, at a proper distance from each other; and having placed a shilling at the bottom of the vessel, it was nearly as low as the paper. Pouring water into the vessel, and viewing the shilling through the medium of water, with one eye, whilst I beheld with the other eye, where the edges of the shilling were projected on the paper, and its parallels, I found the shilling appear larger, at every additional inch depth of the water; and this was the case, if either eye was used; and the same, when the eye was removed far from the surface, or near to it, or in any position thereto.

16. I took large vessels, filled them with water, placed different bodies at the bottoms of those vessels.
It always followed, that the greater depth of the water I looked through, in the direction from my eye, to the objects in the water, the nearer those objects appeared to me. Thus light bodies appeared more mellow and faint, and dark bodies rather better defined, than out of the water, when they were not deeply immersed. And thus they appeared under whatever directions or positions I viewed the bodies.

17. I placed different bodies in proper vessels of fair water, and immersed my face in the water; viewing the bodies in and through the water, they all appeared to me plain, when not too far from the eye, and although a little hazy at the edges, they appeared much enlarged, and always larger through a greater depth of water. Thus, a shilling appeared nearly as large as half a crown, with a red glowing arch on that side opposite to the Sun, when the Sun shined on the water. From this experiment, I concluded, that divers see light objects not only larger, but very distinctly in the water.

18. These, and several other circumstances, being considered, they left me with but little doubt, whether the atmosphere refracts horizontally or not, as the afore-mentioned protuberances in the Sun's limb must have been wholly owing to such a cause, and the nearly allied strata in the atmosphere. That the apparently formed mountains of trees and bushes at sunrising, so easily comparable with other trees and bushes of equal magnitude at other times, but in their affected state as much larger, must also be owing to the same cause. That seeing the nature and properties of those strata of horizontal vapours cannot differ from each other, as much as the whole quantity of
medium, through which a ray of light passeth in coming from the horizontal Sun, differs from the whole quantity of medium, through which another ray of light passeth in coming from the vertical Sun; and the latter experiment being so consonant to the manner in which the rays of light pass, through the medium of the atmosphere, into the eye; I therefore concluded, that these were proofs, that objects seen through a medium of greater depth, or density, do appear more large; and that therefore, not only the Sun and Moon, but that all other objects, seen at great distances, under an horizontal direction, do appear larger to the naked eye, than objects of equal magnitude and distance do appear, when seen under a vertical direction.

23. For the reasons which have been communicated, it has appeared to me, that the Sun and Moon, and distances of stars near the horizon, do appear enlarged to the naked eye; because they then appear nearer to us. That they then appear nearer to us, and more faint; because then their rays pass through a greater length of the atmosphere and horizontal vapours. That in so passing through a greater length of the atmosphere and horizontal vapours, those rays are so reflected, refracted, inflected, attracted, refilled, accelerated, or retarded, so as to become more divergent than they otherwise would, at their entrance into the eye. That as a consequence hereof, that part of the sky which is near the horizon, appears nearer to us than that part near the zenith. That the apparent nearness of the sky near the horizon, is always in a certain, though variable proportion, to the state and properties of the horizontal vapours. That as objects appear larger and nearer, when they are near the horizon, than when they are near the zenith, so they appear larger through the medium of the atmosphere, than they would through a space devoid of air. That the

- **Hint to Problem 2** The explanation of the illusion: for atmospheric reasons the rays from the sun and moon become more di-
vergent in an unspecified way. This results in a decrease in distance. As they are nearer, they appear larger. It is unclear a) what mechanism results in the divergence of rays, b) how does divergence contribute to smaller apparent distance, and c) how does the taking account of distance influence size perception.

3.3 Text 3: Ptolemy’s *Optics*

- The zenith of the science of optics in Antiquity is undoubtedly Ptolemy’s *Optics*. In this work Ptolemy gives a radically different solution to the moon illusion. This type of explanation, operating with the position of the head and the “angle of regard” was rediscovered in the eighteenth century, and even today has some supporters. Although the Greek original was probably still in circulation in the middle ages, the only surviving text of the *Optics* is a twelfth century, badly mangled Latin version of an Arabic translation from a Greek copy, with possibly a Syriac intermediary (Smith 1996). As a result, the first book of *Optics* is missing, and there are several places where the text is corrupted or an absence of text elements (lacunae) can be hypothesized. Book III, §59 also starts with a lacuna in Smith’s new translation (Smith 1996):

... will be concerning distances, insofar as the senses weaken in proportion to the convergence. Generally speaking, in fact, when a visual ray falls upon visible objects in a way other than is inherent to it by nature and custom, it perceives less clearly all the characteristics belonging to them. So too, its perception of the distances it apprehends will be diminished. This seems to be the reason why, among celestial objects that subtend equal visual angles, those that lie near the zenith appear smaller, whereas those that lie near the horizon are seen in another way that accords with custom. Things that are high up seem smaller than usual and are seen with difficulty.

3.4 Text 4: Alhazen’s *Optics*

- Alhazen’s psychological explanation of the illusion is in the last chapter of Book VII in his *Optics*. Sabra’s translation is based on an Arabic manuscript, and expressions in parentheses were in this Arabic manuscript. Angle brackets are additions and corrections using a Latin text, Sabra’s own additions are in square brackets (Sabra 1987).
We say that sight will perceive any star at the zenith to be smaller than in any region of the sky through which the star travels; that the farther the star is from the zenith the larger its magnitude will appear than it does at the zenith: that the star looks largest at the horizon; and that the same is true of the intervals between the stars. Now this is found to be so in fact, namely, that the stars, and their mutual distances, appear to be smaller at the middle of the sky than when they are far from it, and that the star (or interval between two stars) appears largest at the horizon. It remains for us to show the reason why this is so.

We say: It has been shown in Book II of this work, in our discussion of size, that sight perceives size from the magnitudes of the angles subtended at the centre of the eye and from the magnitudes of the distances of the visible objects and from comparing the magnitudes of the angles to those of the distances. We have also shown there that sight neither perceives nor ascertains the magnitudes of the objects’ distances unless these distances extend along near and contiguous bodies; that distances that do not so extend are not ascertainable in magnitude by sight; and that when sight cannot ascertain the magnitudes of the objects’ distances, then it fails to ascertain the objects’ sizes. We showed there, too, that when sight fails to ascertain the distance of an object, then it makes a guess in regard to the distance’s magnitude by likening it to the distances of familiar objects at which it can perceive objects similar to that object in form and figure, then perceives the size of that object from the magnitude of the angle subtended by it at the eye-centre as compared to the distance it has conjectured. But the distances of the stars do not extend along near bodies. Sight does not, therefore, perceive or ascertain their magnitudes, but merely conjectures their magnitudes by assimilating the stars’ distances to the distances of very remote earthly objects which it can perceive and whose magnitudes it conjectures.

Moreover, the body of the heavens is not seen as a sphere whose concavity faces the eye; nor is sight aware of the corporeality of the heavens, and only perceives of the heavens a certain blue colour. As to the heavens’ corporeality, extension in the three <dimensions>, circularity and concavity—these sight has no way of perceiving. And when sight cannot identify something it likens it to one of the familiar objects that resemble it. Thus it perceives the sun and moon as flat, and perceives convex and concave bodies from an excessively great distance as flat, and also perceives arcs whose convexity or concavity faces the eye as straight. For when sight does not perceive [in the case of convex objects] the nearness of their middle points and the remoteness of their extremities, or (in the case of concave objects) the remoteness of their middles and the nearness of their extremities, it likens the convex and concave surfaces to flat surfaces and likens arcs to straight lines, because most familiar objects have flat surfaces and straight edges.

Nor, when the form of a star reaches the eye, is sight aware that it is a refracted form, or that it has been refracted from a concave surface, or that
the body in which the star is rarer than that in which the eye is. Sight rather perceives the form of a star as it perceives the forms that come to it in straight lines from objects located in the air. The forms of visible objects, when they encounter a body whose transparency differs from that of the body in which they were, do not undergo refraction for the sake of the [seeing] eye; nor will the eye be aware of their refraction or of the surface of the differently transparent body; rather, the refraction occurs in virtue of a natural property that is peculiar to the forms of light and colour. The refracted forms of the stars thus reach the eye just as the forms of visible objects in the air reach it, and sight perceives them just as it does the forms of objects in the air.

Sight perceives the colour of the sky and the extension of that colour in length and breadth without perceiving the shape of the sky or identifying its figure by pure sensation. And when sight perceives a colour as extended in length and breadth without perceiving its shape or identifying its figure, then it perceives it as flat, because it assimilates it to familiar flat surfaces that extend in length and breadth, such as those of walls and the ground. Similarly, sight perceives convex and concave surfaces from a large distance as flat, and also perceives the surface of the earth in wide areas as flat and is not aware of its convexity in the absence of hills and protrusions and depressions.

Sight, therefore, perceives the surface of the heavens as flat and perceives the stars in the same way as it perceives familiar objects scattered over wide areas of large dimensions, and assimilates the distances of the stars to those of familiar objects scattered over vast areas on the earth’s surface the ends of which [areas] sight perceives as farther away than their middles and perceives those points that are close to the middle as less distant than those that are further removed from it. But if sight perceives under equal angles a number of objects scattered over a large area, while perceiving the magnitudes of the distances of those objects, then it will perceive the farther of those objects as larger. For it will perceive the size of the far object from comparing the angle subtended by that far object at the eye-centre to a large distance, and will perceive the size of the nearer object from comparing the angle subtended by this near object (which is the same as the angle subtended by the far object) to a smaller distance.

Now this is found to be clearly the case, namely, that when two objects are viewed under the same angle (or under two equal angles), and their distances differ appreciably, then the farther object will look bigger. For let someone stand in front of a wide wall and raise his hand before one eye while closing the other, then look with one eye while holding his hand between it and the wall. His hand will screen a portion of the width of that wall. But since the hand and the [screened] width of the wall are seen at the same time, then they are seen through the same angle; and sight will at the same time perceive the [screened] width of the wall to be many times larger than the hand. And if the
person moves his hand to one side so as to expose the hidden portion of the wall, and looks at the exposed portion and at his hand, he will perceive that portion to be many times larger than his hand. And he perceives his hand and the exposed portion by two equal angles. This experiment therefore shows that sight perceives size from comparing the angle to the magnitude of the distance.

Now sight perceives the surface of the heavens as flat and does not perceive its concavity, and it perceives the stars scattered over it. It therefore perceives separate and equal stars as having unequal sizes, because it compares the angle subtended at the eye-centre by the extreme star near the horizon to a large distance, while comparing the angle subtended by the star at or near the middle of the sky to a small distance. Thus it will perceive the star at or near the horizon to be larger than the star at the middle of the sky or near the zenith; and will perceive one and the same star (or interval between two stars) at different points in the sky to be of different magnitudes. It will thus perceive one and the same star to be larger at or near the horizon than at or near the middle of the sky, and will perceive the interval between two given stars to be larger at the horizon than at or near the middle of the sky, because it compares the angle subtended at the eye-centre by the horizon-star to be a large distance and compares the angle subtended by the zenith-star to a small distance. And there is no great discrepancy between the two angles; rather, they are <close> though different. And the case is similar with intervals between the stars. But if the sense [of sight] compares two angles close in magnitude to two distances of greatly different magnitudes, then it will perceive the farther [object] to be larger.

The proof of the truth of this explanation is that the angles subtended at the eye-centre by a given star from all regions in the sky <when these angles are contained by refracted lines> are equal <to those through which the star is perceived> by means of straight, unrefracted lines. For the eye being located at the centre of the heavens, these angles will not be greatly reduced by refractions of the stars’ forms. And if these reductions are not greatly different in magnitude, then the difference between the refraction angles through which the star (or interval between two stars) is perceived in different locations will not be great. And if so, then the size of the star (or interval between two stars) will not seem to differ greatly at different points in the sky on account of the difference between the angles. That the refraction angles are not greatly smaller than the angles contained by the straight lines has been shown in the experiment described in the chapter on refraction, where it was made clear that sight perceives the star by refraction…. From which it is manifest that the refracted angles are too small to bring about a great difference between the angles through which sight perceives the star at different points in the sky. But there is a great difference in size between the star (or interval between two stars) at the horizon and at the middle of the sky. Therefore the difference between the refraction angles cannot be the cause of that difference in size at the various
points in the sky. And it has been shown that sight perceives the sizes of visible objects by comparing the angles to the distances. Therefore, if the difference between the angles is small, and the difference between the distances is great, and a visible object appears larger from the greater distance—if all that is so, then the reason why the stars (and intervals between them) are seen to be larger at the horizons than at or near the middle of the sky is the conjecture made by the sense [of sight] regarding their greater distance at the horizons than when they are at the middle of the sky.

What sight perceives regarding the difference in the size of the stars at different positions in the sky is one of the errors of sight. It is one of the constant and permanent errors because its cause is constant and permanent. The explanation of this is [as follows]: Sight perceives the surface of the heavens that faces the eye as flat, and thus fails to perceive its concavity and the equality of the distances [of points on it] from the eye. And it has been established in the mind that flat surfaces that extend in all directions about the eye are unequally distant from it, <and that those directly above are closer to the eye than those to the right and, left of it>. Now sight perceives those parts of the sky near the horizon to be farther away than parts near the middle of the sky; and there is no great discrepancy between the angles subtended at the eye-centre by a given star from any region in the sky; and sight perceives the size of an object by comparing the angle subtended by the object at the eye-centre to the distance of that object from the eye; therefore, it perceives the size of the star (or interval between two stars) at or near the horizon from comparing its angle to a large distance, and perceives the size of that star (or interval) at or near the middle of the sky from comparing its angle (which is equal or close to the former angle) to a small distance; <and it perceives a great difference between the distance of the middle and that of the horizon>. We have thus stated the cause on account of which sight errs in regard to the difference in the size of the stars (or their mutual intervals); and it is a constant, permanent and unchanging cause.

That is also the cause of sight’s perception of the stars as small on account of their remoteness. For the star, when remote, subtends a small angle at the centre of the eye; and, failing to ascertain the magnitude of the star’s distance, the sentient [faculty] merely makes a conjecture in regard to this magnitude, comparing the distances of the stars to the distances of familiar but excessively far objects on the surface of the ground. Thus [the sense faculty] compares the angle produced by the star at the eye (which is a small angle) to a distance like the earthly distances, and in consequence of this comparison perceives the star as small. If sight had a true perception of the magnitude of the star’s distance, it would perceive the star as large. Similarly, sight perceives all excessively far objects on the surface of the ground as small because it does not ascertain their distances. We have thoroughly explained this in Book III of this work.

And just as sight errs in regard to the magnitude of a star’s distance because
it fails to ascertain it and because it assimilates it to distances on the surface of the earth, so also it errs in regarding the distances of the star at different positions in the sky as unequal in magnitude, though these distances are equal, because, again, it assimilates them to certainly unequal distances to the right or left or in front of it on the surface of the ground. And just as its error in regard to the star’s distance and size is permanent and constant, so also its error in the difference between the star’s distances and size at various positions [in the sky] is permanent and constant. For the form of these distances does not vary in the eye from time to time but is always the same, and it is sight that assimilates it to the distances of familiar and excessively far objects on the surface of the earth.

The enlargement of heavenly objects at the horizon may frequently have another cause. This cause occurs when a thick vapour stands between the eye and the star positioned at or near the horizon, if the vapour is at or near the horizon and does not continue to the middle of the sky but rather forms a section of a sphere whose centre is the centre of the world because it surrounds the earth. If such a section terminates before [reaching] the middle of the sky, then the surface of it that faces the eye will be plane. But if the surface of the vapour facing the eye is plane, then the form[s] of the stars (and intervals between them) will be seen behind the vapour as larger than before the vapour occurred. Because the form of the star will [first] occur at the place on the heavens concavity from which it will be refracted to the eye.

Then, in the absence of the thick vapour, the form would extend from this place to the eye in straight lines. But, in the presence of the thick vapour at the horizon, this form will extend to the surface of the vapour that faces the [middle of] the sky, and thus occur in that surface. Sight will therefore perceive this form just as it would perceive objects placed in the vapour: that is, the form will extend through the thick vapour on straight lines then will be refracted at the surface of the vapour facing the eye, this refraction being away from the normal to the vapour’s surface (which is a plane surface), since the air near the eye is rarer than the thick vapour. It follows from this that the form will appear to be larger than it would if it were seen rectilinearly. (This was shown in proposition 1 of this chapter, namely, that when the eye is in the rarer medium and the visible object in the denser medium, and the surface of the denser medium is plane [then the object will look bigger than it is]). Thus the form that occurs in the surface of the vapour facing the middle of the sky is the visible object, and the medium in which this form is the thick vapour, and the eye is in the rarer medium of air.

The principal cause on account of which the stars (and their mutual intervals) are seen at the horizons to be larger than at or near the middle of the sky is the one stated earlier. It is the inseparable and permanent cause. When, however, a thick vapour happens to rise at the horizons, it increases
their magnification. But this is an accidental cause which always occurs [only] in certain regions of the earth and occasionally in others but is not permanent.

- **Hint to Problem 3** Alhazen’s psychological explanation operates with a flattened dome-shaped sky (see also the figure on page 320). The distance perception of heavenly bodies is different on the horizon and on the zenith. As perceived distance and the angle subtended by an object determines perceived size according to Alhazen, the more distantly perceived stars near the horizon appear larger. Coupled with this permanent cause, often accidental causes (thick vapour between the observer and the observed bodies) magnify the effect. The latter is a physical explanation. None of these can account for the secondary (distance) aspect of the illusion, and there is no reason to believe that Alhazen was conscious of this, today generally accepted part of the moon illusion.

3.5 **Text 5: Descartes on Natural Geometry**

- In the second edition of Berkeley’s work in 1709 (see Text 6) an Appendix appeared with the most important Cartesian quotes from Descartes’s *Dipotrica*, first appearing in 1637. We can read the following about distance-perception (VI, 13.).

Figure X.2: Descartes’s blind man using sticks and his tactile sensation for apprehending distance.
We apprehend distance, moreover, through a sort of joint activity of the eyes. For in the same way as our blind man (see figure X.2), holding two sticks of indeterminate length, AE and CE, and knowing only the distance between his hands, A and C, together with the size of the angles ACE and CAE, can thence determine the position of E by a sort of innate geometrical knowledge shared by all men, so, when both our eyes (see figure X.3), RST and rst, are focused on X, the length of the line Ss and the size of the angles XsS and XsS let us know the position of the point X. We can also discover that position by means of either one of our eyes alone, by changing its location. If we keep the eye fixed on X and hold it first at point S and then immediately afterwards at point s, that will be enough for the length of the line Ss and the size of the angles XsS and XsS to be present together in the imagination and thus to inform us of the distance of the point X. They do so in virtue of an act of the mind which, while it may seem to be a simple judgment, nevertheless includes within itself a certain reasoning process like that by which geometers calculate inaccessible
positions from two separate given points.

3.6 Text 6: Berkeley on the Illusion

- George Berkeley’s *An Essay Towards a New Theory of Vision* was first published in 1709 (Berkeley 1709). The excerpts are from the fourth edition (1732).

34. First, any radiating point is then distinctly seen when the rays proceeding from it are, by the refractive power of the crystalline, accurately reunited in the retina or fund of the eye: but if they are reunited, either before they arrive at the retina, or after they have passed it, then there is confused vision.

35. Secondly, suppose in the adjacent figures NP represent an eye duly framed and retaining its natural figure. In Fig. 1 the rays falling nearly parallel on the eye, are by the crystalline AB refracted, so as their focus or point of union F falls exactly on the retina: but if the rays fall sensibly diverging on the eye, as in Fig. 2, then their focus falls beyond the retina: or if the rays are made to converge by the lens QS before they come at the eye, as in Fig. 3, their focus F will fall before the retina. In which two last cases it is evident from the foregoing section that the appearance of the point Z is confused. And by how much the greater is the convergency, or divergency, of the rays falling on the pupil, by so much the farther will the point of their reunion be from the retina, either before or behind it, and consequently the point Z will appear by so much the more confused. And this, by the bye, may shew us the difference between confused and faint vision. Confused vision is when the rays proceedings from each distinct point of the object are not accurately recollected in one corresponding point on the retina, but take up some space thereon, so that rays from different points become mixed and confused together. This is opposed to a distinct vision, and attends near objects. Faint vision is when by reason of the distance of the object or grossness of the interjacent medium few rays arrive from the object to the eye. This is opposed to vigorous or clear vision, and attends remote objects. But to return.

56. Now in order to discover by what means the magnitude of tangible objects is perceived by sight. I need only reflect on what passes in my own mind, and observe what those things be which introduce the ideas of greater or lesser into my thoughts, when I look on any object. And these I find to be, first, the magnitude or extension of the visible object, which being immediately perceived by sight, is connected with that other which is tangible and placed at a distance. Secondly, the confusion or distinctness. And thirdly, the vigorousness or faintness of the aforesaid visible appearance. Ceteris paribus, by how much the greater or lesser the visible object is, by so much the greater or lesser do I conclude the tangible object to be. But, be the idea immediately perceived by
sight never so large, yet if it be withal confused, I judge the magnitude of the thing to be but small. If it be distinct and clear, I judge it greater. And if it be faint, I apprehend it to be yet greater. What is here meant by confusion and faintness hath been explained in sect. 35.

