香港中文大學
The Chinese University of Hong Kong

與自然對話

In Dialogue with Nature

通識教育基礎課程讀本

Textbook for
General Education
Foundation Programme

Second edition, 2012

香港中文大學　大學通識教育部
Office of University General Education
The Chinese University of Hong Kong

In Dialogue with Nature
與自然對話

In Dialogue with Nature (與自然對話) is the textbook of "UGFN 1000 In Dialogue with Nature", one of the two courses of the General Education Foundation Programme of The Chinese University of Hong Kong.

Second edition 2012

Managing editor / Ng Hiu Chun Emily
Editorial assistants / Lam Yee Ki Dora, Ng Tsz Ying Tracy
Cover and layout design / Ng Hiu Chun Emily
ISBN: 978-988-16660-1-7

Published by Office of University General Education
The Chinese University of Hong Kong
Shatin, Hong Kong
http://www.cuhk.edu.hk/oge/gef　　gefp@cuhk.edu.hk
Printed in Hong Kong by Wellfit Production & Printing Company (唯美印刷製作公司)

本書以不經漂白和增白的原漿紙印刷
This book is printed on unbleached, with no brightening agents, original pulp paper.

Acknowledgements

pp. 5–10: From Plato, *Republic*, translated by C.D.C. Reeve. Copyright © 2004 by Hackett Publishing Company, Inc. Reprinted by permission of Hackett Publishing Company, Inc. All rights reserved.

pp. 11–48: From David C. Lindberg, *The Beginnings of Western Science: The European Scientific Tradition in Philosophical, Religious, and Institutional Context, Prehistory to A.D. 1450* (Second Edition). Published by The University of Chicago Press. Copyright ©1992, 2007 by The University of Chicago. All rights reserved.

pp. 49–62: From I. Bernard Cohen, *The Birth of a New Physics: Revised and Updated*. Copyright ©1985 by I. Bernard Cohen. Copyright © 1960 by Educational Services Incorporated. Used by permission of W. W. Norton & Company, Inc.

pp. 63–70: From Isaac Newton, *The Principia: Mathematical Principles of Natural Philosophy*. Copyright © 2012 by University of California Press. Reproduced with permission of University of California Press via Copyright Clearance Center.

pp. 73–96: From Charles Darwin, *On the Origin of Species by Means of Natural Selection* (First Edition) (1859). Reproduced from the public domain.

pp. 97–142: From James D. Watson and with Andrew Berry, *DNA: The Secret of Life.* Copyright © 2003 by DNA Show LLC. Used by permission of Alfred A. Knopf, a division of Random House, Inc.

pp. 143–158: From Rachel Carson, *Silent Spring: 40th Anniversary Edition*

Copyright © 1962 by Rachel L. Carson.

Copyright © renewed 1990 by Roger Christie.

This selection originally appeared in SILENT SPRING: 40th Anniversary Edition by Rachel Carson.

This edition of “pages 65 to pages 83 from SILENT SPRING: 40th Anniversary Edition by Rachel Carson” is published by Office of University General Education, The Chinese University of Hong Kong.

Table of Contents

Letter to Students

Dear Students,

Welcome to the community of learners that make up the General Education Foundation (GEF) Programme. We are about to embark on a challenging journey—through the running waters of different cultures and disciplines, and out to the vast open seas of knowledge. The two courses offered in GEF, namely, *In Dialogue with Humanity* and *In Dialogue with Nature*, are designed to widen perspectives and lay open contrasting worlds. We will be reflecting upon issues and concerns of crucial importance, and as we do so, seeking to build a fuller understanding of ourselves and the universe in which we live.

Most people attend university to acquire specific knowledge or a set of skills in a major discipline as a way toward a professional or academic career. In the current social environment, however, with its rapidly changing technology, communications integration, and volatile economics, which we call the knowledge economy, traditional employment qualifications are no longer enough. A wide range of perspectives are needed to understand the complexity of the globalized world. Multiple fields of knowledge often have to be called into play simultaneously and in coordination to solve increasingly complex problems.

At the root of what influences our lives in the world is not only what knowledge or skills we have, but also what kind of people we are. That is why the ultimate goal of general education is to nurture the qualities of responsible intellectuals and global citizens. We must be aware of different ways of knowing, responsive to the common concerns of human existence, sensitive to cultural diversity, and most importantly, we must develop the intellectual ability and disposition to deal with and make informed judgements about the unfamiliar.

In Dialogue with Humanity and *In Dialogue with Nature* constitute the core component of the General Education Programme at CUHK. These are theme-based seminar courses guided by the reading of classic texts, and can be seen as "core" for a few reasons.

In terms of content, the GEF courses cover two key—or *core*—domains of human knowledge and endeavor: the search for the meaning and value of being human, and the achievements and limitations of the human understanding of nature. The ideas contained in the classics we will be reading address critical problems and issues in human existence and knowledge, and define long-lasting beliefs and values which are relevant to contemporary existence.

In addition, both courses facilitate the development of important skills such as active engagement in reading, discussion, reflection, and writing, the indispensable attributes of a lifelong learner, and the intellectual toolkit required for a fulfilling experience at university and beyond.

Above all, the GEF programme is a space in which *core* attitudes can begin to develop. First and foremost is a sense of fearlessness before intellectual challenges. We will engage in direct dialogues with and make sense of original texts from very different traditions, cultures and disciplines.

Equally important quality is a spirit of open-mindedness. As we engage in dialogue with the texts and with one another, we will find our views are often contrastive. Our beliefs and values may be challenged. We may come to see that, for many important questions, there is no definite or final answer; that human knowledge and endeavor is a never-ending process of search and research. And in this process we will lay the foundations for responsible and proactive participation in the quests for knowledge and well being.

GEF participants are *active* learners. This process should not and will not, be summed up by a letter grade. GEF learning outcomes are a sharpened awareness of cultural differences and a sensitivity to the diversity of values; enhanced awareness of universal concerns and sensitivity to the scope and limitations of human knowledge; confidence in confronting difficult texts and complex ideas; capability in expressing informed arguments orally or in writing. To make this intellectual journey a successful one, your active participation is the key.

Be ready!