67. There is a celebrated phenomenon, the solution whereof I shall attempt to give by the principles that have been laid down, in reference to the manner wherein we apprehend by sight the magnitude of objects. The apparent magnitude of the moon when placed in the horizon is much greater than when it is in the meridian, though the angle under which the diameter of the moon is seen be not observed greater in the former case than in the latter: and the horizontal moon doth not constantly appear of the same bigness, but at some times seemeth far greater than at others.

68. Now in order to explain the reason of the moon’s appearing greater than ordinary in the horizon, it must be observed that the particles which compose our atmosphere intercept the rays of light proceeding from any object to the eye; and by how much the greater is the portion of atmosphere interjacent between the object and the eye, by so much the more are the rays intercepted; and by consequence the appearance of the object rendered more faint, every object
appearing more vigorous or more faint in proportion as it sendeth more or fewer rays into the eye. Now between the eye and the moon, when situated in the horizon, there lies a far greater quantity of atmosphere than there does when the moon is in the meridian. Whence it comes to pass that the appearance of the horizontal moon is fainter, and therefore by sect. 56 it should be thought bigger in that situation than in the meridian, or in any other elevation above the horizon.

69. Farther, the air being variously impregnated, sometimes more and sometimes less, with vapours and exhalations fitted to retard and intercept the rays of light, it follows that the appearance of the horizontal moon hath not always an equal faintness, and by consequence that luminary, though in the very same situation, is at one time judged greater than at another.

70. That we have here given the true account of the phenomena of the horizontal moon will, I suppose, be farther evident to anyone from the following considerations. First, it is plain that which in this case suggests the idea of greater magnitude must be something which is itself perceived; for that which is unperceived cannot suggest to our perception any other thing. Secondly, it must be something that does not constantly remain the same, but is subject to some change or variation, since the appearance of the horizontal moon varies, being at one time greater than at another. And yet, thirdly, it cannot be the visible figure or magnitude, since that remains the same, or is rather lesser, by how much the moon is nearer to the horizon. It remains therefore that the true cause is that affection or alteration of the visible appearance which proceeds from the greater paucity of rays arriving at the eye, and which I term faintness: since this answers all the forementioned conditions, and I am not conscious of any other perception that doth.

71. Add to this that in misty weather it is a common observation that the appearance of the horizontal moon is far larger than usual, which greatly conspires with and strengthens our opinion. Neither would it prove in the least irreconcilable with what we have said, if the horizontal moon should chance sometimes to seem enlarged beyond its usual extent, even in more serene weather. For we must not only have regard to the mist which happens to be in the place where we stand; we ought also to take into our thoughts the whole sum of vapours and exhalations which lie betwixt the eye and the moon: all which cooperating to render the appearance of the moon more faint, and thereby increase its magnitude, it may chance to appear greater than it usually does, even in the horizontal position, at a time when, though there be no extraordinary fog or haziness, just in the place where we stand, yet the air between the eye and the moon, taken all together, may be loaded with a greater quantity of interspersed vapours and exhalations than at other times.

72. It may be objected that in consequence of our principles the interposition of a body in some degree opaque, which may intercept a great part of
the rays of light, should render the appearance of the moon in the meridian as large as when it is viewed in the horizon. To which I answer, it is not faintness anyhow applied that suggests greater magnitude, there being no necessary but only an experimental connexion between those two things. It follows that the faintness which enlarges the appearance must be applied in such sort, and with such circumstances, as have been observed to attend the vision of great magnitudes. When from a distance we behold great objects, the particles of the intermediate air and vapours, which are themselves unperceivable, do interrupt the rays of light, and thereby render the appearance less strong and vivid: now, faintness of appearance caused in this sort hath been experienced to coexist with great magnitude. But when it is caused by the interposition of an opaque sensible body, this circumstance alters the case, so that a faint appearance this way caused doth not suggest greater magnitude, because it hath not been experienced to coexist with it.

73. Faintness, as well as all other ideas or perceptions which suggest magnitude or distance, doth it in the same way that words suggest the notions to which they are annexed. Now, it is known a word pronounced with certain circumstances, or in a certain context with other words, hath not always the same import and signification that it hath when pronounced in some other circumstances or different context of words. The very same visible appearance as to faintness and all other respects, if placed on high, shall not suggest the same magnitude that it would if it were seen at an equal distance on a level with the eye. The reason whereof is that we are rarely accustomed to view objects at a great height; our concerns lie among things situated rather before than above us, and accordingly our eyes are not placed on the top of our heads, but in such a position as is most convenient for us to see distant objects standing in our way. And this situation of them being a circumstance which usually attends the vision of distant objects, we may from hence account for (what is commonly observed) an object’s appearing of different magnitude, even with respect to its horizontal extension, on the top of a steeple, for example, an hundred feet high to one standing below, from what it would if placed at an hundred feet distance on a level with his eye. For it hath been shewn that the judgment we make on the magnitude of a thing depends not on the visible appearance alone, but also on divers other circumstances, any one of which being omitted or varied may suffice to make some alteration in our judgment. Hence, the circumstances of viewing a distant object in such a situation as is usual, and suits with the ordinary posture of the head and eyes being omitted, and instead thereof a different situation of the object, which requires a different posture of the head taking place, it is not to be wondered at if the magnitude be judged different: but it will be demanded why an high object should constantly appear less than an equidistant low object of the same dimensions, for so it is observed to be: it may indeed be granted that the variation of some circumstances may vary the
Judgment made on the magnitude of high objects, which we are less used to
look at: but it does not hence appear why they should be judged less rather than
greater? I answer that in case the magnitude of distant objects was suggested by
the extent of their visible appearance alone, and thought proportional thereto, it
is certain they would then be judged much less than now they seem to be (vide
sect. 79). But several circumstances concurring to form the judgment we make
on the magnitude of distant objects, by means of which they appear far larger
than others, whose visible appearance hath an equal or even greater extension;
it follows that upon the change or omission of any of those circumstances
which are wont to attend the vision of distant objects, and so come to influence
the judgments made on their magnitude, they shall proportionally appear less
than otherwise they would. For any of those things that caused an object to be
thought greater than in proportion to its visible extension being either omitted
or applied without the usual circumstances, the judgment depends more entirely
on the visible extension, and consequently the object must be judged less. Thus
in the present case the situation of the thing seen being different from what
it usually is in those objects we have occasion to view, and whose magnitude
we observe, it follows that the very same object, being an hundred feet high,
should seem less than if it was an hundred feet off on (or nearly on) a level with
the eye. What has been here set forth seems to me to have no small share in
contributing to magnify the appearance of the horizontal moon, and deserves
not to be passed over in the explication of it.

74. If we attentively consider the phenomenon before us, we shall find the
not discerning between the mediate and immediate objects of sight to be the
chief cause of the difficulty that occurs in the explication of it. The magnitude
of the visible moon, or that which is the proper and immediate object of vision,
is not greater when the moon is in the horizon than when it is in the meridian.
How comes it, therefore, to seem greater in one situation than the other? What
is it can put this cheat on the understanding? It has no other perception of the
moon than what it gets by sight: and that which is seen is of the same extent,
say, the visible appearance hath the same, or rather a less, magnitude when
the moon is viewed in the horizontal than when in the meridional position: and
yet it is esteemed greater in the former than in the latter. Herein consists the
difficulty, which doth vanish and admit of a most easy solution, if we consider
that as the visible moon is not greater in the horizon than in the meridian, so
neither is it thought to be so. It hath been already shewn that in any act of
vision the visible object absolutely, or in itself, is little taken notice of, the
mind still carrying its view from that to some tangible ideas which have been
observed to be connected with it, and by that means come to be suggested by
it. So that when a thing is said to appear great or small, or whatever estimate
be made of the magnitude of any thing, this is meant not of the visible but of
the tangible object. This duly considered, it will be no hard matter to reconcile
the seeming contradiction there is, that the moon should appear of a different bigness, the visible magnitude thereof remaining still the same. For by sect. 56 the very same visible extension, with a different faintness, shall suggest a different tangible extension. When therefore the horizontal moon is said to appear greater than the meridional moon, this must be understood not of a greater visible extension, but of greater tangible or real extension, which by reason of the more than ordinary faintness of the visible appearance, is suggested to the mind along with it.

75. Many attempts have been made by learned men to account for this appearance. Gassendus, Descartes, Hobbes, and several others have employed their thoughts on that subject; but how fruitless and unsatisfactory their endeavours have been is sufficiently shewn in The Philosophical Transactions, where you may see their several opinions at large set forth and confuted, not without some surprize at the gross blunders that ingenious men have been forced into by endeavouring to reconcile this appearance with the ordinary Principles of optics. Since the writing of which there hath been published in the Transactions another paper relating to the same affair by the celebrated Dr. Wallis, wherein he attempts to account for that phenomenon which, though it seems not to contain anything new or different from what had been said before by others, I shall nevertheless consider in this place.

76. His opinion, in short, is this; we judge not of the magnitude of an object by the visual angle alone, but by the visual angle in conjunction with the distance. Hence, though the angle remain the same, or even become less, yet if withal the distance seem to have been increased, the object shall appear greater. Now, one way whereby we estimate the distance of anything is by the number and extent of the intermediate objects: when therefore the moon is seen in the horizon, the variety of fields, houses, etc., together with the large prospect of the wide extended land or sea that lies between the eye and the utmost limb of the horizon, suggest unto the mind the idea of greater distance, and consequently magnify the appearance. And this, according to Dr. Wallis, is the true account of the extraordinary largeness attributed by the mind to the horizontal moon at a time when the angle subtended by its diameter is not one jot greater than it used to be.

77. With reference to this opinion, not to repeat what hath been already said concerning distance, I shall only observe, first, that if the prospect of interjacent objects be that which suggests the idea of farther distance, and this idea of farther distance be the cause that brings into the mind the idea of greater magnitude, it should hence follow that if one looked at the horizontal moon from behind a wall, it would appear no bigger than ordinary. For in that case the wall interposing cuts off all that prospect of sea and land, etc. which might otherwise increase the apparent distance, and thereby the apparent magnitude of the moon. Nor will it suffice to say the memory doth even
then suggest all that extent of land, etc., which lies within the horizon; which suggestion occasions a sudden judgment of sense that the moon is farther off and larger than usual. For ask any man who, from such a station beholding the horizontal moon, shall think her greater than usual, whether he hath at that time in his mind any idea of the intermediate objects, or long tract of land that lies between his eye and the extreme edge of the horizon? And whether it be that idea which is the cause of his making the aforementioned judgment? He will, I suppose, reply in the negative, and declare the horizontal moon shall appear greater than the meridional, though he never thinks of all or any of those things that lie between him and it. Secondly, it seems impossible by this hypothesis to account for the moon’s appearing in the very same situation at one time greater than at another; which nevertheless has been shewn to be very agreeable to the principles we have laid down, and receives a most easy and natural explication from them. For the further clearing’ up of this point it is to be observed that what we immediately and properly see are only lights and colours in sundry situations and shades and degrees of faintness and clearness, confusion and distinctness. All which visible objects are only in the mind, nor do they suggest ought external, whether distance or magnitude, otherwise than by habitual connexion as words do things. We are also to remark that, beside the straining of the eyes, and beside the vivid and faint, the distinct and confused appearances (which, bearing some proportion to lines and angles, have been substituted instead of them in the foregoing part of this treatise), there are other means which suggest both distance and magnitude; particularly the situation of visible points of objects, as upper or lower; the one suggesting a farther distance and greater magnitude, the other a nearer distance and lesser magnitude: all which is an effect only of custom and experience; there being really nothing intermediate in the line of distance between the uppermost and lowermost, which are both equidistant, or rather at no distance from the eye, as there is also nothing in upper or lower, which by necessary connexion should suggest greater or lesser magnitude. Now, as these customary, experimental means of suggesting distance do likewise suggest magnitude, so they suggest the one as immediately as the other. I say they do not (vide sect. 53) first suggest distance, and then leave the mind from thence to infer or compute magnitude, but suggest magnitude as immediately and directly as they suggest distance.

78. This phenomenon of the horizontal moon is a clear instance of the insufficiency of lines and angles for explaining the way wherein the mind perceives and estimates the magnitude of outward objects. There is nevertheless a use of computation by them in order to determine the apparent magnitude of things, so far as they have a connexion with, and are proportional to, those other ideas or perceptions which are the true and immediate occasions that suggest to the mind the apparent magnitude of things. But this in general may, I think, be observed concerning mathematical computation in optics: that it can
never be very precise and exact since the judgments we make of the magnitude of external things do often depend on several circumstances, which are not proportionable to, or capable of being defined by, lines and angles.

79. From what has been said we may safely deduce this consequence; to wit, that a man born blind and made to see would, at first opening of his eyes, make a very different judgment of the magnitude of objects intromitted by them from what others do. He would not consider the ideas of sight with reference to, or as having any connexion with, the ideas of touch: his view of them being entirely terminated within themselves, he can no otherwise judge them great or small than as they contain a greater or lesser number of visible points. Now, it being certain that any visible point can cover or exclude from view only one other visible point, it follows that whatever object intercepts the view of another hath an equal number of visible points with it; and consequently they shall both be thought by him to have the same magnitude. Hence it is evident one in those circumstances would judge his thumb, with which he might hide a tower or hinder its being seen, equal to that tower, or his hand, the interposition whereof might conceal experimental means the firmament from his view, equal to the firmament: how great an inequality soever there may in our apprehensions seem to be betwixt those two things, because of the customary and close connexion that has grown up in our minds between the objects of sight and touch; whereby the very different and distinct ideas of those two senses are so blended and confounded together as to be mistaken for one and the same thing; out of which prejudice we cannot easily extricate ourselves.

- **Hint to Problem 4** To evade computational mechanisms, Berkeley has to find a property correlating with distance that can be perceived directly and learnt. For Berkeley this property is the result of the particles of the atmosphere. The thicker the layer the more faint the image of objects seen through them. As the horizon moon is seen through the thickest layer of atmosphere it is judged to be most distant, and thus bigger (if it subtends the same visual angle). The distal cue is faintness. This physical cue, through a long process of learning triggers a perception of distance.
3.7 Text 7: Kaufman and Rock

Reprinted with permission from Science 136 (3520) 953-961 and (3521) 1023-1031 ©2005.

15 June 1962, Volume 136, Number 3520

The Moon Illusion, I

Explanation of this phenomenon was sought through the use of artificial moons seen on the sky.

Lloyd Kaufman and Irvin Rock

Since antiquity men have puzzled about the fact that the moon and sun appear larger over the horizon than at the zenith. The distance to the moon is actually somewhat greater at the horizon (because the tangent to the earth is greater than the distance to the earth's surface along the line joining the centers of earth and moon), so the retinal image of the horizon moon is a fraction smaller than that of the zenith moon.

The phenomenon cannot be explained by differences in refraction based upon differences in the angle of incidence to the earth's atmosphere. Logically, refraction could be thought to displace an image, but enlargement would require refraction in different directions, as is the case with a lens. In point of fact there is a displacement effect, often noticeable in the setting sun, where the vertical diameter is foreshortened so as to yield the impression of an ellipse. But, again, this effect actually produces a smaller image of the horizon object.

In any case, all lingering doubt about this explanation can be laid to rest with the realization that photographs of the moon in the different positions yield no measurable difference. Hence, the effect has nothing to do with atmospheric optics.

We are, therefore, dealing with an illusion in which retinal images of approximately equal size yield quite different phenomenal impressions of size, based upon differences in perceived direction. There are, however, two ways of understanding the phrase “differences in perceived direction” in this context. In relation to the observer—that is, egocentrically defined—the horizon moon is “straight ahead,” whereas, to view the zenith moon, the observer must raise his head or his eyes, or both. When the observer changes his position—as when he is lying supine—the zenith moon, not the horizon moon, is egocentrically “straight ahead.” On the basis of the egocentric definition one would have to predict that the zenith moon would appear larger than the horizon moon to an observer lying supine. On the other hand, from an objective standpoint, one can say that the horizon moon is seen at the point where the terrain meets the sky, whereas the zenith moon is seen surrounded by sky on all sides, apart from horizon or horizon. Here the two positions of the moon are distinguished in terms of their geographical location relative to an observer on the earth's surface, and changes in the position of the observer's eye or head are considered irrelevant.

The two ways of defining the difference in direction have led to different theories in explanation of the moon illusion.

A theory based on the second approach was put forth by Ptolemy, among others, to the effect that the presence of the terrain creates the impression that the horizon moon is farther away than the zenith moon because the filled space between the observer and the horizon produces an impression of greater extensity than the unfilled space between the observer and the zenith. If this were true it would follow that the horizon moon would look larger than the zenith moon, as shown in Fig. 1. Many people find this point hard to grasp because it seems paradoxical: If something is perceived as farther away shouldn't it appear smaller, not larger? The confusion lies in the fact that in this case the size of the optical image remains unchanged. It would require a larger object to yield the same image from a greater distance.

This theory can now be restated in the language of the psychology of size perception. It is a well-known fact that the apparent size of an object depends not only on the size of the retinal image or visual angle but on the distance as well. Within certain limits, objects do not appear to vary substantially in size when viewed from varying distances, despite the fact that the size of the optical image varies inversely with distance. This phenomenon is known as size constancy. It is as if the observer took the distance into account in perceiving the size of the object. In line with this same reasoning, where the visual angle remains constant but where the distances are registered as different, the apparent size will change. In the case of an afterimage projected on surfaces at different distances, the apparent size is a direct function of the distance of the surface, a relation known as Emmert's law. The moon illusion can be considered a special manifestation of Emmert's law, or a manifestation of a similar functional relationship, if it is true that the registered distance to the horizon moon is greater than that to the zenith moon, since the visual angle of the moon remains approximately constant. In fact, on the basis of such a possible difference in registered distance, psychologists would have to pre-
dict a moon illusion even if one had never been observed (1).

But is it true that the horizon moon appears farther away than the zenith moon? Boring concluded not too long ago that quite the contrary is the case (2). He asked observers which moon appeared nearer, and they were unanimous in saying that the horizon moon did. We will return later to this point, since we believe there is a fallacy involved in putting the question this way. For the moment it will be worth while to follow Boring’s reasoning. If it is the case that the horizon moon does not seem more distant, then theories such as Ptolemy’s are invalid. Hence it seemed to Boring that there was little to be gained from sticking to the distinction between “horizon” and “zenith” as objectively defined. He therefore proceeded to examine the implication of the egocentric definition, that the crucial difference for the two moon directions is a difference in the angle of regard.

This, however, is a hypothesis requiring experimental proof, and if one attempts to test it, the question of method arises. It is hardly satisfactory to compare the apparent sizes of the moon in its different celestial positions under natural conditions, for two reasons. First, one must depend on memory in comparing the size of the moon as seen previously in one position with the size of the moon as seen in a different position. Second, no measure of the effect is obtained. Although there have been various attempts to measure the illusion in the past, Boring is to be credited with the first thoroughgoing experiments on the moon illusion. His method was as follows. While viewing either the horizon or the zenith moon, the observer was shown a series of disks projected on a screen 3.5 meters away and off to one side. He was told to select the disk which seemed to match the moon in size. Comparison of the average value for many such determinations made by each subject for the horizon moon with the average for determinations for the zenith moon gave a measure of the moon illusion.

The Angle-of-Regard Hypothesis

Using this method, Holway and Boring (3) proceeded to test the hypothesis that the illusion is based on the angle of regard. They first showed that an illusion was obtained. For three observers the average ratio of the diameter of the horizon-moon matching disk to that of the zenith-moon matching disk was 1.67. They next showed that the illusion was reversed (with a ratio of 1.47) when two observers viewed the zenith moon from a supine position, so that from the egocentric viewpoint it was “straight ahead,” and viewed the horizon moon by bending their heads backward from the supine position, so that it was “elevated.” It should be noted, however, that in spite of these quantitative results, to these observers the zenith moon viewed from a supine position seemed less large than the horizon moon viewed from an upright position. In a second paper (4) Holway and Boring reported an attempt to create the illusion by using a reflected image of the moon in a front-surface mirror, so that the observer’s response would be independent of the actual position of the moon in the sky. Using a reflected moon as the standard, the investigators compared the illusion obtained with the eyes raised and the head level with that obtained with the head raised and the eyes level. As their own subjects they found an illusion of 2.0 with the eyes raised and a ratio of 1.0, or no illusion, with only the head raised. In further experiments they found that, regardless of the actual position of the moon, they could obtain an illusion if, and only if, the eyes were raised with respect to the head in one case and level in the other. They found a similar effect for eyes lowered with respect to the head. In other words, Holway and Boring analyzed the angle-of-regard hypothesis and, using the method described, found the crucial factor in looking “up” to be the angle of the eyes with respect to the head. In a further experiment Taylor and Boring (5) found that the illusion is dependent upon binocular vision and disappears for monocular observation, provided this is not preceded by binocular observation. This fact suggested that the cause of the illusion is certain changes, either in convergence innervation or in torsion, when the eyes are raised. But since such changes would presumably yield cues of nearer distance, they should make the zenith moon appear nearer (and this, it will be recalled, is the opposite of what Boring and his colleagues believed to be the case). Hence, they concluded that there is still no satisfactory theory of the moon illusion.