Leung Mei Yee
Director
General Education Foundation Programme

Notes on editing

1. This anthology is published as a textbook for the course “In Dialogue with Nature”, part of a core-text programme that is theme-based in orientation. Selections and omissions of passages from the source texts are therefore made with the goal of fostering a clear focus on the course themes.
2. Reprinted texts in this anthology are reproduced without change except for correction of typographical errors, alteration of typing styles for editorial consistency, and omission or alteration of references to page numbers or appendices of the source texts.
3. All paragraph numbers on the margins are added by this publication for ease of reference during class discussion. They follow the sequence in which the paragraphs appear in the source texts. Thus, where paragraphs are omitted, the numbers are skipped.
4. All footnotes and texts in square brackets are from the source texts, except those marked as “*Ed.*”.
5. Omissions from the source texts are indicated by bracketed ellipsis [. . .]. All omissions so indicated are this publication’s.
6. All headings in Text 10b “Brush Talks from Dream Brook 夢溪筆談” are added for this publication and not present in the source texts. In the Chinese text, entry numbers, which indicate the order of appearance of the entries in the source text, are also given together with the headings.
7. Special attention has been given to Text 11b “Elements”:
   — This anthology reproduces only Heath’s translation of the *Elements* but omitted his “scholia”, commentary on the Greek text and notes on his translation.
   — In the original publication, Heath uses a system of different types of brackets to indicate texts added by him for clarity of translation but not literally present in the Greek text. For ease of reading for the general reader, the brackets are removed from this book while the texts are kept. The only exceptions are the square brackets added by Heath to refer the reader to the relevant parts of the *Elements*, e.g., “[Def. 10]” refers to Definition 1, “[Post.1]” refers to Postulate 1, “[C.N. 3]” refers to Common Notion 3, “[I.4]” refers to Proposition 4 of Book I. These are all retained.

In Dialogue with Nature

Introduction

Humans have long been curious about nature. Since before the dawn of civilization, we have struggled to come to terms with the vast array of phenomena that the world presents. Some of the behavior of the cosmos seems to be orderly; sunrise and sunset, the seasons, the rotations of the heavens, for example. Some seems chaotic and random; for example, storms and volcanic eruptions. Perhaps, behind these patterns, or orders, are some rules of nature. In attempting to make sense of these observations, mediated by particular circumstance and culture, in various parts of the world, vastly different approaches to the understanding of nature were developed and established as the foundations of the ancient cultures.

The endeavour by the early natural philosophers and their descendants, today's modern scientists, to discover rules of nature has been a dialogue between nature and us. This dialogue is conducted through experience and reasoning. It began in ancient Greece, beginning about 500 BCE, where philosophers began applying the intellectual tools of logic and reason to their observations of natural phenomena in their attempts to explain existence. Their attitude towards nature and their methods of investigation laid down the foundation of our modern, Western, empirical sciences and the technologies they support. Later, more numerous and more systematic observations refined our understanding of nature. Modern scientists generalize natural laws from these observations.

But in the course of this process, our understanding of nature underwent numerous modifications and even some profoundly radical changes. Some of these new modes of thinking have even occasionally challenged the very concept of understanding, forcing humankind to reflect on the meaning of knowledge.

Meanwhile, at roughly the same time in China, the quest to understand nature met with very different solutions. By developing concepts such as *yin* and *yang* and the five elements, Chinese scholars created a set of understandings very different from those of their Greek counterparts. It wasn't until the early modern historical period that concepts of western science began to infiltrate the Chinese ethos, and yet China, without the supposed benefits of science, was for many centuries the acknowledged super-power of the ancient world. Clearly the ongoing encounter between the cultures of east and west still has much to offer to our understanding of nature.

This course invites students to retrace the train of thought of our predecessors in this quest for knowledge, and with whose writings students will engage in dialogues. By following the footsteps of great minds, students shall develop informed views about nature and human interactions with it.

Course Structure

This course brings students on an intellectual journey. We first reach out to the external world. Our objects of investigation are the physical cosmos and the world of life. These investigations constitute the first two parts of the course: I. Human Exploration of the Physical Universe and II. Human Exploration of the World of Life. We shall learn about some major discoveries of western science: The universe is governed by physical laws. Species are undergoing natural selection and DNA is the code of life.

Then we shall return to examine our own minds and understanding. This is the third part of the course: III. Our Understanding of Human Understanding, a journey of reflection.

First, we shall reflect on our understanding of nature, and on the validity and limitation of scientific knowledge. We shall try to understand the nature of the human mind, which has the faculty of rational thinking. Second, through the comparison between Western science and Chinese philosophy, we shall try to understand why modern science did not emerge in China; and in Chinese philosophy, we shall find perspectives that may help us to reflect on modern science's approach to the relation between humans and nature.

Wong Wing Hung
Deputy Director
General Education Foundation Programme

Part I
Human Exploration of the Physical Universe

This is the first part of the journey. We shall start with Plato's *Republic*. In the excerpt (Text 1a), which is usually referred to as the Allegory of the Cave, through the dialogue between Socrates and Glaucon, Plato makes a distinction between perception and reality. This distinction sets the basic tone of Western science.

According to Plato, what we see in the "visible realm" are shadows. Only in the "intelligible realm" can one find the Form of the Good, which is truly real and is "the cause of all that is correct and beautiful in anything". It is the ultimate concern of those who seek true understanding. For modern scientists, there is also an "intelligible realm" of natural laws. The ultimate concern of scientists is not a comprehensive description of natural phenomena but understanding of the underlying laws. David Lindberg is a leading historian of medieval and early modern science. In the excerpt (Text 1b) from his popular textbook, *The Beginnings of Western Science*, he explicates Plato's notion of the two realms and its relevance to science.

The second text (Text 2) is also an excerpt from *The Beginnings of Western Science*. It introduces Aristotelian physics and gives a concise historical account of the subsequent developments of the concept of motion by medieval scientists, who laid the foundation for the emergence of Newtonian physics.

Text 3a is an excerpt from *The Birth of a New Physics*, which was written for common readers by I. Bernard Cohen, an internationally renowned Newton scholar. In this excerpt, Cohen expounds the background of Newtonian physics and the main ideas behind the great work, *The Principia: Mathematical Principles of Natural Philosophy*, which has long been regarded as one of the most important books in human history. Text 3b is the first few pages of this great work, in which Newton constructed a mathematical system of the world. The reader will appreciate Newton's effort in giving precise definitions to abstract concepts like mass, force and motion, and also his attempt to draw a line between mathematics and philosophy.

Text 1a

from

Republic

by Plato

BOOK 7

SOCRATES' NARRATION CONTINUES:

SOCRATES: Next, then, compare the effect of education and that of the lack of 1
it on our nature to an experience like this. Imagine human beings living in an underground, cavelike dwelling, with an entrance a long way up that is open to the light and as wide as the cave itself. They have been there since childhood, with their necks and legs fettered, so that they are fixed in the same place, able to see only in front of them, because their fetter prevents them from turning their heads around. Light is provided by a fire burning far above and behind them. Between the prisoners and the fire, there is an elevated road stretching. Imagine that along this road a low wall has been built—like the screen in front of people that is provided by puppeteers, and above which they show their puppets.

GLAUCON: I am imagining it.

SOCRATES: Also imagine, then, that there are people alongside the wall carrying 2
multifarious artifacts that project above it—statues of people and other animals,

made of stone, wood, and every material. And as you would expect, some of the carriers are talking and some are silent.
GLAUCON: It is a strange image you are describing, and strange prisoners.

3 SOCRATES: They are like us. I mean, in the first place, do you think these prisoners have ever seen anything of themselves and one another besides the shadows that the fire casts on the wall of the cave in front of them?
GLAUCON: How could they, if they have to keep their heads motionless throughout life?

4 SOCRATES: What about the things carried along the wall? Isn't the same true where they are concerned?
GLAUCON: Of course.

5 SOCRATES: And if they could engage in discussion with one another, don't you think they would assume that the words they used applied to the things they see passing in front of them?
GLAUCON: They would have to.

6 SOCRATES: What if their prison also had an echo from the wall facing them? When one of the carriers passing along the wall spoke, do you think they would believe that anything other than the shadow passing in front of them was speaking?
GLAUCON: I do not, by Zeus.

7 SOCRATES: All in all, then, what the prisoners would take for true reality is nothing other than the shadows of those artifacts.
GLAUCON: That's entirely inevitable.

8 SOCRATES: Consider, then, what being released from their bonds and cured of their foolishness would naturally be like, if something like this should happen to them. When one was freed and suddenly compelled to stand up, turn his neck around, walk, and look up toward the light, he would be pained by doing all these things and be unable to see the things whose shadows he had seen before, because of the flashing lights. What do you think he would say if we told him that what

he had seen before was silly nonsense, but that now—because he is a bit closer to what is, and is turned toward things that *are* more—he sees more correctly? And in particular, if we pointed to each of the things passing by and compelled him to answer what each of them is, don't you think he would be puzzled and believe that the things he saw earlier were more truly real than the ones he was being shown?
GLAUCON: Much more so.

SOCRATES: And if he were compelled to look at the light itself, wouldn't his eyes 9
be pained and wouldn't he turn around and flee toward the things he is able to see, and believe that they are really clearer than the ones he is being shown?
GLAUCON: He would.

SOCRATES: And if someone dragged him by force away from there, along the 10
rough, steep, upward path, and did not let him go until he had dragged him into the light of the sun, wouldn't he be pained and angry at being treated that way? And when he came into the light, wouldn't he have his eyes filled with sunlight and be unable to see a single one of the things now said to be truly real?
GLAUCON: No, he would not be able to—at least not right away.

SOCRATES: He would need time to get adjusted, I suppose, if he is going to see 11
the things in the world above. At first, he would see shadows most easily, then images of men and other things in water, then the things themselves. From these, it would be easier for him to go on to look at the things in the sky and the sky itself at night, gazing at the light of the stars and the moon, than during the day, gazing at the sun and the light of the sun.
GLAUCON: Of course.

SOCRATES: Finally, I suppose, he would be able to see the sun—not reflections of 12
it in water or some alien place, but the sun just by itself in its own place—and be able to look at it and see what it is like.
GLAUCON: Necessarily.

SOCRATES: After that, he would already be able to conclude about it that it 13
provides the seasons and the years, governs everything in the visible world, and is

in some way the cause of all the things that he and his fellows used to see.
GLAUCON: That would clearly be his next step.

14 SOCRATES: What about when he reminds himself of his first dwelling place, what passed for wisdom there, and his fellow prisoners? Don't you think he would count himself happy for the change and pity the others?
GLAUCON: Certainly.

15 SOCRATES: And if there had been honors, praises, or prizes among them for the one who was sharpest at identifying the shadows as they passed by; and was best able to remember which usually came earlier, which later, and which simultaneously; and who was thus best able to prophesize the future, do you think that our man would desire these rewards or envy those among the prisoners who were honored and held power? Or do you think he would feel with Homer that he would much prefer to "work the earth as a serf for another man, a man without possessions of his own," and go through any sufferings, rather than share their beliefs and live as they do?
GLAUCON: Yes, I think he would rather suffer anything than live like that.

16 SOCRATES: Consider this too, then. If this man went back down into the cave and sat down in his same seat, wouldn't his eyes be filled with darkness, coming suddenly out of the sun like that?
GLAUCON: Certainly.

17 SOCRATES: Now, if he had to compete once again with the perpetual prisoners in recognizing the shadows, while his sight was still dim and before his eyes had recovered, and if the time required for readjustment was not short, wouldn't he provoke ridicule? Wouldn't it be said of him that he had returned from his upward journey with his eyes ruined, and that it is not worthwhile even to try to travel upward? And as for anyone who tried to free the prisoners and lead them upward, if they could somehow get their hands on him, wouldn't they kill him?
GLAUCON: They certainly would.

18 SOCRATES: This image, my dear Glaucon, must be fitted together as a whole with what we said before. The realm revealed through sight should be likened to

the prison dwelling, and the light of the fire inside it to the sun's power. And if you think of the upward journey and the seeing of things above as the upward journey of the soul to the intelligible realm, you won't mistake my intention—since it is what you wanted to hear about. Only the god knows whether it is true. But this is how these phenomena seem to me: in the knowable realm, the last thing to be seen is the form of the good, and it is seen only with toil and trouble. Once one has seen it, however, one must infer that it is the cause of all that is correct and beautiful in anything, that in the visible realm it produces both light and its source, and that in the intelligible realm it controls and provides truth and understanding; and that anyone who is to act sensibly in private or public must see it.
GLAUCON: I agree, so far as I am able.

SOCRATES: Come on, then, and join me in this further thought: you should not be 19
surprised that the ones who get to this point are not willing to occupy themselves with human affairs, but that, on the contrary, their souls are always eager to spend their time above. I mean, that is surely what we would expect, if indeed the image I described before is also accurate here.
GLAUCON: It is what we would expect.

SOCRATES: What about when someone, coming from looking at divine things, 20
looks to the evils of human life? Do you think it is surprising that he behaves awkwardly and appears completely ridiculous, if—while his sight is still dim and he has not yet become accustomed to the darkness around him—he is compelled, either in the courts or elsewhere, to compete about the shadows of justice, or about the statues of which they are the shadows; and to dispute the way these things are understood by people who have never seen justice itself?
GLAUCON: It is not surprising at all.

[. . .]

from

The Beginnings of Western Science

by David C. Lindberg

CHAPTER TWO
THE GREEKS AND THE COSMOS

[. . .]

Plato's World of Forms

The death of Socrates in 399 B.C., coming as it did around the turn of the century 28
(not on their calendar, of course, but on ours), has made it a convenient point of demarcation in the history of Greek philosophy. Thus Socrates' predecessors of the sixth and fifth centuries (the philosophers who have occupied us until now in this chapter) are commonly called the "pre-Socratic philosophers." But Socrates' prominence is more than an accident of the calendar, for Socrates represents a shift in emphasis within Greek philosophy, away from the cosmological concerns of the sixth and fifth centuries toward political and ethical matters. Nonetheless, the shift was not so dramatic as to preclude continuing attention to the major problems of

pre-Socratic philosophy. We find both the new and the old in the work of Socrates' younger friend and disciple, Plato (fig. 2.4).