Grounds for Caution

There are, however, three reasons for caution in accepting the conclusion that the angle of the eyes with respect to the head is crucial for the moon illusion. The first concerns the basis for rejection of the “apparent-distance” hypothesis. Suppose that the moon illusion is based on a difference in the perceived or registered distance of the two moons, and that an observer experiences the illusion. If he is now asked which moon seems the more distant, he is comparing moons which already differ in apparent size. It is natural for him to reason that the larger moon is closer. This is more of a judgmental than a perceptual reaction, and the judgment may dominate when the observer is asked this question. He may not be aware that the horizon moon appears larger because of the stimulus of the terrain. As noted in (1), the observer need not necessarily be aware that a stimulus correlates of distance is registering, and thereby affecting apparent size. When asked about the distances of the two moons, therefore, he may fall back upon what seems an obvious clue—namely, the different apparent sizes of the moons.

We have performed two experiments that support this argument. In the first we sought to demonstrate that subjects will report as farther away whichever moon is smaller. Eight subjects were shown a horizon and a zenith disk or “moon” (by a method described later in more detail), one disk being set so as to yield a much larger impression than the other. The subjects were unanimous in reporting the smaller-appearing disk as farther away, regardless of whether it was seen over the horizon or at the zenith. This finding deprives Boring’s argument on the apparent distance of the two moons of its crucial significance. It shows that whichever moon appears larger will be judged nearer, quite apart from any other factors that produce differences in perceived distance.

In the second experiment we sought to show that if one does not use the moon itself (which so readily produces the effect just described) as the object with which to gauge apparent distance, it is possible to obtain direct evidence in support of the assumption that the horizon “sky” seems farther away than the zenith (we are assuming the moon to be coplanar with the sky).
Ten subjects, looking at a sky without moons, were asked to scan the sky and to try to perceive it as a surface. They were then to report whether an imaginary point over the horizon seemed nearer or farther than such a point at the zenith. Nine of the ten reported that the horizon sky seemed farther away. The tenth subject said the horizon sky and the zenith seemed equidistant. This finding corresponds to the belief, which goes back at least to the 11th century, that the sky appears somewhat flat. Smith, in 1738, actually tried to determine the degree of flatness by measuring the half-arc angle (6). The observer points to that spot in the sky which is perceived as bisecting the arc of sky connecting the horizon with the zenith. The angle which this direction forms with the ground is the half-arc angle. A hemispherically shaped sky would yield a 45° angle a₁, but a flattened sky would yield a smaller angle a₀ (see Fig. 1). The most recent attempt to determine this angle was made by Müller (7), a student of the meteorologist Neuberger (8); Müller found its angle to be in the neighborhood of 30° and to vary inversely with the distance of the horizon and directly with the elevation. In other words, his data supported the hypothesis that the apparent flattening is a function of perceived distance along the ground plane.

Thus we have shown that the distance hypothesis is viable, but in the face of the evidence on eye elevation it may seem superfluous—the relevant factor has already been isolated. This leads to the second reason for caution in accepting this evidence as conclusive—namely, the question of methodology. Is it possible meaningfully to compare, in size, an infinitely distant object, whose size is therefore more or less indeterminate, and a nearby object? It would seem that the two are incommensurate. If the observers are comparing linear size, their matches imply that the distant moon appears to them to be between 5 and 17 inches in diameter. It seems unlikely that many observers would agree that the apparent size of the moon lies within these limits, and even unlikely that the same observer would make a similar judgment under somewhat different conditions of measurement. In fact, Taylor and Boring noted that if the observer moved backward from the screen so as to view from a point farther than 3.5 meters away the disk which he had selected as a match, it was no longer acceptable. Yet we know that the apparent lineal size of this disk would remain fairly constant.

This suggests that the observers were attempting to match visual angle (thus, backing away from the screen would cause the disk to appear too small), but the fact is that the visual angle of the disk selected for even the zenith moon was 4 to 10 times that of the moon. Therefore, it appears that the choice was not a match in linear size or a very accurate match in visual angle. Hence there seem to be serious difficulties connected with such a method. Boring and his associates were well aware of this problem and in one experiment sought to overcome it by requiring the observer to compare a zenith and a horizon moon directly with one another. They accomplished this by using the front-surface mirror to show the moon in a position different from its actual position in the sky. The observer then compared two reflected moons. Boring and his co-workers obtained an estimated illusion ratio by this technique and also again obtained evidence in support of the eye-elevation hypothesis. This technique is clearly preferable to their earlier method because it duplicates the conditions of the moon illusion—that is, the observer is comparing two moons each located in the sky at optical infinity, in different positions. In the method that we

Fig. 1. Effect of the apparent distance of the moon on its apparent size. Top arc. The true position of the moon; all points on the curve are at the same distance from the observer. Bottom arc. The apparent distance of the moon according to the theory. Solid circles: Resultant differences in perceived size of the moon. The figure also illustrates the effect of apparent shape of the sky on the half-arc angle measure a. M₀, Midpoint of actual arc connecting zenith and horizon; M₁, midpoint of perceived arc connecting zenith and horizon. [After Smith (6)].

15 JUNE 1962
Table 1. Average illusion ratios for the eyes-level and eyes-elevated conditions. Aperture diameter for the standard, 0.093 inch.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Horizon standard</th>
<th>Zenith standard</th>
<th>Combined</th>
<th>Horizon standard</th>
<th>Zenith standard</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.72</td>
<td>2.35</td>
<td>2.03</td>
<td>1.72</td>
<td>2.35</td>
<td>2.03</td>
</tr>
<tr>
<td>2</td>
<td>1.72</td>
<td>1.74</td>
<td>1.73</td>
<td>1.72</td>
<td>1.57</td>
<td>1.65</td>
</tr>
<tr>
<td>3</td>
<td>1.00</td>
<td>1.11</td>
<td>1.06</td>
<td>1.17</td>
<td>0.84</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>1.36</td>
<td>1.44</td>
<td>1.46</td>
<td>1.26</td>
<td>1.23</td>
<td>1.25</td>
</tr>
<tr>
<td>5</td>
<td>1.05</td>
<td>0.87</td>
<td>0.95</td>
<td>1.14</td>
<td>0.96</td>
<td>1.05</td>
</tr>
<tr>
<td>6</td>
<td>1.15</td>
<td>1.12</td>
<td>1.13</td>
<td>1.15</td>
<td>0.93</td>
<td>1.02</td>
</tr>
<tr>
<td>7</td>
<td>1.54</td>
<td>1.27</td>
<td>1.41</td>
<td>1.72</td>
<td>1.65</td>
<td>1.67</td>
</tr>
<tr>
<td>8</td>
<td>1.72</td>
<td>1.74</td>
<td>1.73</td>
<td>1.72</td>
<td>2.00</td>
<td>1.86</td>
</tr>
<tr>
<td>9</td>
<td>1.72</td>
<td>1.54</td>
<td>1.63</td>
<td>1.55</td>
<td>1.57</td>
<td>1.56</td>
</tr>
<tr>
<td>10</td>
<td>1.72</td>
<td>1.40</td>
<td>1.56</td>
<td>1.46</td>
<td>2.00</td>
<td>1.73</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td>1.46</td>
<td></td>
<td></td>
<td>1.48</td>
</tr>
<tr>
<td>Std dev</td>
<td></td>
<td></td>
<td>0.12</td>
<td></td>
<td></td>
<td>0.36</td>
</tr>
</tbody>
</table>

* These subjects were tested after dark.

The Eye-Evaluation Hypothesis

We sought a technique whereby the observer would compare a horizon object and a zenith object localized on the sky and where, in addition, one object could be varied in size to achieve a subjective match. The device used has two major features. One is a combining glass through which the observer looks out at the sky (see Fig. 2). When the glass is placed at a 45° angle to the line of sight it can reflect into the eye a beam of light which is perpendicular to the line of sight. If a slide containing a small circular aperture is placed between a light source and the combining glass, the observer will see a luminous disk in the direction of the sky. The other feature is a collimator placed in front of the aperture so that the virtual image of the aperture seen through the combining glass will be at optical infinity (9). Otherwise the observer might localize the disk somewhere between himself and the distant sky. Two identical units were constructed; one serves as a horizontal moon and the other as a zenith moon. Either can be used as the standard or, by changing the size of the aperture, as the variable. In this way it is possible to measure the illusion. With batteries as the source of power for the light, the apparatus can easily be taken to any desired outdoor location.

In the first series of experiments we sought to determine whether an illusion was obtained with our apparatus and, if so, whether it was based upon elevation of the eyes. In these experiments the collimating lens was 3 5/16 inches in diameter; this made binocular viewing possible. The combining glass was 10 by 12 by 1/16 inches. Although the glass was not silvered, it did not produce an objectionable double image. The apertures were drilled into aluminum slides which could be inserted at one end of a box containing a diffuse light source. The slides were placed 6 5/16 inches behind the collimating lens, since this was the focal length of the lens. The distance from the lens to the combining glass was 15 inches. The distance from the observer’s eyes to the combining glass was between 10 and 15 inches (10) (see Fig. 3). One assembly, for viewing the horizon disk, was placed on a table. The assembly for viewing the zenith disk was mounted on an inclined board behind the observer; it projected an image onto a combining glass placed over his head. In these and other experiments the two assemblies were interchanged occasionally, and thus each was used to produce a horizon disk for some subjects and a zenith disk for other subjects. The apertures varied in size from 0.040 to 0.152 inch in diameter, in steps of from 0.010 to 0.012 inch, and subtended visual angles ranging from 0.34° to 1.38°. The aperture of the standard was 0.095 inch in diameter, subtending a visual angle of 0.8°, somewhat larger than the visual angle subtended by the moon, which is 0.5°.

Eyes raised versus eyes level with respect to the head. Two experiments were performed, with similar procedures and similar results, but we will describe here only the second, in which a certain improvements were incorporated into its design. The apparatus was set up in Freeport, Long Island, in a spot yielding an unobstructed view for more than a mile to the horizon. Ten neighbors of one of us (eight men and two women) served as subjects, and all were naive concerning the point of the experiment. The experiment was conducted in the pre-twilight hours, except for testing of three subjects after dark.

Each subject was tested under two conditions—eyes level with respect to the head and eyes raised with respect to the head. In each case he viewed the horizon disk, or “moon” (hereafter referred to without quotation marks), from an erect position. In the control or eyes-level condition, in viewing the zenith moon, his head was maintained in position by means of a biting board, which he gripped with his teeth. The biting board was adjusted so as to force him to tilt his head to the desired position (so that his eyes were level with respect to his head) for viewing a near-zenith moon at an elevation of about 70°. In the experimental or eyes-elevated condition the biting board was not used for viewing the zenith moon; instead, the subject was required to stand erect and gaze upward naturally. Nor was a biting board or head rest used for viewing the horizon moon. Careful observation attested to the fact that each subject noticeably raised his eyes with respect to his head in viewing the zenith moon.
The subjects were first instructed to view the standard disk and to form an impression of its size. They then shifted their gaze to the variable, which, in the starting position, was much smaller or much larger than the standard. Their task was to indicate when the variable appeared equal in size to the standard. They were permitted to refer back to the standard at any time, and most subjects checked their match by doing so. In this, as in all subsequent experiments reported, the time lapse between viewing the standard and viewing the variable was only the length of time it took the observer to move his head from one unit to the other, no more than a few seconds at most. For viewing the zenith moon, the order of the eyes-raised and the eyes-level trials was balanced for the ten subjects. Under each condition the subject was required to make a setting for the horizon moon as the standard and one for the zenith moon as the standard; again the order of presentation was balanced. For each of these four conditions two determinations were made; in one the experimenter gradually increased the size of the aperture of the variable (ascending method) and in one he gradually decreased the size of the aperture (descending method). Moreover, the order of trials with the ascending and the descending methods were also balanced. The two measures were averaged in computing the results.

The results are given in Table 1 in terms of the average ratio of aperture diameters (illusion ratio) obtained for each subject under the eyes-raised and the eyes-level conditions. It should be made clear that when the horizon moon is the standard the subject must select a zenith moon larger than the standard in order to compensate for the illusion. The illusion is here expressed by the fraction

\[ \frac{\text{Variable zenith setting}}{\text{Standard}} \]

When the zenith moon is the standard the subject must select a horizon moon smaller than the standard to compensate for the illusion. The illusion then is expressed by the fraction

\[ \frac{\text{Standard}}{\text{Variable horizon setting}} \]

These ratios are given in Table 1, as are the averages for each subject.

The first thing to note is that under both conditions a considerable illusion is obtained. Both of the mean illusion ratios differ significantly from unity at the 0.01 level of confidence. In fact, there is reason to believe the ratio would have been even greater with the time for the fact that six subjects one or more times selected the variable at the end of the available series of apertures. The magnitude of the effect obtained can perhaps be realized by comparing two circles whose diameters are in the ratio of 1.5:1, as in Fig. 4. This comparison brings out the fact that the ratios pertain to diameters, whereas the observer is comparing areas. The area ratio for the average setting in the experiment is over 2.1:1. Although many authors have estimated the moon illusion to be greater than this, it must be noted that they were depending upon memory, without the opportunity for measurement of any kind. Thus, we cannot be sure what the average magnitude of the true moon illusion is. Undoubtedly it varies for different conditions, as we shall try to
bring out later. Sufficient it to say, therefore, that we obtained a considerable illusion, which, incidentally, is not much smaller than that obtained by Boring et al.

The second thing to note is that an illusion is obtained without eye elevation—one of the same magnitude, in fact, as that obtained with eye elevation. This finding is in contradiction to the findings discussed earlier, according to which there should be no illusion at all under the eyes-level condition.

Both moons in the same region of the sky. Because of the importance of this issue, it was decided to examine the eye-elevation hypothesis in an even more direct way. The subject was asked to compare two moons located in exactly the same region of the sky (at an elevation of approximately 20°), viewing one with his eyes level and one with his eyes raised with respect to his head. (The two assemblies were mounted side by side on a table which was tilted upward at a slight angle.) Such a test is independent of other parameters which might conceivably affect the outcome, such as differences in the geographical positions of the two moons and changing cloud conditions.

The subject viewed one moon, the standard, with his eyes raised and the other, the variable, with his eyes level, or he viewed the standard with his eyes level, and the variable with his eyes raised. He made ascending and descending determinations in each case, and thus made four settings in all. The biting board was placed before one unit and mounted in such a way that the observer had to tip his chin downward and raise his eyes approximately 30° with respect to his head to view the moon. The elevation of the eyes was carefully monitored by the experimenter. In viewing through the other unit, the observers were instructed to raise their chins slightly so that their eyes would be level with respect to the head. This elevation too was carefully monitored. Again the subjects (seven men and three women) were neighbors or friends; some of them had served in the first experiment. However, they were all naive with respect to the hypothesis under investigation. The experiment was carried out in the late afternoon (as were all the other experiments reported here, except for those conducted indoors).

In summary, the subjects viewed two moons in the same region of the sky at the same elevation, viewing one with eyes raised and one with eyes level. According to the eye-elevation hypothesis, the moon viewed with eyes level should appear larger. For this reason, the illusion ratios were derived as follows. The average value for settings of the variable when the "eyes-level" moon was the standard were divided by the value for the standard; the average value for settings of the variable when the "eyes-raised" moon was the standard were divided into the value for the standard. In Table 2 these two ratios are shown for each subject (values for the ascending and the descending methods are averaged), together with the combined average.

It is clear that there is no illusion here. We therefore conclude that, under these conditions of measurement, eye elevation does not yield the moon illusion.

Experiments in Darkness

If it is safe to assume that the eye-elevation hypothesis does not explain the moon illusion, there seem to be certain facts that are difficult to understand. Many years ago, Schur (11) obtained an artificial moon illusion indoors in large rooms, such as a Zeppelin hangar, in darkness. Holway and Boring report a moon illusion obtained by means of a dense filter which obscures all objects but the sun. This illusion was also found to be a function of eye elevation. More recently, Hermans (12) found that an object at a distance of 4 feet in a light-free room appeared to be approximately 6 percent smaller in area when viewed with the eyes raised 30° than when viewed with the eyes straight ahead. An explanation such as Ptolemy’s would seem to be ruled out by such findings. Since we had been unable to demonstrate an effect of eye elevation out-of-doors, we decided to examine the hypothesis under conditions of darkness.

An experiment in a planetarium. In one experiment we sought to obtain a moon illusion in the Hayden Planetarium. We used slides with circular apertures in 1/5 projectors, taking care to have the two projectors equally distant from ceiling and horizon wall, respectively. The images were projected from the center platform to the zenith dome and to a horizon screen, each of which was 37.5 feet away. The image of the standard subtended a visual angle of about 1.5°. A series of eight comparison slides was used. The subject stood beside the two projectors so that he was approximately in the center of the planetarium. The room was dark. Even the small red fire lights were put out for a moment, and each projector was covered over, except for the lens. The instructions to the subjects and the measurement procedure were similar to those in the experiments described earlier, except that no attempt was made to control eye position. The average ratio for five male subjects, naive with respect to the experiment, was 1.03.

An experiment in a dark field. We were at a loss to explain why—in the light of Schur’s data—we did not obtain an illusion. It can be argued, however, that even if Schur’s experiment had also failed to yield an effect, such experiments do not preclude the possibility of obtaining an illusion under conditions of total darkness where the object seen is at optical infinity. The effect might not make its appearance when the objects compared are at a (discriminably) finite distance. Therefore we decided to ascertain whether an illusion is obtainable in total darkness when the objects compared are at optical infinity.

To do this, we simply placed the collimator apparatus in a light-free room and had the subject view one disk
straight ahead and the other over his head. In this experiment the apertures were viewed directly through the lens, since there was no need to use the combining glass. In order that the subject should not be prevented from having an impression of great distance through prior perception of the wall and ceiling of the room, he was brought in blindfolded and did not open his eyes until the room was dark. He could not see the apparatus. He was seated on a chair with a head rest at the top that enabled him to rest his head tilted backward at an angle of 60° from the vertical. This meant that he had to raise his eyes 30° to view the overhead disk. The collimator assembly was mounted on a board placed vertically directly over the subject's head. The other assembly was mounted horizontally on a table in front of him, and there was a nose-rest for positioning his head. To minimize any glow from the scattering of light on the surfaces of the lens, the outer surface of each lens was masked with black paper, except for two holes of 1/4-inch diameter placed 2.5 inches apart from center to center. By properly positioning the subject's head and having him check with each eye separately, it was possible to guarantee binocular viewing of the disk through the two holes. In other respects the procedure was like that of the experiments already described. Each subject made eight comparison settings, two ascending and two descending with the horizon disk as the standard and two ascending and two descending with the zenith disk as the standard. The order of presentation was balanced, as before. Ten graduate students of Yoshiva University participated; of these, seven were completely naive with respect to the experiment and three were fairly sophisticated concerning the moon illusion problem.

Table 3 gives the average illusion ratios for the horizon standard and the zenith standard. Although the ratio for most of the subjects is greater than 1 and the mean ratio of 1.03 is significantly different from unity, the magnitude of the effect is negligible (note the similarity to the results of the planetarium experiment). Whatever factor is operating here certainly cannot account for the moon illusion, since it leads to a barely discriminable difference in size. The effect is so slight that it could easily be due to an artifact, such as the diminution of the intensity of the image (13) through partial coverage of the pupil by the eyelid when the eyes are raised (4). Such an effect might show up in a dark-field experiment where other factors are eliminated, in spite of our failure to demonstrate an effect of intensity (as discussed later), or an effect of eye elevation under outdoor conditions. Our finding does confirm that of Hermans concerning eye elevation, and the magnitude of our effect is quite similar to his when the proper conversion from an area comparison to a diameter comparison is made (14). But although we confirm his finding, our observation fails to support his prediction that eye elevation would be found to yield a sufficiently large illusion to account for the moon illusion when the objects compared are at distances greater than 4 feet—the distance employed in his experiment.

In view of the negligible effect obtained, we may conclude that we do not substantiate Schur's findings, or the findings of Holway and Boring on a sun illusion with a dense filter. Although we cannot account for the difference, we do feel that our experiment constitutes a crucial test for any effect based on purely egocentric considerations, such as eye or head position. Since the subject tilts his head back 60° in viewing the overhead disk, this experiment is a test of the role of head elevation as well as of eye elevation. The results also can be taken to rule out gravity as a factor in the moon illusion, since it is obvious that two disks do not appear appreciably different in size simply because they are oriented in different directions with respect to gravity.

Color and Brightness

We have shown that the illusion cannot be understood in terms of an egocentric definition of the different positions of the moon or in terms of the direction of gravity. We have also shown that there is no reason to reject the alternative possibility—namely, that the important factor is the presence or absence of a visible terrain between observer and moon. Before we present additional evidence bearing on this alternative, however, it might be well to consider two other possibly significant differences between the horizon and the zenith moon that have often been noted: (i) the horizon moon occasionally appears reddish in color, whereas the zenith moon does not; and (ii) a full moon rises over the horizon at sunset and therefore appears less luminous than the zenith moon, seen against a dark sky. Although our knowledge about the effects of such differences does not lead to any clear-cut predictions concerning apparent size, these are nevertheless differences which must be examined.