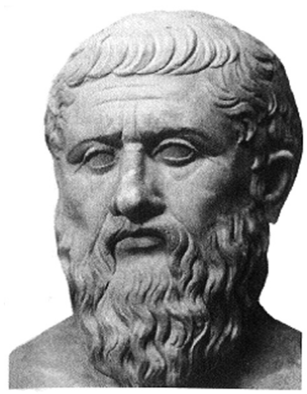


Fig. 2.4. Plato (1st c. A.D. copy). Museo Vaticano, Vatican City. Alinari/Art Resource N.Y.

29 Plato (427–348/47) was born into a distinguished Athenian family, active in affairs of state; he was undoubtedly a close observer of the political events that led up to Socrates' execution. After Socrates' death, Plato left Athens and visited Italy and Sicily, where he seems to have come into contact with Pythagorean philosophers. In 388 Plato returned to Athens and founded a school of his own, the Academy, where young men could pursue advanced studies (see fig. 4.1). Plato's literary output appears to have consisted almost entirely of dialogues, the majority of which have survived. We will find it necessary to be highly selective in our examination of Plato's philosophy; let us begin with his quest for the underlying reality.

30 In a passage in one of his dialogues, the *Republic*, Plato reflected on the relationship between the actual tables constructed by a carpenter and the idea or definition of a table in the carpenter's mind. The carpenter replicates the mental idea as closely as possible in each table he makes, but always imperfectly. No two manufactured tables are alike down to the smallest detail, and limitations in the material (a knot here, a warped board there) ensure that none will fully measure up to the ideal.

31 Now, Plato argued, there is a divine craftsman who bears the same relationship to the cosmos as the carpenter bears to his tables. The divine craftsman (the Demiurge) constructed the cosmos according to an idea or plan, so that the cosmos

and everything in it are replicas of eternal ideas or forms—but always imperfect replicas because of limitations inherent in the materials available to the Demiurge. In short, there are two realms: a realm of forms or ideas, containing the perfect form of everything; and the material realm in which these forms or ideas are imperfectly replicated.

Plato's notion of two distinct realms will seem strange to many people, and 32
we must therefore stress several points of importance. The forms are incorporeal, intangible, and insensible; they have always existed, sharing the property of eternality with the Demiurge; and they are absolutely changeless. They include the form, the perfect idea, of everything in the material world. One does not speak of their location, since they are incorporeal and therefore not spatial. Although incorporeal and imperceptible by the senses, they objectively exist; indeed, true reality (reality in its fullness) is located only in the world of forms. The sensible, corporeal world, by contrast, is imperfect and transitory. It is less real in the sense that the corporeal object is a replica of, and therefore dependent for its existence upon, the form. The form has primary existence, its corporeal replica secondary existence.

Plato illustrated this conception of reality in his famous "allegory of the cave," 33
found in book VII of the *Republic*. Men are imprisoned within a deep cave, chained so as to be incapable of moving their heads. Behind them is a wall, and beyond that a fire. People walk back and forth behind the wall, holding above it various objects, including statues of humans and animals; the objects cast shadows on the wall that is visible to the prisoners. The prisoners see only the shadows cast by these objects; and, having lived in the cave from childhood, they no longer recall any other reality. They do not suspect that these shadows are but imperfect images of objects that they cannot see; and consequently they mistake the shadows for the real.

So it is with all of us, says Plato. We are souls imprisoned in bodies. The 34
shadows of the allegory represent the world of sense experience. The soul, peering out from its prison, is able to perceive only these flickering shadows, and the ignorant claim that this is all there is to reality. However, there do exist the statues and other objects of which the shadows are feeble representations and also the humans and animals of which the statues are imperfect replicas. To gain access to these higher realities, we must escape the bondage of sense experience and climb out of the cave, until we find ourselves able, finally, to gaze on the eternal realities, thereby entering the realm of true knowledge.

35 What are the implications of these views for the concerns of the pre-Socratic philosophers? First, Plato equated his forms with the underlying reality, while assigning derivative or secondary existence to the corporeal world of sensible things. Second, Plato has made room for both change and stability by assigning each to a different level of reality: the corporeal realm is the scene of imperfection and change, while the realm of forms is characterized by eternal, changeless perfection. Both change and stability are therefore genuine; each characterizes something; but changelessness belongs to the forms and thus shares their fuller reality.

36 Third, as we have seen, Plato addressed epistemological questions, placing observation and true knowledge (or understanding) in opposition. Far from leading upward to knowledge or understanding, the senses are chains that tie us down; the route to knowledge is through philosophical reflection. This is explicit in the *Phaedo*, where Plato maintains the uselessness of the senses for the acquisition of truth and points out that when the soul attempts to employ them it is inevitably deceived.

37 Now the short account of Plato's epistemology frequently ends here; but there are important qualifications that it would be a serious mistake to omit. Plato did not, in fact, dismiss the senses altogether, as Parmenides had done and as the passage from the *Phaedo* might suggest Plato did. Sense experience, in Plato's view, served various useful functions. First, sense experience may provide wholesome recreation. Second, observation of certain sensible objects (especially those with geometrical properties) may serve to direct the soul toward nobler objects in the realm of forms. Plato used this argument as justification for the pursuit of astronomy. Third, Plato argued (in his theory of reminiscence) that sense experience may actually stir the memory and remind the soul of forms that it knew in a prior existence, thus stimulating a process of recollection that will lead to actual knowledge of the forms.

38 Finally, although Plato firmly believed that knowledge of the eternal forms (the highest, and perhaps the only true, form of knowledge) is obtainable only through the exercise of reason, the changeable realm of matter is also an acceptable object of study. Such studies serve the purpose of supplying examples of the operation of reason in the cosmos. If this is what interests us (as it sometimes did Plato), the best method of exploring it is surely to observe it. The legitimacy and utility of sense experience are clearly implied in the *Republic*, where Plato acknowledged that a prisoner emerging from the cave first employs his sense of sight to apprehend living creatures, the stars, and finally the most noble of visible (material) things, the

sun. But if he aspires to apprehend "the essential reality," he must proceed "through the discourse of reason unaided by any of the senses." Both reason and sense are thus instruments worth having; which one we employ on a particular occasion will depend on the object of study.

There is another way of expressing all of this, which may shed light on Plato's 39
achievement. When Plato assigned reality to the forms, he was, in fact, identifying reality with the properties that classes of things have in common. The bearer of true reality is not (for example) this dog with the droopy left ear or that one with the menacing bark, but the idealized form of a dog shared (imperfectly, to be sure) by every individual dog—those characteristics by virtue of which we are able to classify all of them as dogs. Therefore, to gain true knowledge, we must set aside all characteristics peculiar to things as individuals and seek the shared characteristics that define them into classes. Now stated in this modest fashion, Plato's view has a distinctly modern ring. Idealization is a prominent feature of a great deal of modern science; we develop models or laws that overlook the incidental in favor of the essential. However, Plato went beyond this, maintaining not merely that true reality is to be found in the common properties of classes of things, but also that this common property (the idea or form) has objective, independent, and indeed prior existence.