Color. The red color of the horizon moon is a result of selective scattering of light of shorter wavelengths in the atmosphere. That the harvest moon appears large and red is well known, and the setting sun is redder than the zenith sun and seems enormous. To determine the role of color, an experiment was performed in which, for each subject, our artificial horizon moon appeared red during half the measurements and of normal color during the other half. The color was achieved by placing a blue Kodak Wratten Filter No. 29 between the light source and the combining glass. The effect was not unlike that of the setting sun.
The apparatus in this and the subsequent experiments differed slightly from that described previously in that a smaller collimator was used, with a smaller front-surface, half-silvered, 2-by 2-inch combining glass which could be viewed through only one eye (see Fig. 5). The observer placed his eye directly in front of the glass so that the edges of the glass were barely visible in the peripheral field. With the other eye he viewed the sky directly; he was not aware that the artificial moon stimulated only one eye. The artificial moon appeared infinitely distant, as was the case when the binocular apparatus was used. The new apparatus had the advantage of compactness, since each unit was enclosed in a tube 2 inches in diameter and 7 inches long (15). Each unit was mounted on a tripod, and this was set at heights which permitted convenient viewing of the horizon moon and the zenith moon, respectively, from a standing position. The apertures in this apparatus varied from 0.015 to 0.132 inch in diameter. There were 16 sizes, which increased as a geometric progression. The aperture of the standard was either 0.047 or (in other experiments) 0.055 inch in diameter, subtending visual angles quite close to the visual angle subtended by the moon. The series of apertures provided a much greater range of sizes for comparison than was available with the earlier apparatus. The extreme values provided for deviation from the standard by a factor of approximately 3. Aside from the difference in apparatus, this experiment was similar to those described earlier. The horizon moon was viewed across Mitchel Air Field from the Hofstra College campus, a visible distance of about 2 miles on a clear day. Each subject was given two trials with the horizon moon red and two trials with the horizon moon of normal color (with the horizon moon as the standard in one of the two trials, the zenith moon as the standard in the other). The order of presentation was balanced. The zenith moon was normal in color for all trials. The results for seven male students were as follows: the mean ratio for horizon moon of normal color was 1.37 (standard deviation, 0.22); the mean ratio for horizon moon red was 1.34 (standard deviation, 0.19). There is thus no evidence that color can even partially explain the moon illusion, and the reddish color which the horizon moon often has must be a coincidental concomitant of the phenomenal size.

Brightness. As noted above, there are also certain differences in the appearance of the moon with respect to brightness. It has been reported that brightness affects apparent distance (16), and thus there would seem to be some basis for Berkeley's belief (stated in 1709) that the moon illusion is based on the fainter appearance of the horizon moon (17). Berkeley reasoned that, through experience, we have come to associate distant visible objects which

Fig. 5. Cross section of the improved moon-illusion instrument. The combining glass was at a 45° angle to the optical axis. The drum contains apertures of various sizes which can be moved successively in front of the diffusing reflector by turning the selector knob. The aperture selected is indicated by a number engraved on the knob. The collimating lens, 2 inches in diameter, is set at a point approximately 5 inches from the aperture (this distance is its focal length). The filter channel makes it possible to insert a colored or a neutral density filter. The entire instrument can be threaded into a conventional camera tripod.
appear dim with large tangible objects. But there is the contradictory laboratory finding that dimmer objects actually look smaller than brighter objects of the same size (13). In fact, we have speculated that this might explain why we obtained a slight illusion in our dark-field experiment. This is an effect opposite to the one Berkeley suggests.

The full moon is two or three times brighter than the horizon moon. It appears dimmer because the sky is darker later at night (that is, it appears brighter by contrast, and the brighter moon is due less than Berkeley thought to greater attenuation of the atmosphere in viewing the horizon moon), but the moon illusion exists whether or not the moon is full, and in general, the moon appears on the horizon at all times of day. Therefore, the horizon moon appears bright and dim equally often. Hence, there is little support for the brightness theory in purely logical considerations. Nevertheless, we performed an experiment analogous to the one on color. There was only one difference: in this case, for half the trials we reduced the brightness of the horizon moon by a factor of 0.5 or by placing a half-silvered mirror between light source and combining glass. The horizon moon then appeared about half as bright as the zenith moon. For the other half of the trials the two moons were of equal brightness. The sky characteristics changed during the experiment from heavy cumulus clouds to clear, but the appearance of the sky was always the same for the two conditions of the horizon moon for each subject. The mean illusion ratios for eight Hofstra students (seven male and one female) were as follows: horizon moon of normal brightness, 1.41 (standard deviation, 0.30); horizon moon dim, 1.40 (standard deviation, 0.27).

We performed another experiment in which the difference in appearance approximated somewhat more closely the difference between the luminous zenith moon and the dimmer horizon moon. Both moons were set at the zenith, but one combining glass was placed before a double polaroid filter set in the positions of maximum attenuation (90° out of phase). This caused the sky to appear a dark blue-purple, and caused the artificial moon to seem luminous against it. The filter was 6 inches square; its sides were attached to black cloth which hung down over the observer's head on all sides so that he could see the sky only through the filter. The other moon was viewed against a normal daylight sky. The aperture in each unit was set at 0.047 inch, and the observers were asked to say whether the disks were equal or, if they were not, to say which was larger. Of ten graduate students of Yeshiva University, six said the disks were equal and four said the luminous moon was larger. This experiment therefore tends not to support the explanation of the moon illusion in terms of brightness.

There is thus no evidence that brightness can even partially explain the moon illusion. The dimmer appearance of the full horizon moon is, therefore, also a coincidental concomitant of its phenomenal size (18).

(This is part 1 of a 2-part article)

References and Notes

1. Although there has been some controversy as to the exact meaning of Simmel's law, we are interpreting it to mean that the apparent size of an image of constant angular size is a function of perceived distance. Recent research by Simmel (1935) has shown that size perception is based on the image area per unit of distance has been challenged by various authors (H. E. Gruber, Am. J. Psychol. 67, 411 (1954); W. L. Jepson, J. Parke, A. Canev, Psychol. Bull. 58, 491 (1935)); it has been found that perceived distance may be measured by judgments of distance does not correspond to the variations in "distance" demanded by changes in perceived size. We believe this challenge stems from a misunderstanding of the meaning of "distance" in this context. What is taken into account is the distance as perceived by the observer. This distance is determined by certain stimulus correlates, regardless of whether or not the observer is always consciously aware of such correlates. Stimulus correlates that normally accompany changes in distance may be represented by the brain and automatically affect size perception without conscious recognition of this effect on the part of the observer. For a recent illustration of this fact, see E. G. H. Heiderman, E. T. Orlin, J. Nachmias, Am. J. Psychol. 72, 25 (1959). Hence, distance correlates which affect Judgments of distance may not be accurately reported in judgments of distance.

9. A collaborator in a positive converting lens with a small field of view when its focal point. The rays of light coming through the lens from any point on the surface are then parallel, so that the image, which is formed, is larger by a large. The large is obtained with different apparatuses. We have also performed a little experiment to demonstrate that varying the distance of the observer from the combining glass has no effect; such a procedure should have an effect if the glass were playing any role at all.
13. Herrmann interprets his findings in terms of horizontal movements associated with convergence. When the eyes are raised, one piece of evidence supporting this interpretation (and contradicting our speculation concerning change in intensity) is that we obtained a significant positive correlation of 0.52 between the experiment (as described in the experiment) and the angle of the collimator was used, so that the disc could be seen with only one eye. This suggests that the convergence was of the small size of the effect, however, it would seem wise to suspend judgment as to the value of the experiments can be undertaken with a greater number of subjects, to test these hypotheses.
14. The new units were designed to our specifications by Erwin Streicher of Research Laboratories.
17. Part 2 of this study will appear in the next issue of Science. A portion of the data was completed while we were both at the New School for Social Research. Some of the findings were reported at the meeting of the Eastern Psychological Association of April 1960. We wish to thank J. M. Chamberlain for making it possible for us to conduct experiments at the Harvard University and J. F. Chappell for permitting us to experiment on the Hofstra College campus. We also wish to thank David Bergman, Robyn Poulain, and Stanley Richin for their help in performing certain of the experiments.
The Moon Illusion, II

The moon’s apparent size is a function of the presence or absence of terrain.

Irvin Rock and Lloyd Kaufman

The horizon moon appears to be larger than the zenith moon; this is called the moon illusion. In the last issue of Science (1) a new technique for studying this illusion was described. It consists of a device which permits the observer to view a disk of light or artificial moon on the sky. The size of the disk can be varied. Using two such devices the observer can compare a standard disk set in one position of the sky (for example, on the horizon) with a variable disk set in another position (for example, at the zenith) (see Figs. 1 and 2). The variable selected by the observer to match the standard in size gives a measure of the magnitude of the illusion. Experiments carried out with this technique failed to support the earlier finding that the illusion was based on the elevation of the eyes with respect to the head (2). It was also shown that the illusion was not based on changes in the color or brightness of the moon. Here, in part 2, we discuss work on the apparent-distance hypothesis.

The Apparent-Distance Hypothesis

We are now ready to consider the hypothesis that the illusion is based on the sense of great distance which the observer has when viewing the moon directly above the horizon. This sense of distance is created by the terrain, which, for present purposes, may be defined as a stimulus which yields the impression of a plane receding from the observer. (It should be noted here that the sense of distance or apparent distance need not necessarily correspond to the subject’s report of distance. This point was discussed earlier (1, ref. 1) and is covered in more detail later.) We have already shown that the moon illusion cannot be explained by factors operating in complete darkness, even when the moons compared are at optical infinity. This finding can be taken as supporting the apparent-distance hypothesis, because, in the case of the ordinary illusion, the zenith moon is essentially a disk at optical infinity surrounded by a homogeneous field. Therefore, in effect, the dark-field experiments may be said to eliminate the visible terrain in viewing the horizon moon, and in so doing, to abolish the illusion.

The effect of obscuring the terrain from view. We also tested this deduction somewhat more directly by obscuring the terrain under outdoor conditions. The observer compared an artificial moon set near the horizon with another artificial moon seen through a 1/4-inch aperture in a large cardboard mounted in front of the second instrument. The latter “reduction moon” was located approximately 10° above the horizon to make sure that no part of the terrain would be seen through the opening. The observer was required to place his head against the cardboard and look through the aperture at the moon reflected by the combining glass. Because the reduction moon could be viewed with only one eye, the observer was required to view the unobstructed horizon moon with one eye. The observer also compared the reduction moon with the (monocularly viewed) zenith moon. For this purpose the assembly for viewing the unobstructed moon was tilted back on the tripod so as to locate the disk in the zenith. With these exceptions the procedure for each of these comparisons was identical to procedures followed previously. Ten male students from Hofstra College were used as observers. They viewed the scene across Mitchel Air Field; the sky in that direction was clear throughout the afternoon of the experiment.

The mean ratio obtained in a comparison of the normal moon with the reduction moon was 1.34 (standard deviation, 0.08), where the reduction moon was considered to be the zenith moon of the previous experiments. In other words, we obtained an illusion with both moons in a horizontal direction merely by eliminating the visible terrain in viewing one of them. Moreover, the magnitude of the effect was about the same as that obtained, with other subjects, for the ordinary moon illusion for this same scene (1.38), under similar sky conditions. It is worth noting in passing that this result, obtained with monocular viewing, fails to support the conclusion of Taylor and Boring (3) that binocular vision is essential for a moon illusion. We know of two monocular individuals who report experiencing a moon illusion. Binocular viewing should not be crucial, according to the apparent-distance hypothesis, because the important stimulus to distance, the terrain, is received monocularly. The mean ratio obtained in a comparison of the reduction moon with the zenith moon was 0.99 (standard deviation, 0.04). There is thus no illusion when the terrain is blocked from view.

Reversal of the illusion by means of mirrors. If the presence of the visible terrain is indeed crucial, as the foregoing evidence suggests, it should be possible to reverse the illusion by giving the zenith moon a terrain, so to speak, and at the same time depriving the horizon moon of its terrain. In other words, if one were to see the zenith moon over a horizon at the end of an apparent terrain, it should appear larger than a moon viewed horizontally but not in connection with a terrain—a moon surrounded on all sides by sky.

We achieved this condition by requiring the subject to view each artificial moon through a right-angle prism, which is essentially a mirror at a 45° angle to the line of sight. The observer is seated with his back to the terrain scene he is to view. To see one moon he tilts his head and eyes upward to 90° so as to view the scene through the prism. The prism opening is 5/4 inches long and 1 3/8 inches high. The observer places his eye as close as possible to the prism. Directly behind the prism and...
Fig. 1. The artificial moon as it would be seen by an observer looking at the horizon through the combining glass, with one eye. The observer would view the scene directly with his other eye; thus any disturbing images of the edges of the combining glass, or of the clamp, would tend to be washed out.

off to one side is the small combining glass which reflects an artificial moon so as to make it appear within the mirrored scene. The observer then sees the terrain stretching vertically upward. The artificial moon is made to appear on top of the perceived horizon. To see the other moon, the observer looks through a second prism (below a combining glass), which reflects the zenith sky in a horizontal direction. Hence he sees the artificial moon straight ahead, in a horizontal direction, but instead of seeing terrain below it he seems sky surrounding it.

This particular experiment was performed on the roof of the Graduate School of Education Building of Yeshiva University, on 57th Street in Manhattan. Because it was not necessary to remove the apparatus daily, as was the case on the Hofstra campus, the instruments were attached to a wooden framework. The zenith unit was clamped to a horizontal board, which the observer could view from a sitting position. The horizontal unit was attached to a vertical board, which the observer could view from the same sitting position. The framework provided a firm anchorage for the right-angle prisms used in this experiment. The view facing east from the roof included the street, lined with buildings, and the horizon at the end of 57th Street, which was at a distance of well over a mile (Fig. 3).

Ten students of Yeshiva University were used as subjects. To compare the illusion obtained with the mirrors with that obtained under normal conditions, we had the subjects of the mirror experiments view the same scene without the mirrors. This control condition indicated the magnitude of the illusion to be expected under conditions that were comparable except for the use of mirrors. Half of the subjects viewed the mirror scene one day and the regular scene the following day. For the other half the order was reversed. The procedure of measurement was otherwise identical to that employed in the previous experiments.

The mean ratio for the mirror experiment was 1.37 (standard deviation, 0.28) and for the control variation, 1.56 (standard deviation, 0.25). The difference is statistically significant. Because the scene of a city street with buildings surrounding the horizon sky might be considered a special case (and certain facts support this view), the experiment was also performed at the Hofstra College site, with minor variations in the physical arrangement of the apparatus. For nine subjects the mean ratio was 1.34 (standard deviation, 0.25). This value is significantly lower than the ratio obtained without mirrors at the same location and under similar cloud conditions—namely, 1.54 (standard deviation, 0.19). The ratios obtained with and without reversal are strikingly similar in the two experiments.

The results thus show that we were successful in reversing the illusion, although the magnitude of the effect obtained is not as great as that of the ordinary illusion. Does this mean that the visible terrain is not the whole story—that some other factor, such as angle of regard, also plays a role? It must be borne in mind that, from the standpoint of an angle-of-regard theory, not only should the illusion not have been reversed but the true horizon moon should have continued to appear larger. Hence, if both factors were operating and were of equal strength, we should

Fig. 2. The artificial moon as it would be seen by an observer looking at the zenith through the combining glass. The dark regions in the two pictures are an out-of-focus image of the clamp that holds the combining glass.
expect them to cancel each other out, because they are in opposition. The obtained reversed ratios of 1.37 and 1.34 in the two experiments can then only mean that if an angle-of-regard factor were involved, it must have exerted only a very weak influence. More plausible, therefore, in the light of this reasoning and the evidence cited earlier against an angle-of-regard theory is the conclusion that the reduction in the size of the visible field, as the observer looked through the tank prism, reduced the impression of depth yielded by the scene. (Also, the frame of the prism may have provided a constant reference system for judging visual angle.) Moreover, looking up at a landscape aligned perpendicular to gravity is unnaturally, and this may have been a factor. If these conjectures are correct, one may say that the effect nevertheless obtained is very impressive indeed.

It seems clear, then, that it is the presence of terrain in one case and the absence of terrain in the other that is the major factor in the moon illusion. But the objection can justifiably be made that this in itself is not sufficiently analytical proof of the apparent-distance hypothesis. Perhaps the presence of the terrain stimulus pattern adjacent to a moon creates the effect for reasons other than that the pattern yields a sense of great distance. Although there may be no obvious rationale for such an effect, it still must be established that is the distance aspect of the terrain stimulus which is crucial.

An inverted terrain. It is a fact known to psychologists in the field of perception that an inverted photograph of a landscape often loses much of its effect of depth (Fig. 4). Although this is as yet unexplained, there is no question about the fact, and we decided to make use of it to test the apparent-distance hypothesis. If an observer were to view an inverted scene he would have a sense of less distance to the horizon moon than he has in viewing the scene without inversion. Hence, according to the apparent-distance hypothesis, the illusion should be diminished. Yet the terrain stimulus pattern would remain adjacent to the horizon moon, and thus, if the crucial factor is some aspect of the terrain pattern other than distance, the illusion created should be undiminished.

The observer viewed the 57th Street scene through two large Dove prisms, each 1 9/16 by 1 1/16 inches in cross section, mounted side by side directly in front of the combining glass. The subject sat with his eyes close up against the prisms. The prisms were mounted in a thick cardboard in which a hole had been cut equal in size to the cross-sectional area of the two prisms combined. The cardboard thus surrounded the prisms on all sides, serving as a shield which prevented the observer from seeing the scene in any way except through the prisms. He viewed the zenith moon normally, without prisms. To compare results of observations with and without inversion of the scene, a control condition was included in which the observer viewed the scene through an aperture equal in size to the two prisms combined—that is, an aperture 3 1/8 by 1 1/16 inches. Four subjects were tested first under the experimental condition and six subjects first under the control condition. Otherwise the procedure was identical to the measurement procedures described previously.

The mean ratio for ten naive subjects was 1.66 (standard deviation, 0.32) without the prisms (4) and 1.28 (standard deviation, 0.17) with the prisms. (Three of the control subjects selected apertures at the upper end of the series, so the mean ratio of 1.66 is somewhat conservative.) These two values differ significantly from one another, and the second differs significantly from unity. Thus there are two conclusions to be drawn: (i) the inversion of the scene does very appreciably reduce the moon illusion, and (ii) there is still an illusion even with inversion.

The first conclusion provides important support for the apparent-distance hypothesis. The second leaves us with an unsolved problem. It is probable that the inverted scene does not completely eliminate a sense of depth. This conclusion is especially plausible in the case of this particular scene, which contains a perspective pattern derived from the vertical lines of buildings as well as one derived from the horizontal elements along the ground plane. The perspective based on the vertical elements is not changed with inversion. Furthermore, there are other possible cues to the scene's true depth, such as monocular parallax. We retained some sense of depth in viewing the inverted scene. But there is another factor to be considered in the case of this particular scene. The moon is seen between tall buildings. Thus, it is framed on three sides, and this frame of reference might very well affect the moon's apparent size (5). Some additional evidence on this point was obtained in experiments conducted in the laboratory with slides of outdoor scenes. In one such experiment a slide of the 57th Street scene was shown, and, as a control, a slide of a pattern virtually identical to that scene with respect to line elements but
drawn so as not to convey a sense of three-dimensionality. The moon seen in the control slide was, therefore, also framed on three sides. This slide yielded an illusion ratio close to 1.2, quite similar to the ratio obtained with the inverted scene. Such a relational effect would be in no way changed by inversion of the scene. Thus, it is possible that the slight illusion obtained with the inverted scene is due to a residual depth impression or a relational effect of the surrounding buildings, or both. It would be instructive to repeat this experiment with slides of a more typical landscape, such as the scene at Hofstra College, which does not produce a framing effect.

Various authors have commented on the apparent destruction of the moon illusion that occurs when an observer views the horizon moon between his legs. Boring interpreted this as evidence supporting the angle-of-regard hypothesis. Our finding of a diminution of the illusion with an inverted scene (which does not involve any change in angle of regard) suggests an alternative explanation: inverting the head brings about an inversion of the retinal image, and the latter inversion, for whatever reason, diminishes the impression of depth yielded by the landscape. (It might also be pointed out that looking between one’s legs inevitably lowers the observer’s vantage point. Looking at the terrain from a point nearer the ground would also decrease the apparent distance to the horizon.)

Variation of distances and cloud conditions. It should follow from the hypothesis under investigation that the moon illusion would increase with apparent distance to the horizon. We therefore performed an experiment in which the illusion was compared for two scenes which differed substantially in the impression of distance to the horizon. One scene extended north from the Hofstra College campus and encompassed Mitchel Air Force Base; for this scene the apparent distance from the viewing point to the horizon was roughly 2 miles. The other scene was 20° west of the first. In this direction trees and shrubbery obscure the horizon at a distance of about 2000 feet. Thus, the apparent distance (D1) for the first scene was much greater than that for the second (D2).

Another factor which might conceivably contribute to the differential impression of distance to horizon and zenith is the presence of clouds, as Helmholtz (6) and others have speculated. In fact, Miller (7) found the half-arc angle to vary inversely with the degree of cloudiness. We therefore decided to include a test of the effect of cloudiness on the illusion. This was done by testing different subjects on totally clear days, on totally overcast days (with structured stratuscumulus cloud coverage), and on days with broken coverage (that is, with clouds predominantly cumulus, and with coverage judged to be between 0.3 and 0.7). It was expected that the illusion would be maximal on totally overcast days, minimal on totally clear days.