[. . .]

from

The Beginnings of Western Science

by David C. Lindberg

CHAPTER THREE
ARISTOTLE'S PHILOSOPHY OF NATURE

Life and Works

Aristotle (fig. 3.1) was born in 384 B.C. in the northern Greek town of Stagira, into 1
a privileged family. His father was personal physician to the Macedonian king, Amyntas II (grandfather of Alexander the Great). Aristotle had the advantage of an exceptional education: at age seventeen, he was sent to Athens to study with Plato. He remained in Athens as a member of Plato's Academy for twenty years, until Plato's death about 347. Aristotle then spent several years in travel and study, crossing the Aegean Sea to Asia Minor and its coastal islands. During this period he undertook biological studies, and he encountered Theophrastus (from the island of Lesbos), who was to become his pupil and lifetime colleague. He returned to Macedonia in 342 to become the tutor of the young Alexander (later "the Great"). In 335, when Athens fell under Macedonian rule, Aristotle returned to the city and

From David C. Lindberg, *The Beginnings of Western Science: The European Scientific Tradition in Philosophical, Religious, and Institutional Context, Prehistory to A.D. 1450* (Second Edition), published by The University of Chicago Press ©1992, 2007. Reproduced with permission.

began to teach in the Lyceum, a public garden frequented by teachers. He remained there, establishing an informal school, until shortly before his death in 322.

2 In the course of his long career as student and teacher, Aristotle systematically and comprehensively addressed the major philosophical issues of his day. He is credited with more than 150 treatises, approximately 30 of which have come down to us. The surviving works appear to consist mainly of lecture notes or unfinished treatises not intended for wide circulation; whatever their exact origin, they were obviously directed to other philosophers, including advanced students. In modern translation, they occupy well over a foot of bookshelf, and they present a philosophical system overwhelming in power and scope. It is out of the question for us to survey the whole of Aristotle's philosophy, and we must be content with examining the fundamentals of his philosophy of nature—beginning with his response to positions taken by the pre-Socratics and Plato.

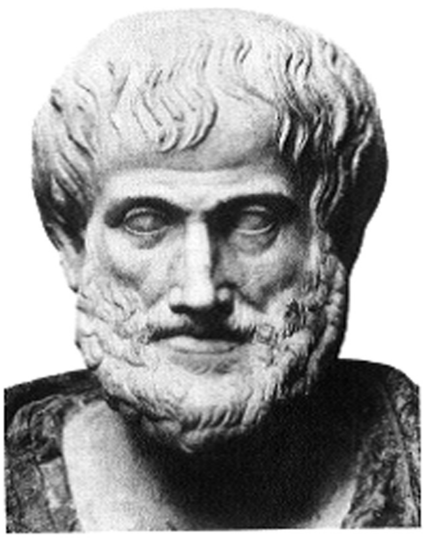


Fig. 3.1. Aristotle. Museo Nazionale, Rome. Alinari/Art Resource N.Y.

Metaphysics and Epistemology

3 Through his long association with Plato, Aristotle had, of course, become thoroughly versed in Plato's theory of forms. Plato had drastically diminished (without totally rejecting) the reality of the material world observed by the senses. Reality in

its perfect fullness, Plato argued, is found only in the eternal forms, which are dependent on nothing else for their existence. The objects that make up the sensible world, by contrast, derive their characteristics and their very being from the forms; it follows that sensible objects exist only derivatively or dependently.

Aristotle refused to accept this diminished, dependent status that Plato assigned 4
to sensible objects. They must exist fully and independently, for in Aristotle's view they were what make up the real world. Moreover, the traits that give an individual object its character do not, Aristotle argued, have a prior and separate existence in a world of forms, but belong to the object itself. There is no perfect form of a dog, for example, existing independently in the world of forms and replicated imperfectly in individual dogs, imparting to them their attributes. For Aristotle, there were just individual dogs. These dogs certainly shared a set of attributes—for otherwise we would not be entitled to call them "dogs"—but these attributes exist in, and belong to, individual dogs.

Perhaps this way of viewing the world has a familiar ring. Making individual 5
sensible objects the primary realities ("substances," Aristotle called them) will seem like good common sense to most readers of this book, and probably struck Aristotle's contemporaries the same way. But if it makes good common sense, can it also be good philosophy? That is, can it deal successfully, or at least plausibly, with the difficult philosophical issues raised by the pre-Socratics and Plato—the nature of the fundamental reality, epistemological concerns, and the problem of change and stability? Let us take up these problems one by one.

The decision to locate reality in sensible, corporeal objects does not yet tell 6
us very much about reality—only that we should look for it in the sensible world. Already in Aristotle's day, any philosopher would demand to know more: one thing he would demand to know was whether the corporeal materials of daily experience (wood, water, air, stone, metal, flesh, etc.) are themselves the fundamental, irreducible constituents of things, or whether they are composites of still more fundamental stuff. Aristotle addressed this question by drawing a distinction between properties and their subjects. He maintained (as most of us would) that a property has to be the property *of* something; we call that something its "subject." To be a property is to belong to a subject; properties cannot exist independently.

Individual corporeal objects, then, have both properties (color, weight, 7
texture, and the like) and something other than properties to serve as their subject.

These two roles are played by "form" and "matter," respectively. Corporeal objects are "composites" of form and matter—form consisting of the properties that make the thing what it is, matter serving as the subject or substratum for the form. A white rock, for example, is white, hard, heavy, and so forth, by virtue of its form; but matter must also be present, to serve as subject for the form, and this matter brings no properties of its own to its union with form. (Aristotle's doctrine will be further discussed in chap. 12, below, in connection with medieval attempts to clarify and extend it.)

8 We can never, in actuality, separate form and matter; they are presented to us only as a unitary composite. If they were separable, we should be able to put the properties (no longer the properties *of* anything) in one pile, the matter (absolutely propertyless) in another—an obvious impossibility. But if form and matter can never be separated, is it not meaningless to speak of them as the real constituents of things? Isn't this a purely logical distinction, existing in our minds, but not in the external world? Surely not for Aristotle, and perhaps not for us; most of us would think twice before denying the real existence of cold or red, although we can never collect a bucket of either one. In short, Aristotle once again surprises us by using commonsense notions to build a persuasive philosophical edifice.

9 Aristotle's claim that the primary realities are concrete individuals surely has epistemological implications, since true knowledge must be knowledge of truly real things. By this criterion, Plato's attention was naturally directed toward the eternal forms, knowable through reason or philosophical reflection. Aristotle's metaphysics of concrete individuals, by contrast, directed his quest for knowledge toward the material world of individuals, of nature, and of change—a world encountered through the senses.