The design of this experiment involved testing subjects under six sets of conditions—combinations of the two apparent distances and the three types of cloudiness. It was not feasible to use subjects as their own controls by testing them under all conditions of cloudiness, because we obviously could not manipulate the cloud conditions at will (although we were able to test some subjects for both distances). Altogether, 55 Hofstra College students, male and female, served as subjects, 20 on overcast days, 20 on days with broken cloud coverage, and 15 on clear days. Half the subjects tested under overcast and broken-cloud conditions viewed the horizon moon over scene D1 and half over scene D2. Five of the subjects tested under the clear-sky condition viewed the moon over both scenes, and the remaining ten viewed it over one or the other. In this experiment each subject made two ascending and two descending series of matches for each moon that served as the standard. The procedure followed was otherwise the same as in the other experiments. The zenith moon was set at an elevation of 30°.

The results are given in Table 1 in terms of the average ratios for each of the six subgroups and for overall distance and overall cloud conditions. The data of Table 1 reveal an increase in the illusion with increasing cloudiness and a greater illusion with greater apparent distance to the horizon (D1). An analysis of variance shows that the cloud-condition ratios differ significantly, as do the distance-condition ratios. (As noted above, five of the subjects were tested for D1 as well as for D2. An additional four subjects were also tested for both D1 and D2. In these tests the sky was clear. For eight of the nine subjects the illusion was larger for D1. For the ninth subject the illusion was the same for D1 and D2. The average for these nine subjects was 1.25 for D1 and 1.14 for D2.) These results support the apparent-distance hypothesis and confirm Helmholtz’s speculations on the role of cloudiness, as well as the findings of Miller, Neuberger (8), and others.

Error of the Standard

In most of the experiments reported in this article the illusion obtained was greater when the horizon moon was the standard than when the zenith moon was the standard. This is apparent in Table 2, where the results are given separately for the two cases for all experiments. Only in the eye-elevation experiment in which the "binocular" collimator was used, and in the experiments in which no illusion was obtained, does this difference fail to appear. We would be inclined to believe that the exceptional result in the former case is a function of the slight differences in the apparatus were it not for the fact that in other experiments, not reported in this article, in which the binocular collimator was employed a similar effect was obtained. The absolute magnitude of the effect is considerably smaller than that of the moon illusion itself. A rough approximation is yielded by dividing by 2 the average difference between

<table>
<thead>
<tr>
<th>Variable zenith setting</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>Variable horizon setting</td>
</tr>
</tbody>
</table>

for those experiments where such an effect occurred. This yields a value of approximately 15 percent; that is, a comparison object would have to be made 15 percent greater in size than a standard object if only such an error were operating.

One way of viewing this finding is in terms of a tendency to exaggerate the standard; this tendency has in recent years been discovered by others working on size judgments and called "the error of the standard" (9). In our experiments this tendency would increase the magnitude of the illusion when the horizon moon is the standard because it makes the already phenome-
nally large moon seem even larger. It would decrease the magnitude of the illusion when the zenith moon is the standard because it makes the smaller-appearing zenith moon seem larger, thus offsetting the illusion to some extent. This constant error might be considered to be a positive time error for size except for two considerations. (i) It has not been experimentally demonstrated that the error is a function of the order of presentation, only that it is associated with the stimulus made to serve as the standard. In fact, in many of our experiments the subject was allowed to check the standard after setting his variable. (ii) The effect does not seem to appear in our experiments when the moon illusion itself does not appear, although this should provide an ideal opportunity for observing the operation of a time error if one exists.

The fact that, in our experiments, the effect does not seem to be present unless the moon illusion itself is present suggests another interpretation relating to certain phenomenal differences between the horizon and the zenith moons. We will therefore return to this problem.

Discussion

Methodology. Since our results on the matter of eye elevation fail to substantiate previous findings, the difference in method employed becomes crucial. In addition to the points made earlier concerning our reasons for dissatisfaction with the method used by Boring and his colleagues, we would like to make a comment concerning our method. Assuming that viewing the sky through glass does not affect the results—an assumption that we think justified (see §, note 10, to which we may now add the observation that in our dark-field experiment the observer does not see the apparatus and is looking directly at the artificial moon—we believe we have duplicated the conditions of the moon illusion in nature)—comparison of a moon in one region of the sky with a moon in another region. Our observers merely have to compare one moon with the other, they do not have to compare either moon with anything else. The one remaining difference between our experimental conditions and the conditions in daily life is that of immediate versus delayed comparison. But this is a difference which we deliberately introduced in order to eliminate any dependence on memory; impressions of the moon illusion in daily life may be somewhat spurious because of the unknown role of memory.

In support of our contention that we have duplicated the conditions found in nature, we performed an experiment in which three observers were asked to compare the real horizon moon (viewed over the ocean) with our artificial moon pointed at the zenith. The average illusion ratio obtained was 1.83, a value slightly inflated by the lack of a control for an error of the standard. Observations by these same subjects yielded no illusion ratio whatever when the artificial moon was pointed at the horizon but 40° to one side of the real moon. In the latter comparison the subjects selected an aperture identical to the one they had selected when the artificial moon was directly superimposed on the real moon. In other words, the aperture of our apparatus, known to subtend approximately the same visual angle as the real moon, yielded a phenomenal disk equal in size to the phenomenal moon when the two were seen at the same elevation. But when that aperture was viewed at the zenith it appeared much too small. These checks demonstrate the phenomenal equivalence of our artificial moon and the real moon.

If our reasons for questioning the method used by Boring and his associates are valid, and if our method is indeed a duplication of the illusion as it exists in nature, two problems remain unsolved: how their observers were able to arrive at a satisfactory match, and why these matches revealed a moon illusion dependent on eye elevation. It is a fact worth noting that, for the most part, either Boring, his colleagues, or other persons familiar with the problem under investigation served as subjects. A more serious contradiction exists, however—one between our findings and those of Holway and Boring in experiments carried out with their direct-comparison (reflected mirror-image) method. The contradiction is serious because, as we have noted, the essential conditions of the moon illusion are successfully duplicated in this method. Boring and his colleagues obtained only a verbal estimate of the difference in size, and again it should be noted that the observers were familiar with the problem under investigation. The same is true for the findings of Holway and Boring concerning an illusion of the sun seen through dense filters. Nor can we shed any light at this time on Schar's findings of an illusion based on differ-

<table>
<thead>
<tr>
<th>Distance</th>
<th>Clear Mean</th>
<th>Clear SD</th>
<th>Broken Mean</th>
<th>Broken SD</th>
<th>Overcast Mean</th>
<th>Overcast SD</th>
<th>Overall Mean</th>
<th>Overall SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Near (D2)</td>
<td>1.28</td>
<td>0.17</td>
<td>1.35</td>
<td>0.27</td>
<td>1.45</td>
<td>0.21</td>
<td>1.36</td>
<td></td>
</tr>
<tr>
<td>Far (D1)</td>
<td>1.40</td>
<td>0.22</td>
<td>1.54</td>
<td>0.19</td>
<td>1.58</td>
<td>0.28</td>
<td>1.51</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>1.34</td>
<td>1.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.52</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Average illusion ratios for various distances and cloud conditions (ten observers for each cloud-and-distance condition).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Horizon Average</th>
<th>Zenith Average</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eyes raised versus eyes level</td>
<td>1.46</td>
<td>1.51</td>
<td>1.48</td>
</tr>
<tr>
<td>Eyes level</td>
<td>1.47</td>
<td>1.46</td>
<td>1.46</td>
</tr>
<tr>
<td>Both moons in same region of sky</td>
<td>1.07</td>
<td>1.01</td>
<td>1.04</td>
</tr>
<tr>
<td>Dark field</td>
<td>1.03</td>
<td>1.03</td>
<td>1.03</td>
</tr>
<tr>
<td>Color (red and white horizon moons combined)</td>
<td>1.49</td>
<td>1.23</td>
<td>1.36</td>
</tr>
<tr>
<td>Brightness</td>
<td>1.53</td>
<td>1.28</td>
<td>1.40</td>
</tr>
<tr>
<td>Obstructed versus unobstructed terrain</td>
<td>1.41</td>
<td>1.27</td>
<td>1.34</td>
</tr>
<tr>
<td>Obstructed terrain versus zenith</td>
<td>1.02</td>
<td>0.97</td>
<td>0.99</td>
</tr>
<tr>
<td>Mirror reversal</td>
<td>1.54</td>
<td>1.20</td>
<td>1.37</td>
</tr>
<tr>
<td>57th Street</td>
<td>1.79</td>
<td>1.33</td>
<td>1.56</td>
</tr>
<tr>
<td>Control</td>
<td>1.49</td>
<td>1.20</td>
<td>1.54</td>
</tr>
<tr>
<td>Hovstera</td>
<td>1.37</td>
<td>1.19</td>
<td>1.28</td>
</tr>
<tr>
<td>Inverted terrain</td>
<td>1.95</td>
<td>1.38</td>
<td>1.66</td>
</tr>
</tbody>
</table>

Table 2. Average ratios given separately for the horizon standard, the zenith standard, and average.

The various distances and cloud conditions:

- **Long distance (D1):**
  - Clear | 1.56 | 1.23 | 1.39 |
  - Broken | 1.27 | 1.33 | 1.54 |
  - Overcast | 1.73 | 1.43 | 1.58 |

- **Short distance (D2):**
  - Clear | 1.40 | 1.15 | 1.27 |
  - Broken | 1.44 | 1.25 | 1.34 |
  - Overcast | 1.61 | 1.28 | 1.44 |

*Moon viewed with eyes level taken as the standard. **Moon viewed with eyes raised taken as the standard.
ences in perceived direction inside darkened buildings (10). As already noted, it would seem that our dark-field experiment is the ideal test for such an effect, yet the result we obtained was only negligible.

Very recently Leibowitz and Hartman reported an experiment (11) in which subjects in a darkened theater made size comparisons of disks seen overhead with disks seen straight ahead. The disks were 35 feet away. The overhead disk was underestimated by 19.1 percent by adults and by 32 percent by children of 5 to 8 years of age. We are at a loss to explain this finding, in the light of our planetarium and dark-field experiment, except to note that some stray light from the projector enabled the observers to detect the outlines of chairs on the ground (12), and that the cues for distance in the horizontal direction were thus probably better than those in the vertical direction. There is nothing but empty space between observer and overhead disk. (The same point is relevant to Schur's experiment.)

Leibowitz and Hartman obtained similar results outdoors with a disk suspended outward from the roof of an 85-foot building. The latter finding could also be a function of the superior cues for distance along the ground or a function of the "framing" of the horizontal disk by the wood backing and by objects behind it.

The context effect. It seems quite clear from the various experiments reported here that a visible terrain is essential for the appearance of the illusion. Have we demonstrated that the terrain produces the illusion because it increases the perceived or registered distance of the horizon moon? There were two findings in support of this conclusion: (i) when the horizon appeared farther away the illusion increased; and (ii) the illusion decreased when the terrain pattern was inverted, presumably because the impression of depth decreased. Logically the only alternative to the distance hypothesis is the theory that the terrain pattern, as a two-dimensional structure or context, increases the apparent size of a disk seen adjacent to it as compared to the apparent size of a disk seen within a homogeneous surround. On the face of it this alternative is not a particularly plausible one, since the typical terrain lies entirely to one side of the moon—what is, it does not frame the moon except in the case of scenes containing tall buildings or the like. It is unlikely that, under these circumstances, such a context effect, even if it existed, could approach in magnitude the moon illusion obtained. In any event, we ruled out this possibility in experiments in which we sought to achieve an illusion indoors by means of slides of terrain patterns. On the whole, only a negligible illusion was obtained when a disk seen above the terrain on the screen was compared with a disk seen within a homogeneous surround. No illusion at all was obtained when a control slide was substituted which duplicated the terrain pattern in all structural essentials but which was deliberately drawn so as not to yield an impression of depth. If the moon illusion is a function of such a context effect, we can see no reason why it should not be easily created in the laboratory. On the other hand, if it is a function of apparent depth, one can readily see why it is difficult to create it in the laboratory. Hence, we may consider these negative results still a third piece of evidence in support of the distance hypothesis.

The size-distance invariance hypothesis. We turn now to certain theoretical questions bearing on the apparent-distance hypothesis. As noted earlier (1, ref. 1), there has been considerable dissatisfaction in the last few years with explanations of size perception based on the taking into account of distance—or with what is being called the size-distance invariance hypothesis. We need not repeat our reasons for questioning the basis for this dissatisfaction. In any case, it is not clear whether those who question the invariance hypothesis wish to argue that phenomenal size is not at all a function of distance or merely that the precise nature of the function is not known. As far as the moon illusion is concerned, our claim is not that every increment in perceived or registered distance will necessarily yield some proportional increment in the phenomenal size of the moon but merely that, in a gross way, the horizon moon appears larger because it appears much farther away, or that a very-distant appearing horizon moon looks larger than a not-so-distant appearing horizon moon.

Recently the so-called paradox concerning the relative apparent distances of the two moons first pointed out by Boring has been cited as further evidence against the invariance hypothesis (13). The horizon moon allegedly appears nearer, not farther away. We dealt with this problem earlier in terms of certain logical considerations and cited experimental evidence in support of our position (1), but it might be well to reiterate our belief that what is crucial is not distance as judged but distance as registered by the nervous system on the basis of certain stimuli. Woodworth and Schlosberg have made this very point in discussing the seemingly paradoxical results of stereoscopic studies of changes in convergence. They proposed a solution "in terms of a multilevel view of perception" (14).

"We may assume" they state, "that convergence and the resulting appropriate distance are registered at a low level of the perceptual sequence and serve as cues for judgments of size, although the cues themselves are not directly available through introspection. The size judgments then serve as cues for another judgment of distance, which may conflict with the lower-level cue." This is precisely the way in which we have tried to deal with the paradox reported by Boring. We propose that changes in phenomenal size may be a better index of changes in registered distance than of reportable changes in perceived distance. To support the invariance hypothesis one need only show that specifiable changes in registered distance (as indicated by convergence, accommodation, and so on) yield predictable changes in phenomenal size; not that changes in phenomenal or judged distance yield predictable changes in phenomenal size (15). Nevertheless, in the case of the moon illusion, when judgment can be eliminated as a factor by removing the moon from view, observers then do report the horizon sky to be farther away.

Stimulus correlates of distance. We have not tried to tackle the question of what the important stimulus correlates of distance are in the case of the moon illusion, except indirectly. The importance of clouds, and of scenes which allow one to view the horizon at very great distance, suggest that configurational properties of the stimulus are crucial, because physiological correlates cease to be important at great distances. By configurational we mean relationships within the stimulus pattern, such as perspective, interposition, and the like (16). The effect of inverting the scene supports this line of reasoning. In any case, we can rule out convergence and ac-
accommodation, because these adjustments are the same for horizon and zenith moons in daily life as well as in our experiments. One can easily eliminate other nonconfigurational correlates of distance perception, such as retinal disparity and movement parallax, by viewing the real moon with one eye and with the head stationary; an observer viewing it in this way still seems to obtain a substantial illusion—at least we do. If this reasoning is correct, and if, as is plausible, the configurational correlates are a product of experience, then the illusion itself would be indirectly dependent on experience.

The constancy function. In one respect the apparent-distance hypothesis oversimplifies the problem of the moon illusion. On the one hand the horizon moon can be said to take on the size of a region of the terrain of equivalent visual angle at the horizon. That region, in turn, has a large phenomenal size because of the constancy function—the observer's tendency to take distance into account. (This way of stating the matter is similar but not identical to the popular explanation that the horizon moon looks large because we compare it with familiar objects seen adjacent to it on the horizon. For example, the image of the moon is larger than that of a house on the horizon. Erro, the moon is at least larger than a house. The fact is, however, that familiar objects need not be present, as in the case of the moon seen over the ocean. But one can say that the

Fig. 4. The effect of distance on size. The black rectangle on the horizon appears larger than the one in the foreground, although they are identical in size. The effect would be much greater in viewing a truly three-dimensional scene, where binocular and other cues would enhance the impression of depth. Conversely, the effect can be increased by viewing the picture with only one eye, because the impression of the two-dimensional surface of the page can be somewhat reduced.

22 JUNE 1962
moon must be at least as large as an extent of water of equivalent visual angle at the horizon, and that that extent is seen to be quite large because of the constancy function. See Fig. 4.)

By contrast, the zenith moon cannot be related to any other regions of the field, and in that sense its distance is essentially indeterminate. In fact, it is more or less a reduction object. We have shown that the horizon moon viewed through an aperture appears to be the same size as the zenith moon. Although the distance of the zenith moon is indeterminate, relative to the horizon moon the zenith moon nevertheless seems to register as "nearer."

If this way of putting the matter is correct, it suggests an interesting explanation of the error of the standard, discussed earlier. When the observer views the horizon moon as the standard he approaches the zenith variable with the immediate memory of a disk, which, at least to some extent, is seen as a thing with an objective size (if not a thing of any particular linear size). When, however, the zenith moon is the standard, because his viewing of it is more of a "pure visual angle" or "pure extensity" experience, he approaches the horizon variable with a visual-angle set. It is known that such a set can reduce the tendency toward constancy (17). In the instance under discussion, it would reduce the illusion when the zenith moon is the standard. According to this interpretation, the error we obtained may be viewed as a special case and not as an instance of a more general error of the standard, as was implied earlier.

This leads to a second reason for stating that the apparent-distance hypothesis is oversimplified. As noted earlier the zenith moon is at an indeterminate distance and is therefore of indeterminate size. The horizon moon appears very far away, and objects at very great distances also are of somewhat indeterminate size (18). Ordinarily we say that distance is taken into account in a particular impression of linear size; thus, as stated in Emmert's law, a greater distance yields an impression of a particular larger size. In the case of the moon illusion we have to extend this reasoning to say that distance influences size perception (in the sense that one moon looks larger than the other) despite the fact that neither moon appears to be of any specifiable size. That is, in the case of the moon illusion it would seem that distance affects the relative-extensity experience, not a relative-linear-size experience.

Miscellaneous considerations. The moon illusion has often been cited as an example of the anisotropy of visual space, in that there is a nonequivalence of phenomenal space in different directions. In our opinion, not only is this term not clear but it presupposes something which may not be true. Anisotropy could refer to direction within two-dimensional space—could pertain to facts such as the phenomenal changes and recognition changes brought about by disorientation of shapes (for example, a square becomes a diamond when tilted 45°, or text is difficult to read upside down). Or it could refer to three-dimensional space. In either case it remains to be proved that phenomenal changes such as those just mentioned, or the moon illusion (or the presumably related fact that size constancy is more marked in the horizontal than in the vertical direction), require an explanation in which visual space is, per se, anisotropic. The two-dimensional effects can perhaps be explained in a different fashion (19), and the apparent-distance explanation of the moon illusion is based on the different content of the visual scene in the different directions. Our failure to achieve anything more than a minute effect in the dark-field experiments argues against any inherent anisotropy of three-dimensional space.

The question of whether the moon illusion is based on an experiential enlargement of the horizon moon or a shrinkage of the zenith moon has often been raised. Such a question presupposes a normative base line against which the effect can be measured. In the Müller-Lyer illusion one might compare either line seen within the arrowheads with one seen in isolation and ask, Is the illusion due to the apparent enlargement of the line seen within the outward-pointing arrowheads, to the apparent diminution of the line seen within the inward-pointing arrowheads, or possibly to both? Here the isolated line is the base line, and the question can be investigated. In the case of the moon illusion, however, there is no "normal"-size moon, and the visual angle remains constant. From the standpoint of the tendency toward size constancy, however, one might say that the horizon moon approaches the veridical size of the moon more closely than the zenith moon does. Of course, the departure from constancy is extremely large in either case, but it is somewhat less for the horizon moon. Therefore, if the illusion is defined in terms of size constancy (or the objectively true size of the moon), one must conclude that it is based on the smaller appearance of the zenith moon resulting from the inadequate registration of distance, which was discussed earlier.

As is evident from Table I of part 1 (1) and from the variability reported for many of the experiments, there are great differences in the illusion for different individuals; this was true in all experiments where an average illusion of any magnitude was found. Findings for individual observers were consistent in repetitions of the experiment (r = .82 for the experiment on the effect of distance and cloud condition) and even from condition to condition (r = .83 for the experiment on the effect of eye elevation). How should we interpret such consistent individual differences? It is possible to argue that some observers respond more readily than others on the basis of visual angle uninfluenced, so to speak, by distance. Thus, in the literature on size constancy it has often been argued that there are "analytical" perceivers who show less than average tendency toward constancy (20). Such persons presumably would not experience much of a moon illusion. Although this may indeed be the case, we are somewhat reluctant to accept the notion that the actual sensory experience of the moon's size differs for different individuals viewing the moon at the same time and in the same place. The alternative is the somewhat radical proposition that, while perceived size does not vary, the judgments nevertheless do vary because the comparison is a more difficult one to make than would be the case if two disks were simultaneously perceived on the same background at the same distance. Faced with the difficulty of making a precise match—where it is necessary for the observer to remember the size of the standard while he inspects the variable series—each observer settles on a particular value and then, in order to be consistent, continues to select a value close to his original choice. Of course, such variability centers around a value which reflects the illusion per se. It is possible that a nonperceptual process of this kind accounts for individual differences in various perceptual tasks reported in the psychological literature.
Summary and Conclusions

We have examined the two types of explanations of the moon illusion—the egocentric, in which the differences in direction of the horizon and the zenith moons are thought of in relation to different angles of regard of the observer, and the objective, in which the presence or absence of the terrain is considered crucial. The former type is exemplified chiefly by the eye-elevation hypothesis in the work of Boring and his colleagues; the latter, by the apparent-distance hypothesis based on the superior cues to distance provided by the terrain. Boring had rejected the apparent-distance hypothesis on the grounds that the horizon moon is reported as nearer, not farther away, by most observers. He then performed experiments which supported the eye-elevation hypothesis.

Our own work started with our reservations about Boring's conclusions because of (i) logical considerations and contradictory data of our own concerning the question of which moon appears to be farther away; (ii) the observation that in daily life eye-elevation does not seem to account for the moon illusion; and (iii) logical difficulties connected with Boring's method of studying the phenomenal size of the moon. We developed an apparatus which made it possible for an observer to view an artificial moon in the sky at optical infinity. The size of this moon could then be varied. Using two such units, one pointed at the horizon and one at the zenith, the observer could compare the moons directly with one another and match the variable with the standard. A series of experiments were performed with this apparatus. The major conclusions from these experiments are as follows:

1) An appreciable illusion is obtained, varying in average magnitude from a diameter ratio of 1.2 to a ratio of 1.6, depending on terrain and sky conditions.

2) Eye-elevation does not account for the moon illusion (nor, for that matter, does head elevation).

3) A minute illusion (ratio 1.03) is obtained in a completely dark field for binocularly viewed moons at optical infinity. The reliability and possible significance of this slight effect warrant further study, but it is clear that whatever produces the effect cannot be considered a factor of any importance in the ordinary moon illusion.

4) Neither apparent color nor brightness can even partially account for the moon illusion. The frequently noted reddish color of the horizon moon, or its lower brightness as compared with the brightness of the zenith moon, or both, are apparently coincidental concomitants of the phenomenal size.

5) The presence of the terrain is crucial for the existence of the illusion. The evidence is as follows: (i) The illusion disappears when the observer's view of the terrain is obstructed; (ii) the illusion can be obtained when a reduced horizon moon (here analogous to a zenith moon) is compared with a normally viewed horizon moon; and (iii) the illusion can be reversed with respect to the direction of regard by means of mirrors, so that the elevated moon seen directly above the terrain's "horizon" looks larger.

6) The apparent (or better, the registered) distance along the terrain plays a causal role. The evidence is as follows: (i) The illusion is considerably reduced when the terrain is optically inverted; (ii) the illusion can be shown to be a function of the apparent distance to the horizon and of the degree of cloudiness; and (iii) only a minute illusion can be produced indoors by means of slides that yield an impression of a terrain, although the patterns of the slides are structurally similar to outdoor patterns which do yield an illusion. The only difference would seem to be that the slides do not convey a sufficient impression of depth.

No illusion at all is produced by control slides which duplicate the structural features of terrain but which do not yield an impression of depth.

7) There is some evidence that a secondary factor contributes to the moon illusion under certain special conditions—namely, a framing or relational effect when the horizon moon is seen between buildings or other large terrestrial objects.

8) The illusion is greater when the horizon moon is the standard than when the zenith moon is the standard, a fact which may be an instance of what has recently come to be known as the error of the standard.

9) The apparent-distance hypothesis as an explanation of the moon illusion requires elaboration. (i) Distance influences the apparent size of the moon despite the fact that the moon does not appear to be of any specifiable linear size (distance here affects the relative-extent experience, not the linear-size experience). (ii) The zenith moon, while appearing less far away than the horizon moon, is to some extent a reduction object—its distance is essentially indeterminate. (iii) The observer may not be consciously aware that he is responding to a greater subjectively registered distance in viewing the horizon moon. In fact, when asked to compare the distances of the two moons, he may even judge the horizon moon to be the nearer. The latter judgment, however, depends strictly upon the relative sizes of the two moons.

References and Notes

4. The high ratio obtained in the nonintension condition despite the restricted field of view of the aperture seems to contradict the explanation offered for the reduced illusion in the mirror experiment. This is not necessarily a contradiction, however, because the aperture is not as deep as that of the right-angle prism and thus does not have to see more of the foreground. Also, different groups of subjects were used in the two control experiments, and the cloud conditions were probably not identical.
15. We are not taking the position that registered distance is the only factor determining size. One of us (I.R.) has sought to demonstrate the importance of an entirely different factor in size perception and size constancy (see 3).
16. The failure to obtain a substantial illusion with slides does not contradict this conclusion, because other distance correlates seem to operate in the direction of revealing the slide's twodimensionality.
18. If the size of the zenith moon is indeterminate and that of the horizon moon is precisely determined, one might guess that it should be as difficult to compare the two moons directly as it is to compare either with a nearby object (Boring's method). This is not the case, however, because the horizon moon, being so very distant, is not perceived as a definite linear size and is perceived, therefore, as more or less commensurate with the zenith moon.
• **Hint to Problem 5** Irvin Rock was one of the most significant psychologists who supported the so called “constructivist” approach. In the introduction of his influential book *The Logic of Perception* (Rock 1983) he wrote: “The thesis of this book is that perception is intelligent in that it is based on operations similar to those that characterize thought”. Thus, similarly to the “natural geometry” of Descartes, Rock also held that certain properties of the objects seen are calculated: “I will argue that the process of achieving constancy is one of deductive inference where the relevant ‘premises’ are immediately known. That is to say, in the case of a specific constancy such as that of size, two aspects of the proximal stimulus are most relevant, one being the visual angle subtended by the object and the other being information about the objects’s distance” (Rock 1983, p. 240). This position means subscribing to the TAD theory.

It is important to realise that the SDIH in itself does not support the claim that distance is *calculated* by the visual system. Therefore a TAD model is not necessarily supported by the experimental data supporting SDIH. (Egan 1998)

In the article there is only weak evidence for the TAD model. When investigating the framing effect of large buildings, Rock and Kaufman register a strong illusion without strong cues of great distance (see page 329). Thus either more than one mechanism is responsible for the illusion, or the mechanism suggested by the writers is not the real one.

The secondary aspect of the illusion necessitates two distance values for a TAD model. One is the registered distance, greater for the horizon moon. Size is calculated from this, and the size affects the perceived distance of the moon, which is greater for the zenith moon. This causes further difficulties in the employment of the TAD model.

The SDIH permits that, as opposed to TAD models, the cues for the visual system are size cues, used to determine the distance of objects. It is therefore possible to accept the SDIH without subscribing to the TAD model.
4 Epilogue

With increasing contemporary knowledge about the existence of two separate visual systems, probably many of our concepts about illusions will have to be reformulated (Norman 2002).

It seems that the problem-centered approach that has characterised most of the 27 centuries in search for an explanation of the illusion is slowly being replaced by a theory-centered approach. First a coherent, general account of size and distance perception has to be accepted—on such a basis the explanation of the moon illusion should simply follow from the general principles.

5 Further Problems

- **Problem 7** Compare Aristotle’s explanation to the two Ptolemaic explanations.

- **Problem 8** The theory of Kaufman and Rock has undergone several changes during the years. Point to these by comparing the original articles (Text 7 with (Kaufman & Rock 1989), or the recent (Kaufman & Kaufman 2000)).

- **Problem 9** Evaluate the arguments for and against a physiological explanation of the moon illusion in (Enright 1989) and (Roscoe 1989), as opposed to (Kaufman & Rock 1989).

6 Suggested Readings

This chapter is partly based on (Zemplén 2003). An excellent introduction to the history of the question is (Plug 1989), analysing 285 sources on the illusion. The recent (Ross & Plug 2002) gives a detailed overview of celestial illusions. It discusses the real sizes of the sun and the moon, ways of measuring the moon illusion. A whole chapter is devoted to atmospheric refraction (ch. 5), and to aerial perspectives (ch. 6). The flattened-dome theory is the topic of chapter 7. The secondary aspect as well as the angle of regard hypothesis is discussed in this detailed study.

Generally on paintings and illusion see (Dunning 1991), on optical illusions (Frisby 1980) and (Robinson 1972) also or the classical (Gombrich 1969) or (Gregory, Gombrich & Blakemore 1973), and newer articles like
(Gregory 1997). On the early history see (Johannsen 1971). The number of illusions now known is staggering, just as how early some of these appear. For an interesting early Roman beaker with an inverted face illusion see (Wade, Kovacs & Vidnyansky 2003).
Chapter XI

MOLYNEUX’S PROBLEM: A CONGENITALLY BLIND
RECOVERING THE FACULTY OF SIGHT

1 INTRODUCTION

Unlike the problem of the horizon-enlargement of the moon, the topic of
the present chapter has a very precise and specific date of birth: William
Molyneux (1656-1698) addressed a letter to John Locke (1623-1704) on 7
July 1688 in which he formulated a problem that has intrigued philosophers,
psychologist, and physiologists for the coming three hundred years.

In the letter he asked whether a congenitally blind person, capable
of distinguishing between a sphere and a cube by touch alone would, if
he suddenly recovered his sight, be able to distinguish and name the two
objects without touching them.

The problem raises questions not only about perception, but also fun-
damental issues in epistemology and the philosophy of mind, some of
which will be treated in this chapter, which investigates some of the responses and solutions given to “Molyneux’s Problem”.

2 The question

William Molyneux, an Irish supporter of the “new learning” had wide-ranging interests, among which optics had a privileged role. His *Dioptrica Nova*, published in 1692 was an important early contribution to optics in the English vernacular. His interest in blindness was probably greatly augmented by the unfortunate illness of his wife: a few months after their marriage in 1678 Lucy Domville suffered a stroke which resulted in the loss of her sight.

Another cause influencing the formulation of Molyneux’s problem was his reading of a French abstract of a book to be published in London in 1790, *An Essay Concerning Humane Understanding*, by Locke, whom he greatly admired. In the chapter on Aristotle we have already discussed the question of special and common sensibles. This distinction was still used by Locke. He claimed that as ‘special sensibles’ (like the ideas of colour, etc.) are only obtained through one of the senses, a lack of this sense (like blindness) necessarily implies the lack of the ideas belonging to that sense. Other ideas, however, are obtainable by more than one sense (like shape). A blind would thus be able to distinguish a cube and a sphere by touch, and using his tactile faculty he would have a concept of shape. Would this person, after his sight was restored, be able to see the same ideas of shape that he has learnt to distinguish by touch?

Molyneux originally addressed the question to Locke in 1688 (see Text 1). Locke has not answered the letter before he had read Molyneux’s flattering words about him in the dedication of his *Dioptrica Nova* to the Royal Society in 1692. After this he took up the correspondence and from the second edition on included the problem in his *Essay* (see Text 2, §8.).

- **Problem 1** Compare the two formulations of the problem, the original letter in Text 1 and the letter as included in Locke’s work (Text 2, §8)! What are the significant differences between them? What could explain these?
3 Answers to a Hypothetical Problem

Eighteenth century European philosophers saw no possibility to cure congenital blindness. Molyneux’s brainchild was thus regarded as purely hypothetical. Their answers to the problem and their conception of its significance reflect their general epistemology and tested how the sensations of two different senses, sight and touch, related to each other. If the sensations are different (which few doubted), would they be necessarily linked to each other, or only arbitrarily? What is the role of experience in connecting the sensations? And is the visual concept of a sphere different from the tactile concept of it? If yes, connecting the two is the result of reasoning or experience?

Locke’s reply, for example, ruled out the possibility of the sort of innately guided, unconscious reasoning that was postulated by Descartes. The empiricist Locke argued that distance perception is an acquired ability and answered Molyneux’s question in the negative. This reply seems to contradict his own doctrine that states: ideas of primary qualities like shape or motion resemble their causes. As the visual and tactile ideas in question have common causes, this implies that they resemble each other. On this count it is less trivial to give a straightforward answer to the question.

Leibniz, on the other hand, held that the connection between the visual and the tactile sensations is not innate, but if a blind person knows the crucial piece of information that in front of him lie a cube and a globe, he can conjecture and find out which is which based on the geometric features that the visual and the tactile ideas share.

In Texts 2-5 are some possible answers to the question from Molyneux, Locke, Berkeley, Reid, and Synge. Text 2, apart from citing the crucial paragraph 8, contains some of the most important parts describing Locke’s basic ideas about perception, and its role in acquiring ideas. Text 3 from (Berkeley 1709) shows that apart from the moon illusion the Molyneux-problem had an important role in the development of Berkeley’s theory of perception. The short abstract should be read with Text 6 of Chapter IX in mind. Thomas Reid in Text 4 drew a distinction between natural and original perception and perception acquired through experience. Text 5 is the synopsis of the argument of Edward Synge, stressing that the images of the object should be different, hence distinguishing them must be possible.

• Problem 2 What are the reasons for answering Molyneux’s question
in the positive or in the negative? What are perceptible properties in these accounts? Find similarities and differences (also with earlier accounts of perception).

4 The (Ir)relevance of Empirical Data

A common reason for the weakening of the faculty of vision and even blindness is the clouding of the eye’s lenses, the result of cataract. The earliest operations were probably carried out three thousand years ago in India. The method of depressing the cataract, also referred to as reclination is mentioned in Celsus’s De re medicina, and is still used in some parts of the world. The surgeon, by pushing the lens out of the line of sight with a needle, can usually greatly improve the sight of the patient. Although the lens is of a higher optical density than the aqueous humor in front of the lens and the vitreous humor behind it, the eye’s refractive capacity is only partially lost as a result of the operation (the convex outer layer of the eye, the cornea contributes significantly more to this capacity than the lens). With the removal of the lens the eye loses its ability to accommodate, to focus on objects that have different distance, and the resulting farsighted (hyperopic) condition has to be corrected by lenses\(^1\).

4.1 The First Publication: Cheselden

A well-known surgeon William Cheselden published an article in the Philosophical Transactions of the Royal Society in 1728 on a successful cataract operation (Cheselden 1728). Although ignorant about the debate Molyneux’s problem ignited (and thus not asking the patient to tell a cube and a sphere apart without touching them), Cheselden played an important role in bringing the discussion to a new level. Below in Text 6 is the original report by Cheselden.

- **Problem 3** Collect aspects that can be used in favour of either negative or positive answers to the question. Are any of these decisive for the debate?

---

\(^1\)The eyes often got infected after the operation, or the opaque lens slipped back to its original position. To escape this latter problem another technique had been developed to extract the cataract. By cutting the capsule of the lens open, the opaque interior could be removed through the opening in the cornea. While already used in the eleventh century, the method was rediscovered and gained popularity in the eighteenth century. Today artificial lenses can be implanted, greatly reducing the negative effects of earlier methods.
4.2 Testing Object Recognition: Home

Unlike Cheselden, later surgeons were generally very well aware of the problem posed by Molyneux. They devised different methods to test patients, but none of the tests were unanimously accepted. One of the most ingenious one was Everard Home (1756-1832), who described two patients in Text 7 (Home 1807).

4.3 A Sphere and a Cube: Franz

In his book The Eye August Franz argued that Molyneux’s question cannot be strictly speaking answered: the patient will see the difference of the objects, but will call one a square and the other a disk (Franz 1839), p. 32-33. His careful experiments confirmed his hypothesis (Franz 1841). In Text 8 the experiments and some of the details of the recovery process can be read.

- Problem 4 Compare the three reports. Among others discuss the following aspects: how is the test situation prepared, what measures are taken to reach conclusive results, how are the questions formulated and described. Note the contradictory features.

5 Texts

5.1 Text 1: William Molyneux

- Molyneux’s letter to Locke from Dublin, July 7th 1688 (Degenaar 1996).

A Problem Proposed to the Author of the Essai Philosophique concernant L’Entendement

A Man: being born blind, and having a Globe and a Cube, nigh of the same bignes, Committed into his Hands, and being taught or Told, which is Called the Globe, and which the Cube, so as easily to distinguish them by his Touch or Feeling; Then both being taken from Him, and Laid on a Table, Let us suppose his Sight Restored to Him; Whether he Could, by his Sight, and before he touch them, know which is the Globe and which the Cube? Or Whether he Could know by his Sight, before he stretchd out his Hand, whether he Could not Reach them, tho they were Removed 20 or 1000 feet from him?

If the Learned and Ingenious Author of the Forementiond Treatise think this problem Worth his Consideration and Answer, He may at any time Direct it to One That Much Esteems him, and is

His Humble Servant
1. Perception the first simple idea of reflection. Perception, as it is the first faculty of the mind exercised about our ideas; so it is the first and simplest idea we have from reflection, and is by some called thinking in general. Though thinking, in the propriety of the English tongue, signifies that sort of operation in the mind about its ideas, wherein the mind is active; where it, with some degree of voluntary attention, considers anything. For in bare naked perception, the mind is, for the most part, only passive; and what it perceives, it cannot avoid perceiving.

2. Reflection alone can give us the idea of what perception is. What perception is, every one will know better by reflecting on what he does himself, when he sees, hears, feels, etc., or thinks, than by any discourse of mine. Whoever reflects on what passes in his own mind cannot miss it. And if he does not reflect, all the words in the world cannot make him have any notion of it.

3. Arises in sensation only when the mind notices the organic impression. This is certain, that whatever alterations are made in the body, if they reach not the mind; whatever impressions are made on the outward parts, if they are not taken notice of within, there is no perception. Fire may burn our bodies with no other effect than it does a billet, unless the motion be continued to the brain, and there the sense of heat, or idea of pain, be produced in the mind; wherein consists actual perception.

6. The effects of sensation in the womb. But though it be reasonable to imagine that children receive some ideas before they come into the world, yet these simple ideas are far from those innate principles which some contend for, and we, above, have rejected. These here mentioned, being the effects of sensation, are only from some affections of the body, which happen to them there, and so depend on something exterior to the mind; no otherwise differing in their manner of production from other ideas derived from sense, but only in the precedency of time. Whereas those innate principles are supposed to be quite of another nature; not coming into the mind by any accidental alterations in, or operations on the body; but, as it were, original characters impressed upon it, in the very first moment of its being and constitution.

7. Which ideas appear first, is not evident, nor important. As there are some ideas which we may reasonably suppose may be introduced into the minds of
children in the womb, subservient to the necessities of their life and being there: so, after they are born, those ideas are the earliest imprinted which happen to be the sensible qualities which first occur to them; amongst which light is not the least considerable, nor of the weakest efficacy. And how covetous the mind is to be furnished with all such ideas as have no pain accompanying them, may be a little guessed by what is observable in children new-born; who always turn their eyes to that part from whence the light comes, lay them how you please. But the ideas that are most familiar at first, being various according to the divers circumstances of children’s first entertainment in the world, the order wherein the several ideas come at first into the mind is very various, and uncertain also; neither is it much material to know it.

8. Sensations often changed by the judgment. We are further to consider concerning perception, that the ideas we receive by sensation are often, in grown people, altered by the judgment, without our taking notice of it. When we set before our eyes a round globe of any uniform colour, e.g. gold, alabaster, or jet, it is certain that the idea thereby imprinted on our mind is of a flat circle, variously shadowed, with several degrees of light and brightness coming to our eyes. But we having, by use, been accustomed to perceive what kind of appearance convex bodies are wont to make in us; what alterations are made in the reflections of light by the difference of the sensible figures of bodies;—the judgment presently, by an habitual custom, alters the appearances into their causes. So that from that which is truly variety of shadow or colour, collecting the figure, it makes it pass for a mark of figure, and frames to itself the perception of a convex figure and an uniform colour; when the idea we receive from thence is only a plane variously coloured, as is evident in painting. To which purpose I shall here insert a problem of that very ingenious and studious promoter of real knowledge, the learned and worthy Mr. Molyneux, which he was pleased to send me in a letter some months since; and it is this:—“Suppose a man born blind, and now adult, and taught by his touch to distinguish between a cube and a sphere of the same metal, and nighly of the same bigness, so as to tell, when he felt one and the other, which is the cube, which the sphere. Suppose then the cube and sphere placed on a table, and the blind man be made to see: quaere, whether by his sight, before he touched them, he could now distinguish and tell which is the globe, which the cube?” To which the acute and judicious proposer answers, “Not. For, though he has obtained the experience of how a globe, how a cube affects his touch, yet he has not yet obtained the experience, that what affects his touch so or so, must affect his sight so or so; or that a protuberant angle in the cube, that pressed his hand unequally, shall appear to his eye as it does in the cube.”—I agree with this thinking gentleman, whom I am proud to call my friend, in his answer to this problem; and am of opinion that the blind man, at first sight, would not be able with certainty to say which was the globe, which the cube, whilst he only
saw them; though he could unerringly name them by his touch, and certainly
distinguish them by the difference of their figures felt. This I have set down,
and leave with my reader, as an occasion for him to consider how much he
may be beholden to experience, improvement, and acquired notions, where he
thinks he had not the least use of, or help from them. And the rather, because
this observing gentleman further adds, that “having, upon the occasion of my
book, proposed this to divers very ingenious men, he hardly ever met with one
that at first gave the answer to it which he thinks true, till by hearing his reasons
they were convinced.”

9. This judgment apt to be mistaken for direct perception. But this is not, I
think, usual in any of our ideas, but those received by sight. Because sight, the
most comprehensive of all our senses, conveying to our minds the ideas of light
and colours, which are peculiar only to that sense; and also the far different
ideas of space, figure, and motion, the several varieties whereof change the
appearances of its proper object, viz. light and colours; we bring ourselves by
use to judge of the one by the other. This, in many cases by a settled habit,—in
things whereof we have frequent experience, is performed so constantly and
so quick, that we take that for the perception of our sensation which is an idea
formed by our judgment; so that one, viz. that of sensation, serves only to excite
the other, and is scarce taken notice of itself;—as a man who reads or hears with
attention and understanding, takes little notice of the characters or sounds, but
of the ideas that are excited in him by them.

...  