10 Aristotle's epistemology is complex and sophisticated. It must suffice here to indicate that the process of acquiring knowledge begins with sense experience. From repeated sense experience follows memory; and from memory, by a process of "intuition" or insight, the experienced investigator is able to discern the universal features of things. By the repeated observation of dogs, for example, an experienced dog breeder comes to know what a dog really is; that is, he comes to understand the form or definition of a dog, the crucial traits without which an animal cannot be a dog. Note that Aristotle, no less than Plato, was determined to grasp the universal traits or properties of things; but, unlike his teacher, Aristotle argued that one must

start with the individual material thing. Once we grasp the universal properties or definition, we can put it to use as the premise of deductive demonstrations.

Knowledge is thus gained by a process that begins with experience (a term broad enough, in some contexts, to include common opinion or the reports of distant observers). In that sense knowledge is empirical; nothing can be known apart from such experience. But what we learn by this "inductive" process does not acquire the status of true knowledge until put into deductive form; the end product is a deductive demonstration (nicely illustrated in a Euclidean proof) beginning from universal definitions as premises. Although Aristotle discussed both the inductive and deductive phases (the latter far more than the former) in the acquisition of knowledge, he stopped considerably short of later methodologists, especially in the analysis of induction. 11

This is the theory of knowledge outlined by Aristotle in the abstract. Is it also the method actually employed in Aristotle's own scientific investigations? Probably not—with perhaps an occasional exception. Like modern scientists, Aristotle did not proceed by following a methodological recipe book, but rather by rough and ready methods, familiar procedures that had proved themselves in practice. Somebody has defined science as "doing your damnedest, no holds barred"; when it came (for example) to his extensive biological researches, this is exactly what Aristotle did. It is not a surprise, and certainly no character defect, that Aristotle should, in the course of thinking about the nature and the foundations of knowledge, formulate a theoretical scheme (an epistemology) not perfectly consistent with his own scientific practice. 12

Nature and Change

The problem of change had become a celebrated philosophical issue (within the quite small community of philosophers) in the fifth century B.C. In the fourth century, Plato had dealt with it by restricting change to the imperfect material replica of the changeless world of forms. For Aristotle, a distinguished naturalist who was philosophically committed to the full reality of the changeable individuals that make up the sensible world, the problem of change was a most pressing one. 13

Aristotle's starting point was the commonsense assumption that change is genuine. But this does not, by itself, get us very far; it remains to be demonstrated 14 that the idea of change can withstand philosophical scrutiny; it must also be shown how change can be explained. Aristotle had various weapons in his arsenal by which to achieve these ends. The first was his doctrine of form and matter. If every object is constituted of form and matter, then Aristotle could make room for both change and stability by arguing that when an object undergoes change, its form changes (by a process of replacement, the new form replacing the old one) while its matter remains unchanged. Aristotle went on to argue that change in form takes place between a pair of opposites or contraries, one of which is the form to be achieved, the other its privation or absence. When the dry becomes wet, or the cold becomes hot, this is change from privation (dry or cold) to the intended form (wet or hot). Change, for Aristotle, is thus never random, but confined to the narrow corridor connecting pairs of contrary qualities; order is thus discernible even in the midst of change.

A determined Parmenidean might protest that to this point the analysis does nothing to escape Parmenides' objection to all change on the ground that inevitably it calls for the emergence of something out of nothing. Aristotle's reply is found in his doctrine of potentiality and actuality. Aristotle would undoubtedly have granted that if the only two possibilities are being and nonbeing—that is, if things either exist or do not exist—then the transition from non-hot to hot would indeed involve passage from nonbeing to being (the nonexistence of hot to the existence of hot) and would thus be vulnerable to Parmenides' objection. But Aristotle believed that the objection could be successfully circumvented by supposing that there are three categories associated with being instead of two: not just being and nonbeing, but (1) nonbeing, (2) potential being, and (3) actual being. If such is the state of things, then change can occur between potential being and actual being without nonbeing ever entering the picture. What Aristotle has in mind is perhaps most easily illustrated by examples from the biological realm. An acorn is potentially, but not actually, an oak tree. In becoming an oak tree, it becomes actually what it originally was only potentially. The change thus involves passage from potentiality to actuality—not from nonbeing to being, but from one kind or degree of being to another. Or for a pair of nonbiological examples, a heavy body held above the earth falls in order to fulfill its potential of being situated with other heavy things about the center of the universe. And a sculptor, with mallet and hammer, reveals in actuality a shape that existed potentially within the original block of marble.

If these arguments allow us to escape the logical dilemmas associated with the idea of change, and therefore to believe in its possibility, they do not yet tell us anything about the cause of change. Why should an acorn move from the status of potential oak tree to that of actual oak tree, or an object change from black to white, rather than remaining in its original state? Aristotle answered with an intricate, subtle, and not always consistent, theory of nature and causation. Given these difficulties, we will spare ourselves the pain of an exhaustive account and treat ourselves to the short version. 16

The world we inhabit is an orderly one, in which things generally behave in predictable ways, Aristotle argued, because every natural object has a "nature"—an attribute (associated primarily with form) that makes the object behave in its customary fashion, provided no insurmountable obstacle intervenes—or, as a modern commentator has put it, "that within a thing which determines basically what that thing does when it is being itself." For Aristotle, a brilliant zoologist, the growth and development of biological organisms were easily explained by the activity of such an inner driving force. An acorn becomes an oak tree because its nature is to do so. But the theory was applicable beyond biological growth and, indeed, beyond the biological realm altogether. Dogs bark, rocks fall, and marble yields to the hammer and chisel of the sculptor because of their respective natures. Ultimately, Aristotle argued, all change and motion in the universe can be traced back to the natures of things. For the natural philosopher, who by definition is interested in change and things capable of undergoing change, these natures are the central object of study. 17

To this general statement of Aristotle's theory of "nature," we need to add a qualification—namely, that an artificially produced object is a special case, for such an object possesses no nature other than the natures of its ingredients. If a chariot is constructed of wood and iron, the nature of wood and nature of iron do not yield to a composite "nature of a chariot." By contrast, in the organic world the natures of the organs and tissues that make up an organism yield to the nature of the organism as a whole. The nature of the human body is not the sum of the natures of its various tissues and organs, but a unique nature characteristic of that living human as an organic whole. 18

With this theory of nature in mind, we can understand a feature of Aristotle's scientific practice that has puzzled and distressed modern commentators and 19 critics—namely, the absence from his work of anything resembling controlled experimentation. Unfortunately, such criticism overlooks Aristotle's aims, which drastically limited his methodological options. If, as Aristotle believed, the nature of a thing is to be discovered through the behavior of that thing in its natural, unfettered state, then artificial constraints will merely interfere and corrupt. If, despite interference, the object behaves in its customary fashion, we have troubled ourselves for no purpose. If we set up conditions that prevent the nature of an object from revealing itself, all we have learned is that it can be interfered with to the point of remaining concealed. Contrived experimentation violates, rather than reveals, the natures of things. Aristotle's scientific practice is not to be explained, therefore, as a result of stupidity or deficiency on his part—failure to perceive an obvious procedural improvement—but as a method compatible with the world as he perceived it and suited to the questions that interested him. Experimental science emerged not when, at long last, the human race produced somebody clever enough to perceive that artificial conditions would assist in the exploration of nature, but when a rich variety of conditions were fulfilled—including the emergence of questions to which such a procedure promised to provide answers.

20 To complete our analysis of Aristotle's theory of change, we must briefly consider the celebrated four Aristotelian causes. To understand a change or the production of an artifact is to know its causes (perhaps best translated "explanatory conditions and factors"). There are four of these: the form of a thing; the matter underlying that form, which persists through the change; the agency that brings about the change; and the purpose served by the change. These are called, respectively, formal cause, material cause, efficient cause, and final cause. To take an extremely simple example—the production of a statue—the formal cause is the shape given the marble, the material cause is the marble that receives this shape, the efficient cause is the sculptor, and the final cause is the purpose for which the statue is produced (perhaps the beautification of Athens or the celebration of one of its heroes). There are cases in which identifying one or another of the causes is difficult, or in which one or more causes merge, but Aristotle was convinced that his four causes provided an analytical scheme of general applicability.

21 We have said enough about the form-matter distinction to make clear what was meant by "formal" and "material" causes, and "efficient" cause is close enough to modern notions of causation to require no further comment; but "final"

cause requires explanation. In the first place, the expression "final cause" is an English cognate derived from the Latin word *finis*, meaning "goal," "purpose," or "end," and it has nothing to do with the fact that it often appears last in the list of Aristotelian causes. Aristotle argued, quite rightly, that many things cannot be understood without knowledge of purpose or function. To explain the arrangement of teeth in the mouth, for example, we must understand their functions (sharp teeth in front for tearing, molars in back for grinding). Or to take an example from the inorganic realm, it is not possible to grasp why a saw is made as it is without knowing the function the saw is meant to serve. Aristotle went so far as to give final cause priority over material cause, noting that the purpose of the saw determines the material (iron) of which it must be made, whereas the fact that we possess a piece of iron does nothing to determine that we will make it into a saw.

Perhaps the most important point to be made about final cause is its clear 22
illustration of the role of purpose (the more technical term is "teleology") in Aristotle's universe. The world of Aristotle is not the inert, mechanistic world of the atomists, in which the individual atom pursues its own course mindless of all others. Aristotle's world is not a world of chance and coincidence, but an orderly, organized world, a world of purpose, in which things develop toward ends determined by their natures. It would be unfair and pointless to judge Aristotle's success by the degree to which he anticipated modern science (as though his goal was to answer our questions, rather than his own); it is nonetheless worth noting that the emphasis on functional explanation to which Aristotle's teleology leads would prove to be of profound significance for all of the sciences and remains to this day a dominant mode of explanation within the biological sciences.

Cosmology

Aristotle not only devised methods and principles by which to investigate and 23
understand the world: form and matter, nature, potentiality and actuality, and the four causes. In the process, he also developed detailed and influential theories regarding an enormous range of natural phenomena, from the heavens above to the earth and its inhabitants below.

Let us start with the question of origins. Aristotle adamantly denied the 24
possibility of a beginning, insisting that the universe must be eternal. The

alternative—that the universe came into being at some point in time—he regarded as unthinkable, violating (among other things) Parmenidean strictures about something coming from nothing. Aristotle's position on this question would prove troublesome for medieval Christian Aristotelians.

25 Aristotle considered this eternal universe to be a great sphere, divided into an upper and a lower region by the spherical shell in which the moon is situated. Above the moon is the celestial region; below is the terrestrial region; the moon, spatially intermediate, is also of intermediate nature. The terrestrial or sublunar region is characterized by birth, death, and transient change of all kinds; the celestial or supralunar region, by contrast, is a region of eternally unchanging cycles. That this scheme had its origin in observation would seem clear enough; in his *On the Heavens*, Aristotle noted that "in the whole range of time past, so far as our inherited records reach, no change appears to have taken place either in the whole scheme of the outermost heaven or in any of its proper parts." If in the heavens we observe eternally unvarying circular motion, he continued, we can infer that the heavens are not made of the terrestrial elements, the nature of which (observation reveals) is to rise or fall in transient rectilinear motions. The heavens must consist of an incorruptible fifth element (there are four terrestrial elements): the quintessence (literally, the fifth essence) or aether. The celestial region is completely filled with this quintessence (no void space) and divided, as we shall see, into concentric spherical shells bearing the planets. It had, for Aristotle, a superior, quasi-divine status.

26 The sublunar region is the scene of generation, corruption, and impermanence. Aristotle, like his predecessors, inquired into the basic element or elements to which the multitude of substances found in the terrestrial region can be reduced. He accepted the four elements originally proposed by Empedocles and subsequently adopted by Plato—earth, water, air, and fire. He agreed with Plato that these elements are in fact reducible to something even more fundamental; but he did not share Plato's mathematical inclination and therefore refused to accept Plato's regular solids and their constituent triangles. Instead, he expressed his own commitment to the reality of the world of sense experience by choosing sensible qualities as the ultimate building blocks. Two pairs of qualities are crucial: hot-cold and wet-dry. These combine in four pairs, each of which yields one of the elements (see fig. 3.2). Notice the use made once again of contraries. There is nothing to forbid

any of the four qualities being replaced by its contrary, as the result of outside influence. If water is heated, so that the cold of water yields to hot, the water is transformed into air. Such a process easily explains changes of state (from solid to liquid to vapor, and conversely), but also more general transmutation of one substance into another. On such a theory as this, alchemists could easily build.

cold and dry = earth
cold and wet = water
hot and wet = air
hot and dry = fire

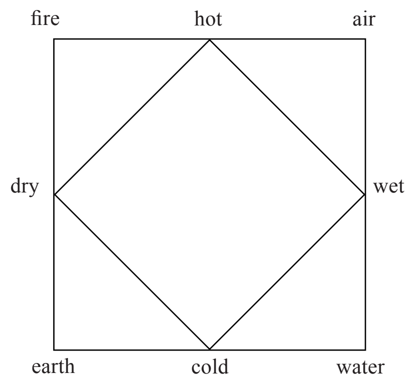


Fig. 3.2. Square of opposition of the Aristotelian elements and qualities. For a medieval (9th c.) version of this diagram, see John E. Murdoch, Album of Science: Antiquity and the Middle Ages*, p. 352.*

The various substances that make up the cosmos totally fill it, leaving 27
no empty space. To appreciate Aristotle's view, we must lay aside our almost automatic inclination to think atomistically; we must conceive material things not as aggregates of tiny particles but as continuous wholes. If it is obvious that, say, a loaf of bread is composed of crumbs separated by small spaces, there is no reason not to suppose that those spaces are filled by some finer substance, such as air or water. And there is certainly no simple way of demonstrating, nor indeed any obvious reason for believing, that water and air are anything but continuous. Similar reasoning, applied to the whole of the universe, led Aristotle to the conclusion that the universe is full, a *plenum*, containing no void space. This claim would be attacked by medieval scholars.