15. Perception the inlet of all materials of knowledge. Perception then
being the first step and degree towards knowledge, and the inlet of all the
materials of it; the fewer senses any man, as well as any other creature, hath; and
the fewer and duller the impressions are that are made by them, and the duller
the faculties are that are employed about them,—the more remote are they from
that knowledge which is to be found in some men. But this being in great variety
of degrees (as may be perceived amongst men) cannot certainly be discovered
in the several species of animals, much less in their particular individuals. It
suffices me only to have remarked here,—that perception is the first operation
of all our intellectual faculties, and the inlet of all knowledge in our minds. And
I am apt too to imagine, that it is perception, in the lowest degree of it, which
puts the boundaries between animals and the inferior ranks of creatures. But
this I mention only as my conjecture by the by; it being indifferent to the matter
in hand which way the learned shall determine of it.

- Hint to Problem 1 Molyneux in the 1688 letter asked two ques-
tions: one about distinguishing the objects by sight, the other about
guessing their correct distance. In the version quoted by Locke the questions are about distinguishing the objects and correctly naming them.

In the first case all combinations of answers are possible (as perception of distance and visible forms can be independent of each other). For the reformulation if the answer to the second question is positive, the answer to the first one must necessarily be also positive. A negative answer to the first question, however, implies a similarly negative answer to the second one.

Omitting the question about the distance of the objects might be the result of Molyneux’s conviction that distance cannot be directly perceived, and its estimation is not inborn. As he writes in 1792: “the Estimate we make of the Distance of Objects...is rather an act of Judgement than of Sense; and acquired by Exercise...rather than Natural” (Molyneux 1692, p. 113). Other differences include mentioning the material of the objects in the second version, as well as the fact that the blind should be an adult—both to further specify and disambiguate the situation.

5.3 **Text 3: George Berkeley**

- Berkeley’s *An essay towards a new theory of vision* first appeared in 1709 (Berkeley 1709). The quotation below is from the fourth edition.

133. Now, if a square surface perceived by touch be of the same sort with a square surface perceived by sight, it is certain the blind man here mentioned might know a square surface as soon as he saw it: it is no more but introducing into his mind by a new inlet an idea he has been already well acquainted with. Since, therefore, he is supposed to have known by his touch that a cube is a body terminated by square surfaces, and that a sphere is not terminated by square surfaces: upon the supposition that a visible and tangible square differ only in numero it follows that he might know, by the unerring mark of the square surfaces, which was the cube, and which not, while he only saw them. We must therefore allow either that visible extension and figures are specifically distinct from tangible extension and figures, or else that the solution of this problem given by those two thoughtful and ingenious men is wrong.
5.4 Text 4: Thomas Reid

- The text is from Thomas Reid’s *An Inquiry into the Human Mind* printed in 1765. Parts of Chapter 6 are also reprinted in (Degenaar 1996).

3. To a man newly made to see, the visible appearance of objects would be the same as to us; but he would see nothing at all of their real dimensions, as we do. He could form no conjecture, by means of his sight only, how many inches or feet they were in length, breadth or thickness. He could perceive little or nothing of their real figure; nor could he discern that this was a cone and that a cylinder. His eye could not inform him, that this object was near, and that more remote...

11. . . . if Dr. Saunderson² had been made to see, and attentively had viewed the figures of the first book of Euclid, he might, by thought and consideration, without touching them, have found out that they were the very figures he was before so well acquainted with by touch.

20. By this sense [sight] we perceive originally the visible figure and colour of bodies only, and their visible place; but we learn to perceive by the eye, almost everything which we can perceive by touch. The original perceptions of this sense, serve only as signs to introduce the acquired.

24. The original perceptions which nature gave [to children] are few, and insufficient for the purpose of life; and therefore she made them capable of acquiring many more perceptions by habit.

5.5 Text 5: Edward Synge

- Edward Synge (1659-1741), later archbishop of Tuam and famous writer of the *Gentleman’s Religion* (Synge 1693) wrote a letter to a friend on 6 September 1695, who sent it to Molyneux, who passed on a copy to Locke. The gist of his argument is summarized below (Locke 1976, vol. V. p. 496).

²Nicolas Saunderson (1682-1739) was Lucasian chair of mathematics from 1711-1739 at the University of Cambridge (the same chair as the one Newton held after Isaac Barrow). As a result of smallpox he became blind at the age of one, and had been educated with help from his family and friends. He was mostly lecturing on algebra (but also on optical problems!) and invented a calculating device using pins placed at the corners and midpoints of each side of a square, and one more point at the center. He used a larger pin to represent one. To represent numbers, he would move the pin to new positions around the perimeter, using several of these squares for large numbers and calculations (Tattersall 1992).
The Image which upon the first View such a man will frame of a Cube, must needs be this, that it is a body which is not alike in all parts of its Superficies which consequently must be agreeable to the idea which before he had of it and different from that idea which he had of a globe

- **Hint to Problem 2** Molyneux’s argument is not strong. It rests on the questionable statement that the visual impressions the healed blind have are insufficient to tell the difference between the two objects.

Neither Molyneux nor Locke argue for the validity of the above claim. To explain Locke’s position one possibility is to accept that he subscribes to an atomistic theory of sensation. In this view the sensory “atoms”, the *simple ideas* have no cross-modal relations. Thus, even though primary qualities like space, figure, and motion are conveyed through both visual and tactile impulses, these have no relations to each other other than co-occurrence. Another view is to argue that unlike visual ideas of colour and light, those of figures are not immediately given to sight. This latter view has Medieval and Renaissance origins, as has been claimed by (Bolton 1994). But we have already seen in the first part of the book that Aristotle’s or Ptolemy’s attitude was much the same. This account was still widely accepted in the seventeenth century. From both of these views follows that the novice will not be able to distinguish the sphere from the cube.

Berkeley sets up a dichotomy: either visible and tangible ideas are unrelated or Molyneux and Locke gave the wrong answer. For him vision directly perceives only light and colour. In §129 he writes “There is no immediate object of sight besides light and colour, and the visual perception of line and surface is incommensurable with the tangible perception of these properties—hence any relationship between the two must be purely contingent and rely on experience”.

Reid presupposes that a ‘perfect’ two-dimensional projection is produced, but the observer is unable to find the distance cues, delineate

---

3 Compare Chapter V, Text 2 (ch. 6. and 7.) and Chapter VII, Text 3 (§5): colours are primarily visible, and also Text 2, (§9): the proper object of sight are light and colours. Berkeley in his account states: “…in a strict sense I see nothing but light and colours” (§130).
the objects: his faculty of sight would provide him with the relevant data, but these signs would be without any significance to him. Interestingly for him figure is also directly perceived (§20.), not only colour. Hence the faculty of mathematical reasoning—which Sanderson obviously possessed—, could bridge the gap between sensory modes: given the time to study the objects he could distinguish and correctly name them. Of course two-dimensional forms are more easily perceived.

Synge claims that as one object has unlike parts and the other has not, this difference is sufficient for the man to decide which object is which. This account also implies that some primitive form of figure- or shape-perception takes place.

5.6 TEXT 6: WILLIAM CHESELDEN

- William Cheselden’s article from the *Philosophical Transactions*, Volume 35 (1727–1728, pp. 447-450) was one of the most important empirical contributions to the debate, even though Molyneux’s question is not discussed and no experiments are carried out.
VII. An Account of some Observations made by a young Gentleman, who was born blind, or lost his Sight so early, that he had no Remembrance of ever having seen, and was couched between 13 and 14 Years of Age. By Mr. Will. Chesfelden, F. R. S. Surgeon to Her Majesty, and to St. Thomas’s Hospital.

THO’ we say of the Gentleman that he was blind, as we do of all People who have Ripe Cataracta, yet they are never so blind from that Cause, but that they can discern Day from Night; and for the most Part in a strong Light, distinguish Black, White, and Scarlet; but they cannot perceive the Shape of any thing; for the Light by which these Perceptions are made, being let in obliquely thro’ the aqueous Humour, or the anterior Surface of the Chry stal l ine (by which the Rays cannot be brought into a Focus upon the Retina) they can discern in no other Manner, than a found Eye can thro’ a Glass of broken Jelly, where a great Variety of Surfaces so differently refract the Light, that the several distinct Pencils of Rays cannot be collected by the Eye into their proper Foci; wherefore the Shape of an Object in such a Case, cannot be at all discern’d, tho’ the Colour may: And thus it was with this young Gentleman, who though he knew these Colours asunder in a good Light; yet when he saw them 

after
after he was couch'd, the faint Ideas he had of them before, were not sufficient for him to know them by afterwards; and therefore he did not think them the same, which he had before known by those Names. Now Scarlet he thought the most beautiful of all Colours, and of others the most gay were the most pleasing, whereas the first Time he saw Black, it gave him great Uneasiness, yet after a little Time he was reconcil'd to it; but some Months after, seeing by Accident a Negro Woman, he was struck with great Horror at the Sight.

When he first saw, he was so far from making any Judgment about Distances, that he thought all Objects whatever touch'd his Eyes, (as he express'd it) as what he felt, did his Skin; and thought no Objects so agreeable as those which were smooth and regular, tho' he could form no Judgment of their Shape, or guess what it was in any Object that was pleasing to him: He knew not the Shape of any Thing, nor any one Thing from another, however different in Shape, or Magnitude; but upon being told what Things were, whose Form he before knew from feeling, he would carefully observe, that he might know them again; but having too many Objects to learn at once, he forgot many of them; and (as he said) at first he learn'd to know, and again forgot a thousand Things in a Day. One Particular only (tho' it may appear trifling) I will relate; Having often forgot which was the Cat, and which the Dog, he was ashamed to ask; but catching the Cat (which he knew by feeling) he was observ'd to look at her steadfastly, and then setting her down, said, So Puifs! I shall know you another Time. He was very much surpriz'd, that those Things which he had lik'd best,
best, did not appear most agreeable to his Eyes, ex-
pecting those Persons would appear most beautiful that
he lov'd most, and such Things to be most agreeable to his
Sight that were so to his Taste. We thought he soon
knew what Pictures represented, which were shew'd to
him, but we found afterwards we were mistaken; for
about two Months after he was couch'd, he discovered
at once, they represented solid Bodies; when to that
Time he consider'd them only as Party-colour'd Planes,
or Surfaces diversified with Variety of Paint; but e-
ven then he was no less surpriz'd, expecting the Pictures
would feel like the Things they represented, and was
amaz'd when he found those Parts, which by their
Light and Shadow appear'd now round and uneven, felt
only flat like the rest; and ask'd which was the lying
Sence, Feeling, or Seeing?

Being shewn his Father's Picture in a Locket at his
Mother's Watch, and told what it was, he acknowledg-
ed a Likeness, but was vastly surpriz'd; asking, how
it could be, that a large Face could be express'd in so
little Room, saying, It should have seem'd as impossible
to him, as to put a Bushel of any thing into a Pint.

At first, he could bear but very little Sight, and the
Things he saw, he thought extremly large; but upon
seeing Things larger, those first seen he conceiv'd less, ne-
ever being able to imagine any Lines beyond the Bounds he
saw; the Room he was in he said, he knew to be but Part
of the House, yet he could not conceive that the whole
House could look bigger. Before he was couch'd, he
expected little Advantage from Seeing, worth under-
going an Operation for, except reading and writing;
for he said, He thought he could have no more Plea-

O 0 0 2  
ure
Hint to Problem 3 The report stresses that even patients with a ripe cataract are not completely blind, they see light, and colours, but have no shape perception. Thus a positive answer to Molyneux's question cannot be decisively drawn on the basis of the report. The patient seems to have impaired distance perception: an argument against natural geometry. His conception of space is limited.

The report is unclear on many points: the boy could see objects (they touched his eyes) but saw no shapes. But what was the condition of the boy's eyes? Without a lens and without accommodation what did
he see? Learning the visual recognition of objects was difficult: was he unintelligent and so unfit to test Molyneux's problem, or is such difficulty inevitable?

The impaired distance-perception described in the account paralleled Berkeley’s (and Molyneux’s) insight that distance is not perceived directly. But this, and the weak perception of shape could be the result of the inability to accommodate. Thus, supporters of Berkeley’s theory of vision applauded the observation. In his book published in 1733 Berkeley saw his theory vindicated (Berkeley 1733). Smith in his popular book on optics (Smith 1738) believed that Locke’s and Molyneux’s answers are correct. The report also supports Reid’s view that a two-dimensional representation of the visual space is insufficient to see three-dimensional objects. Voltaire, who summarized Cheselden’s report in French (Albert 1997) and ‘sGravesande also held similar opinions. These writers usually did not discuss the methods used.

Several writers, however, criticised the methodology of the psychological experiments. Cheselden wrote the report after he had operated the second eye—that is more than a year after the first operation. He did not have extensive notes, and leading questions might have influenced the responses. LaMettrie (1709-1751) drew attention to the condition and the state of the eyes after surgery. Condillac (1714-1780) believed that once the patient started reflecting on what he saw he could connect his ideas of sight with the ideas he already had through touch\(^4\). This view has much in common with some of the most recent theories in cognitive science, the role of language is stressed in intermodal representation and cognition.

5.7 Text 7: Everard Home
- The following excerpts are from Everard Home’s article entitled: “An Account of Two Children Born with Cataracts in Their Eyes, to Shew That Their Sight Was Obscured in Very Different Degrees; With Experiments to Determine the Proportional Knowledge of Objects

\(^4\)While in his early works Condillac held that from birth on we can perceive size, distance, and shape, in his later works he became an empiricist and follower of Berkeley, opposing Locke by denying the existence of unconscious judgement.

strong in the former case. As the eye was not irritable, and was likely to be but little disturbed by this operation, every thing was previously got ready for ascertaining his knowledge of objects, as soon as the operation was over, should the circumstances prove favourable. The operation was attended with success, and gave very little pain. The eye was allowed ten minutes to recover itself: a round piece of card of a yellow colour, one inch in diameter, was then placed about six inches from it. He said immediately that it was yellow, and on being asked its shape, said, “Let me touch it, and I will tell you.” Being told that he must not touch it, after looking for some time, he said it was round. A square blue card, nearly the same size, being put before him, he said it was blue and round. A triangular piece he also called round. The different colours of the objects placed before him he instantly decided on with great correctness, but had no idea of their form. He moved his eye to different distances, and seemed to see best at 6 or 7 inches. His focal distance has been since ascertained to be 7 inches. He was asked whether the object seemed to touch his eye, he said “No;” but when desired to say at what distance it was, he could not tell. These experiments were made in the theatre of the hospital, in which the operation was performed, before the surgeons and all the students. He was highly delighted with the pleasure of seeing, and said it was “so pretty,” even when no object was before him, only the light upon his eye. The eye was covered, and he was put to bed, and told to keep himself quiet, but upon the house-surgeon going to him half an hour afterwards, his eye was found uncovered, and he was looking at his bed curtains, which were close drawn.
The bandage was replaced, but so delighted was the boy with seeing, that he again immediately removed it. This circumstance distressed the house-surgeon, who had been directed to prevent him from looking at any thing till the next day, when the experiment was to be repeated. Finding that he could not enforce his instructions, he thought it most adviseable to repeat the experiment about two hours after the operation. At first the boy called the different cards round; but upon being shewn a square, and asked if he could find any corners to it, he was very desirous of touching it. This being refused, he examined it for some time, and said at last that he had found a corner, and then readily counted the four corners of the square; and afterwards when a triangle was shewn him, he counted the corners in the same way; but in doing so his eye went along the edge from corner to corner, naming them as he went along.

Next day, when I saw him, he told me he had seen “the soldiers with their fifes and pretty things.” The guards in the morning had marched past the hospital with their band; on hearing the music he had got out of bed, and gone to the window to look at them. Seeing the bright barrels of the musquets, he must in his mind have connected them with the sounds which he heard, and mistaken them for musical instruments. On examining the eye 24 hours after the operation, the pupil was found to be clear. A pair of scissors was shewn him, and he said it was a knife. On being told he was wrong, he could not make them out; but the moment he touched them he said they were scissors, and seemed delighted with the discovery. On being shewn a guinea at the distance of 15 inches from his eye, he said it was a seven
shilling piece, but placing it about 5 inches from his eye, he knew it to be a guinea; and made the same mistake, as often as the experiment was repeated.

From this time he was constantly improving himself by looking at, and examining with his hands, every thing within his reach, but he frequently forgot what he had learnt. On the 10th I saw him again, and I told him his eye was so well that he might go about as he pleased without leaving the room. He immediately went to the window, and called out, "What is that moving?" I asked him what he thought it was? He said, "A dog drawing a wheelbarrow. There is one, two, three dogs drawing another. How very pretty!" These proved to be carts and horses on the road, which he saw from a two pair of stairs window.

On the 19th, the different coloured pieces of card were separately placed before his eye, and so little had he gained in thirteen days, that he could not without counting their corners one by one tell their shape. This he did with great facility, running his eye quickly along the outline, so that it was evident he was still learning, just as a child learns to read. He had got so far as to know the angles, when they were placed before him, and to count the number belonging to any one object.

The reason of his making so slow a progress was, that these figures had never been subjected to examination by touch, and were unlike any thing he was accustomed to see.

He had got so much the habit of assisting his eyes with his hands, that nothing but holding them could keep them from the object.

On the 26th the experiments were again repeated on the
couched eye, to ascertain the degree of improvement which had been made. It was now found that the boy, on looking at any one of the cards in a good light, could tell the form nearly as readily as the colour,

From these two cases the following conclusions may be drawn:

That, where the eye before the cataract is removed, has only been capable of discerning light, without being able to distinguish colours, objects after its removal will seem to touch the eye, and there will be no knowledge of their outline; which confirms the observations made by Mr. Cheselden:

That where the eye has previously distinguished colours, there must also be an imperfect knowledge of distances, but not of outline, which however will afterwards be very soon acquired, as happened in Mr. Ware's cases. This is proved by the history of the first boy in the present Paper, who before the operation had no knowledge of colours or distances, but after it, when his eye had only arrived at the same state, that the second boy's was in before the operation, he had learnt that the objects were at a distance, and of different colours: that when a child has acquired a new sense, nothing but great pain or absolute coercion, will prevent him from making use of it.
5.8 Text 8: John Charles August Franz

- Again, only excerpts are reprinted from J. C. August Franz’s article from the *Philosophical Transactions of the Royal Society of London*, Volume 131 (1841, pp. 59-68). The text contain a number of technical terms that are not essential to understanding the experiments.

VI. *Memoir of the Case of a Gentleman born blind, and successfully operated upon in the 18th year of his age, with Physiological Observations and Experiments.* By J. C. August Franz, of Leipsic, M.D., M.R.C.S., &c. Communicated by Sir Benjamin C. Brodie, Bart., F.R.S., &c.

Received April 21.—Read May 6, 1841.

Mr. F. J., the subject of the present memoir, is the son of a physician; of scrofulous diathesis, but otherwise of robust constitution; of irritable temperament, but of contented and happy disposition; and endowed with an excellent understanding, quick power of conception, and retentive memory. In both the eyes of his father, cataract (with the addition, I suspect, of glaucoma) has manifested itself within the last four years, after a severe attack of influenza. The relatives on the paternal side are predisposed to diseases of the eye, but in the mother, and in the relatives on her side, no such predisposition can be traced. With regard to the cause of the ophthalmic affections which form the subject of this paper, the mother seemed to lay much stress on the following circumstance, which, although it may possibly have had some share in the cause of one of them, can have had no influence, in my opinion, in producing the other. She stated to me that in the eighth month of her pregnancy, which up to this period had proceeded favourably, she received from her youngest child, which she was carrying in her arms, a severe blow on the eye. This accident caused inflammation of the eye, accompanied with a curious visual illusion, viz. that all objects which she saw, but especially those situated on the ground, appeared of a deep concave form; an illusion which lasted for several months. The fright experienced from the accident also brought on convulsions, which, recurring several times, extended even to the fetus. The recurrence of these convulsions produced in the mind of the mother a continual anxiety and fear for the health of the child, while the pain arising from the ophthalmia, together with the visual illusion just mentioned, gave her fears a direction more especially towards its eyes. Delivery took place at the proper period, when the eyes of the infant, which was otherwise healthy and well-formed, were found to present a twofold defect of organization. The father, to whose statement, on account of his professional knowledge, more weight is to be attached, informed me that both eyes were turned inwards to such an extent that a portion of the cornea was hidden by the inner canthus, and that in both pupils a yellowish-white discoloration was to be observed, which, being situated behind the iris, could not be the pupillary membrane. That the strabismus and cataract of both eyes in this case were congenital, is evident from the
testimony both of the parents and of the nurse, whom I have closely questioned on
this subject. The latter, who can distinctly remember all the circumstances of the
case, told me that when the child was a few months old, she held a light before its
eyes, of which it took no notice. I ascertained also from her that the eye-balls had
not that restless motion which is generally observed in those who are born blind, but
that both eyes were always turned inwards, and that but rarely either the one or the
other was moved from the internal cauthus.

It was also stated to me, that towards the end of the second year the operation of
keratonyxis was performed on the right eye, upon which a severe iritis ensued, termi-
nating in atrophy of the eye-ball. Within the next four years two similar operations
were performed on the left eye, which did not indeed destroy the organ, but at the
same time did not remove the opacity in the pupil. The colour of the opacity be-
came in time, however, of a clearer white; and the patient acquired a certain sensa-
tion of light, which he did not seem to have had before the operation. Both eyes for
a long time retained a disposition to inflammation, and suffered repeatedly from con-
junctivitis, whence the vessels of the conjunctiva were increased in number and size
to such an extent, that it was necessary they should be several times excised.

from page 63

On opening the eye for the first time on the third day after the operation, I asked
the patient what he could see; he answered that he saw an extensive field of light,
in which everything appeared dull, confused, and in motion. He could not distin-
guish objects. The pain produced by the light forced him to close the eye immedi-
ately. Two days afterwards, the eye, which had been kept closed by means of court-
plaster, was again opened. He now described what he saw as a number of opake
watery spheres, which moved with the movements of the eye, but, when the eye was
at rest, remained stationary, and then partially covered each other. Two days
after this the eye was again opened; the same phenomena were again observed, but
the spheres were less opake and somewhat transparent; their movements more steady;
they appeared to cover each other more than before. He was now for the first time capable, as he said, to look through the spheres, and to perceive a difference, but merely a difference, in the surrounding objects. When he directed his eye steadily towards an object, the visual impression produced by the object was painful and very imperfect, and no clear visual perception of it took place, because the eye, on account of the intolerance of light, could not be kept open long enough for the formation of the idea as derived from visual sensation. The appearance of spheres diminished daily; they became smaller, clearer, and more pellucid, allowed objects to be seen more distinctly, and disappeared entirely after two weeks. The musea volitantes, which had the form of black, immovable, and horizontal stripes, appeared, every time the eye was opened, in a direction upwards and inwards. When the eye was closed, he observed, especially in the evening, in an outward and upward direction, an appearance of dark blue, violet, and red colours; these colours became gradually less intense, were shaded into bright orange, yellow, and green, which latter colours alone eventually remained, and in the course of five weeks disappeared entirely.

As soon as the intolerance of light had so far abated that the patient could regard an object without pain and for a sufficient time to gain an idea of it, the following experiments were made in the presence of Dr. Swain. The first experiments were of that class in which the idea of a visible object is derived merely from pure visual sensation; the succeeding, of that kind in which the idea, in ordinary cases, depends upon the sense of sight combined with the sense of touch, and is gained by reflecting on the impressions made on the organs of both senses. It was necessary to perform these experiments on different days, as otherwise they would have distressed the eye too much.

1st Experiment. Silk ribands of different colours, fastened on a black ground, were employed to show, first the primitive, and then the complementary colours. The patient recognized the different colours, with the exception of yellow and green, which he frequently confounded, but could distinguish when both were exhibited at the same time. He could point out each colour correctly when a variety was shown him at the same time. Grey pleased him best, because this colour he said produced an agreeable and grateful sensation; the effect of red, orange, and yellow was painful, but not disagreeable; that of violet and brown not painful, but very disagreeable; the latter he called ugly. Black produced subjective colours, and white occasioned the recurrence of musea volitantes in a most vehement degree.

2nd Experiment. The patient sat with his back to the light, and kept his eye closed. A sheet of paper, on which two strong black lines had been drawn, the one horizontal, the other vertical, was placed before him, at the distance of about three feet. He was now allowed to open the eye, and, after attentive examination, he called the lines by their right denominations. When I asked him to point out with his finger the horizontal line, he moved his hand slowly, as if feeling, and pointed to the vertical, but after a short time, observing his error, he corrected himself. The outline in black of a square, six inches in diameter, within which a circle had been drawn, and within the latter a triangle, was, after careful examination, recognized and
correctly described by him. When he was asked to point out either of the figures, he never moved his hand directly and decidedly; but always as if feeling, and with the greatest caution; he pointed them out, however, correctly. A line consisting of angles, or in other words, a zigzag, and a spiral line, both drawn on a sheet of paper, he observed to be different, but could not describe them otherwise than by imitating their forms with his finger in the air. He said he had no idea of these figures.

3rd Experiment. The windows of the room were darkened, with the exception of one, towards which the patient, closing his eye, turned his back. At the distance of three feet and on a level with the eye, a solid cube and a sphere, each of four inches diameter, were placed before him. Allowing him to move the head in a lateral direction no further than was necessary to compensate the point of view of the right amaurotic eye, I now let him open his eye, and requested him to state decidedly what he observed. After attentively examining these bodies, he said he saw a quadrangular and a circular figure, and after some consideration he pronounced the one a square and the other a disc. His eye being then closed, the cube was taken away, and a disc of equal size substituted and placed next to the sphere. On again opening his eye, he observed no difference in these objects, but regarded them both as discs. The solid cube was now placed in a somewhat oblique position before the eye, and close beside it a figure cut out of pasteboard, representing a plane outline prospect of the cube when in this position. Both objects he took to be something like flat quadrates. A pyramid, placed before him with one of its sides towards his eye, he saw as a plain triangle. This object was now turned a little, so as to present two of its sides to view, but rather more of one side than of the other; after considering and examining it for a long time, he said that this was a very extraordinary figure; it was neither a triangle, nor a quadrangle, nor a circle; he had no idea of it, and could not describe it; “in fact,” said he, “I must give it up.” On the conclusion of these experiments, I asked him to describe the sensations the objects had produced, whereupon he said that immediately on opening his eye, he had discovered a difference in the two objects, the cube and the sphere, placed before him, and perceived that they were not drawings; but that he had not been able to form from them the idea of a square and a disc, until he perceived a sensation of what he saw in the points of his fingers, as if he really touched the objects. When I gave the three bodies (the sphere, cube, and pyramid) into his hand, he was much surprised that he had not recognized them as such by sight, as he was well acquainted with these solid mathematical figures by his touch. These experiments prove the correctness of the hypothesis I have advanced elsewhere on the well-known question put by Mr. Molyneux to Locke, which was answered by both these gentlemen in the negative, and has been much discussed since their time.

4th Experiment. In a vessel, containing water to about the depth of one foot, was placed a musket-ball, and on the surface of the water a piece of pasteboard, of the same form, size, and colour as the ball. The patient could perceive no difference in the position of these bodies; he believed both to be upon the surface of the water.
Pointing to the ball, I desired him to take up this object; he made an attempt to take it from the plane of the water, but when he found he could not grasp it there, he said he had deceived himself, the objects were lying in the water; upon which I informed him of their real position. I now desired him to touch the ball, which lay in the water, with a small rod; he attempted this several times, but always missed his aim; he could never touch the object at the first movement of his hand towards it, but only by feeling about with the rod. On being questioned with respect to reflected light, he said that he was always obliged to bear in mind, that the looking-glass was fastened to the wall, in order to correct his idea of the apparent situation of objects behind the glass.

When the patient first acquired the faculty of sight, all objects appeared to him so near that he was sometimes afraid of coming in contact with them, though they were in reality at a great distance from him. He saw everything much larger than he had supposed from the idea obtained by his sense of touch. Moving, and especially living objects, such as men, horses, &c., appeared to him very large. If he wished to form an estimate of the distance of objects from his own person, or of two objects from each other, without moving from his place, he examined the objects from different points of view by turning his head to the right and to the left. Of perspective in pictures he had of course no idea; he could distinguish the individual objects in a painting, but could not understand the meaning of the whole picture; it appeared to him unnatural, for instance, that the figure of a man represented in the front of the picture should be larger than a house or a mountain in the background. All objects appeared to him perfectly flat; thus, although he very well knew by his touch that the nose was prominent, and the eyes sunk deeper in the head, he saw the human face only as a plane. Though he possessed an excellent memory, this faculty was at first quite deficient as regarded visible objects; he was not able, for example, to recognize visitors, unless he heard them speak, till he had seen them very frequently. Even when he had seen an object repeatedly, he could form no idea of its visible qualities in his imagination, without having the real object before him. Therefore, when he dreamed of any persons, of his parents, for instance, he felt them and heard their voices, but never saw them; but now, after having seen them frequently, he saw them also in his dreams. The human face pleased him more than any other object presented to his view; the eyes he thought most beautiful, especially when in motion; the nose disagreeable, on account of its form and great prominence; the movement of the lower jaw in eating he considered very ugly. Although the newly-acquired sense afforded him many pleasures, the great number of strange and extraordinary sights was often disagreeable and wearisome to him; he said that he saw too much novelty which he could not comprehend. And even though he could see both near and remote objects very well, he would nevertheless continually have recourse to the use of the sense of touch.

On the 21st of September I operated, in the presence of several medical gentlemen, in one sitting, on both eyes for the congenital strabismus. The lids were fixed by the
fingers of an assistant, the ball of the eye by a pair of forceps, and the tendon of the muscle divided by a pair of curved scissors. The rectus internus of the right eye was, like the organ itself, atrophied. The conjunctiva of the left eye was thickened at the inner angle; the muscle was uncommonly broad and thick; its tendon had a very broad attachment to the ball, and behind it was a separate bundle of muscular fibres attached to the sclerota. The pupils of both eyes assumed immediately after the operation their proper position in the orbits. No inflammation ensued; not even in the left eye, which, from the prior operation, was still rather sensitive. The musceus volitantes became less irksome, and the violent spasms which previously had affected not only the eyelids, but also the whole left side of the face, disappeared entirely. The right eye, which had been amaurotic, gained by this operation the power of perceiving light, so that when the left eye is closed, the patient can now distinguish light and shade, on the hand being moved before this eye. The sight of the left eye likewise was considerably improved in acuteness and clearness, both as regards near and distant objects, but especially the latter. Objects now, however, appeared in a different situation to that which they really held; when, for instance, he directed his eye to an object situated immediately before him, he saw it more to his right, and, if he attempted to grasp it, he moved his hand in this wrong direction*. For this reason in walking across a room he always took a direction to the right, and consequently often came unawares in contact with articles of furniture, &c. This appearance of objects in false positions lasted for two months, after which time he was also capable of walking forwards in a straight direction. The right atrophied eye, which before the operation was deeply sunk in the orbit, is now more prominent, and appears therefore fuller and larger, so that the difference of the two eyes is less perceptible; he has consequently gained considerably in personal appearance. On one occasion when I was honoured with a visit from Mr. Lawrence, Dr. Watson, Dr. Kerrison, and several other medical gentlemen at my residence, I introduced him to them for examination.

In the middle of October I let him try several pair of spectacles at Mr. A. Ross's, in Regent Street. With a double convex lens of 5 1/2 inch focus, he saw both near and remote objects of large size most clearly and distinctly, but for small objects he could find no glasses that improved his sight. He recognized the capitals of a large print with his naked eye, and on looking through a pin-hole made in a card held close before the eye, he could distinguish even the small letters of a very minute print. He had not yet learned to read. The reason for the condition of his sight with respect to small objects, and that his vision is better on cloudy days, is no doubt to be sought in the enlarged pupil and the immobility of the iris.

In the middle of November he was able without spectacles to read the names over the windows of the shops in the streets, and to tell the time to the minute by St.

* This phenomenon I have observed in all eyes operated upon for strabismus of a great degree and long standing, when the other eye was closed. I have mentioned it in the Medical Gazette for June 1840, vol. xxvi. p. 540, where I have also given an explanation of the physiological cause.
Paul's clock. Walking alone in the crowded streets, especially in the City, he found very tedious. He said, seeing so many different things, and the quick movements of the multitude of people, carriages, &c., confused his sight to such a degree, that at last he could see nothing; that the sensation produced by the object last seen had not yet disappeared from the retina, when the next object made its impression thereon, by which means confusion of ideas, great anxiety, and even vertigo were occasioned, from which he could only free himself by closing his eyes for a few moments.

In the middle of December an experiment was again made with spectacles. A lens of seven inches focus was now of the same service as one of 5½ inches had been two months before. After the operation for the strabismus he was accustomed, in speaking with any person, to turn his eye away from the face, as otherwise he said he felt disturbed by the looks of the person; he had now at length learned to look at the eyes of those with whom he conversed. The old habit of using the sense of touch to examine objects he had not yet entirely lost.

In the middle of February 1841, a third experiment was made with spectacles. A lens of ten inches focus was of the same service as one of seven inches had been on the last occasion, and one of 5½ inches four months ago. This proves a slow, but positive amelioration of sight, and permits us to expect a still greater improvement, the more so as the patient has not passed the period of puberty. If the employment of spectacles were begun at the present period, although it is now more than seven months since the operation was performed, there would be no further amelioration of sight; the development of the visual apparatus would be arrested. I am therefore of opinion that the use of spectacles is not to be permitted, until it is, as it were, mathematically demonstrated by similar experiments with lenses, that the sight is no longer improved; by which means the faculty may in time, perhaps, reach such a degree of perfection as not to require any lens at all for remote objects.

This is the only case on record within my knowledge wherein, with a person born blind and afterwards successfully operated upon at a period of life as far advanced as in this instance, such experiments have ever been made. In the well-known case of Cheselden, published in the Philosophical Transactions for the year 1728 (page 447), the patient was only in the fourteenth year of his age, and although the case contains many highly interesting physiological observations, no series of systematic experiments was instituted. Beer has also made some interesting observations, which, however, like those made in rather a superficial manner by Janin and Daviel, tend principally to describe the impressions which the newly-acquired sense had made on the mind of the person operated upon. In Ware's case the patient was not born blind, but had become so at an early period of life. In the present paper I have merely given the simple history of the case, without making any remark on several points interesting to the pathologist and physiologist, to which I shall advert on a future occasion; the explanation and philosophy of the foregoing experiments as to the sense of sight I shall attempt in another paper, which I purpose to lay before this Society.

• Hint to Problem 4 It is probably a good idea to collect the differences in a large table. This makes comparison easier and the dif-
ferences more visible. Headings could include the vision-impairment before the operation, the type of operation (for Cheselden’s patient (C) and Home’s (H) couching), the types of experiments and the way they were recorded, distance perception—though distance perception was impaired in all cases, the measure was different (objects “touching” the eyes (C) or not (H))—, surface perception, shape perception, ability to learn (problematic (C), based on cue-identification (H), at first bad cross-modal coordination in Franz’s patient (F)), 2-D vision (no problem for C, F; but H has to be given cues), size perception (for C and F objects appear large, for H objects from the window appear very small: size constancy breaks down), space-perception, colour-experiences, focus, attitude to the newly obtained ability to see, and the doctor’s theory to explain the empirical data. The details can be worked out using headings not listed in this short list.

6  Epilogue

The investigations became more and more detailed, the technique of couching developed, and cataracts could be removed at an earlier age—thus the number of adults, on whom the original question could be tested, decreased. Although the question was addressed by leading scientists, like William James, Adam Smith, Johannes Müller, John Stuart Mill, and others, its relevance gradually decreased. More and more studies on animals and newborn babies have shown the complexity of the development of the visual system. In nineteenth century psychological discussions (like the nativism-empiricism debate) the problem became an interesting side issue. Berkeley’s theory received a devastating blow from Wheatstone’s 1838 discovery of stereoscopic space-perception. From our present point of view the original question lost most of its significance: deprivation experiments show that depending on the extent of blindness the development of the visual system is more or less retarded. Once a so-called critical period has passed congenitally blind individuals cannot be made to see. Also, intermodal matching of tactile and visual stimuli have been recorded in neonates suggesting much closer connection between the sense-modalities than expected by the British Empiricists (Meltzoff & Borton 1979, pp. 403-404). Although the 300-year-old conundrum remains unanswered, in the recent years an ever increasing number of studies take up the problem. Even though in blind

5These findings were debated (Maurer, Stager & Mondloch 1999) but some recent studies support the original results (Streri & Gentaz 2003).
patients some of the areas formerly used for processing visual information in the brain might be involved e.g. in processing tactile information, this transfer between modalities remains problematic. As a recent study suggests (Hamilton & Pascual-Lene 1998):

Herein may be an answer to Molyneux’s famous question. It seems that without visual input, the occipital cortex can be recruited to serve a role in tactile spatial processing, but this recruitment does not result in intrinsic cross-modal transfer of tactile information to potential visual sensations. On the contrary, visual-spatial association appears to be generated and maintained only by the simultaneous experience of both modalities, so that lifelong cataracts or even temporary loss of sight can impede an individual’s ability to apply spatial information across sensory modalities. Thus, a blind individual suddenly gifted with sight might have abnormalities in visual perception, because the cortical space that would otherwise have been devoted to vision has been subsumed by the modality of touch. Such a person might lack the functional apparatus required for visual analysis and recognition of visual shapes, perhaps even to the point of being unable to distinguish correctly shapes like a sphere and a cube.

Studies on tactile vision substitution systems (Bach-y Rita 1972) show that with use even tactile stimuli can be perceived as non-proximal. In recent years more and more works rephrase Molyneux’s original question (Jac-omuzzi, Kobau & Bruno 2003), but the detailed study of these developments is outside the scope of this chapter.

7 Further Problems

- **Problem 5** Although only touched upon in this chapter, the response of British Empiricist philosophers to Molyneux’s question is a highly intriguing and debated topic (see Further Readings). Discuss in more detail the views of Berkeley, Locke, and Reid. Collect characteristic features of the Lockean model of perception. Compare his discussion of the “perceptible” with that of Aristotle. What is the relation of Berkeley’s discussion to either account (use paragraphs 41, 43, 128-133 from his *New Theory of Vision*).

- **Problem 6** Discuss in more detail the development of psychological experimenting using the three articles cited here, together with
(Abbott 1988 (1864)) and modern reports, like (Gregory & Wallace 1963) and (Gregory 1997).

- **Problem 7** How would earlier writers answer the Molyneux question? Use the hint to Problem 2 of this chapter, and the writings of Aristotle, Ptolemy, and Lucretius in the first part of the book.

8 **Suggested Readings**


A short introduction to Cheselden’s life is given by (Cope 1953). For a general history of cataract surgery, esp. of the method of extraction see (Kwitko & Kelman 1998), or, for a contemporary account, (Neill 1848). For an earlier version of this chapter see (Zemplén 2003).

On the connection of enlightenment philosophy and vision (Levin 1999) is a useful introduction. Berkeley’s *New Theory* appeared in several editions. From these the second (Berkeley 1709) contained an appendix with special attention to Descartes’s theory of seeing, and his employment of a ‘natural geometry’. He later reflected on the empirical data from Cheselden and others, and believed that these vindicate his theory (Berkeley 1733). In the nineteenth century the Berkeleyan theory of vision was attacked from several sides – see e.g. (Bailey 1842, Abbott 1988 (1864)). For an insight into the debate with reprints of works from the 1840s, see (Pitcher 1988). For evaluations of Berkeley’s theory see (Armstrong 1960) or (Atherton 1990). His views on perception are analysed in (Stack 1970), part five in (Sosa 1987) is also devoted to Berkeley’s views on perception.

On the general philosophy of Reid, see (Dalgarno & Matthews 1989), on his ideas on perception see Lehrer’s and Smith’s article in (Capp & MacIntosh 1985). On Locke’s answer, see (Rogers 1994).

In this chapter the main focus was on the answers by British Empiricists as well as on the changes in the techniques of finding conclusive empirical data to Molyneux’s question. This meant that the significant French reactions were mostly overlooked. For an introduction to French attitudes on the soul in the seventeenth-eighteenth centuries see (Rosenfield 1968). On Condillac see (Knight 1968), with bibliographical references. The first volume of (Robinson 1977) includes the *Logie* of Condillac. For an English translation of LaMettrie’s *The natural history of the soul*, see the anthology by
Spicker 1970). Diderot’s connections to the sciences are investigated in (Pucci 1986).

Also, more contemporary approaches and classical experiments like (Hubel & Wiesel 1962) have not been included. An excellent starting point for follow-up reading is (Wade 1998). On the nineteenth century developments, see (Wade 2000). For a more contemporary argument see (Gallagher Forthcoming), esp. Neurons and Neonates: Reflections on the Molyneux Problem.
Bibliography


Cheselden, W.: 1728, An Account of some Observations Made by a Young Gentleman, who was Born Blind, or Lost his Sight so Early, that he had No Remembrance of Ever Having Seen, and was Couch’d between 13 and 14 Years of Age, *Philosophical Transactions of the Royal Society of London* **25**, 447–450.


Dunn, S.: 1762, An attempt to assign the Cause, why the Sun and Moon appear to the naked Eye larger when they are near the Horizon. With an Account of several natural Phenomena, relative to this Subject, *Philosophical Transactions of the Royal Society of London* 25, 447–450.


Hecht, E.: 2001, Optik, Oldenbourg Verlag, München.

Helmholtz, H. v.: 1867, Handbuch der physiologischen Optik I-III.


Home, E.: 1807, An Account of two Children born with Cataracts in their Eyes, to Shew that their Sight was Obscured in Very Different Degrees: ith Experiments to Determine the Proportional Knowledge of Objects Acquired by them Immediately after the Cataracts were Removed, Philosophical Transactions of the Royal Society of London 97, 83–92.


Molyneux, W.: 1687, Concerning the Apparent Magnitude of the Sun and Moon, or the Apparent Distance of Two Stars, when Nigh the Horizon, and When Higher Elevated, Philosophical Transactions (1683-1775) 16, 314–323.


Newton, I.: 1671-72, New Theory about Light and Colors, Philosophical Transactions (80), 3075–3087.


Smith, R.: 1738, *A compleat system of opticks in four books, viz. a popular, a mathematical, a mechanical, and a philosophical treatise. To which are added remarks upon the whole*, Printed for the author, Cambridge.


Boethius: Texte und Abhandlungen zur Geschichte der Mathematik
und der Naturwissenschaften, Franz Steiner Verlag, Stuttgart, pp. 481–
502.