Aristotle defended this conclusion with a variety of arguments, such as the 28
following. The speed of a falling body is dependent on the density of the medium through which it falls—the less the density, the swifter the motion of the falling body. It follows that in a void space (density zero), there is nothing to slow the

descent of the body, from which we would be forced to conclude that the body would fall with infinite speed—a nonsensical notion, since it implies that the body could be at two places at the same time. Critics have frequently noted that this argument can just as well be taken to prove that the absence of resistance does not entail infinite speed as to prove that void does not exist. The point is, of course, well taken. However, we need to understand that Aristotle's denial of the void did not rest on this single piece of reasoning. In fact, this was but one small part of a lengthy campaign against the atomists, in which Aristotle battled the notion of void space (or void place) with a variety of arguments, some more and some less persuasive.

29 In addition to being hot or cold and wet or dry, each of the elements is also heavy or light. Earth and water are heavy, but earth is the heavier of the two. Air and fire are light, fire being the lighter of the two. In assigning levity to two of the elements, Aristotle did not mean (as we might, if we were making the claim) simply that they are less heavy, but that they are light in an absolute sense; levity is not a weaker version of gravity, but its contrary. Because earth and water are heavy, it is their nature to descend toward the center of the universe; because air and fire are light, it is their nature to ascend toward the periphery (that is, the periphery of the terrestrial region, the spherical shell that contains the moon). If there were no hindrances, therefore, earth and water would collect at the center; because of its greater heaviness, earth would achieve a lower position, forming a sphere at the very center of the universe; water would collect in a concentric spherical shell just outside it. Air and fire naturally ascend, but fire, owing to its greater levity, occupies the outermost region, with air as a concentric sphere just inside it. In the ideal case (in which there are no mixed bodies and nothing prevents the natures of the four elements from fulfilling themselves), the elements would thus form a set of concentric spheres: fire on the outside, followed by air and water, and finally earth at the center (see fig. 3.3). But in reality, the world is composed largely of mixed bodies, one always interfering with another, and the ideal is never attained. Nonetheless, the ideal arrangement defines the natural place of each of the elements; the natural place of earth is at the center of the universe, of fire just inside the sphere of the moon, and so forth.

30 It must be emphasized that the arrangement of the elements is spherical. Earth collects at the center to form *the earth*, and it too is spherical. Aristotle defended

this belief with a variety of arguments. Arguing from his natural philosophy, he pointed out that since the natural tendency of earth is to move toward the center of the universe, it must arrange itself symmetrically about that point. But he also called attention to observational evidence, including the circular shadow cast by the earth during a lunar eclipse and the fact that north-south motion by an observer on the surface of the earth alters the apparent position of the stars. Aristotle even reported an estimate by mathematicians of the earth's circumference (400,000 stades = about 45,000 miles, roughly 1.8 times the modern value). The sphericity of the earth, thus defended by Aristotle, would never be forgotten or seriously questioned. The widespread myth that medieval people believed in a flat earth is of modern origin.

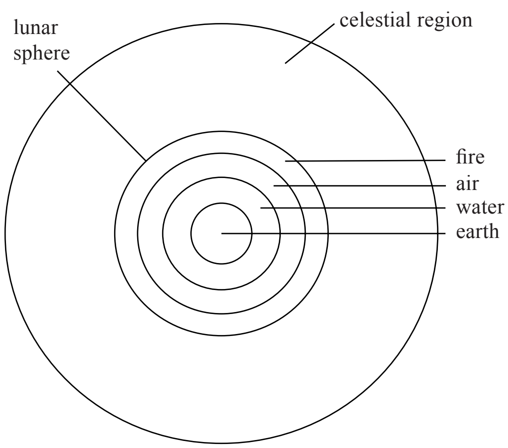


Fig. 3.3.The Aristotelian cosmos.

Finally, we must note one of the implications of this cosmology, namely that 31
space, instead of being a neutral, homogeneous backdrop (analogous to our modern notion of geometrical space) against which events occur, has properties. Or to express the point more precisely, ours is a world of space, whereas Aristotle's was a world of place. Heavy bodies move toward their place at the center of the universe not because of a tendency to unite with other heavy bodies located there, but simply because it is their nature to seek that central place; if by some miracle the center happened to be vacant (a physical impossibility in an Aristotelian universe, but an interesting imaginary state of affairs), it would remain the destination of every heavy body.

Motion, Terrestrial and Celestial

32 We can best understand Aristotle's theory of motion by grasping its two most fundamental claims. The first is that motion is never spontaneous; there is no motion without a mover. The second is the distinction between two types of motion: motion toward the natural place of the moving body is "natural" motion; motion in any other direction occurs only under coercion from an outside force and is therefore a "forced" or "violent" motion.

33 The mover in the case of natural motion is the nature of the body, which is responsible for its tendency to move toward its natural place as defined by the ideal spherical arrangement of the elements. Mixed bodies have a directional tendency that depends on the proportion of the various elements in their composition. When a body undergoing natural motion reaches its natural place, its motion ceases. The mover in the case of forced motion is an external force, which compels the body to violate its natural tendency and move in a direction or manner other than straight-line motion toward its natural place. Such motion ceases when the external force is withdrawn.

34 So far, this seems sensible. One obvious difficulty, however, is to explain why a projectile hurled horizontally, and therefore undergoing forced motion, does not come to an immediate halt when it loses contact with whatever propelled it. Aristotle's answer was that the medium takes over as mover. When we project an object, we also act on the surrounding medium (air, for instance), imparting to it the power to move objects; this power is communicated from part to part, in such a way that the projectile is always in contact with a portion of the medium capable of keeping it in motion. If this seems implausible, consider the greater implausibility (from Aristotle's standpoint) of the alternative—that a projectile, which is inclined by nature to move toward the center of the universe, moves horizontally or upward despite the fact that there is no longer anything causing it to do so.

35 Force is not the only determinant of motion. In all real cases of motion in the terrestrial realm, there will also be a resistance or opposing force. And it seemed clear to Aristotle that the quickness of motion must depend on these two determining factors—the motive force and the resistance. The question arose: what is the relationship between force, resistance, and speed? Although it probably did not occur to Aristotle that there might be a quantitative law of universal applicability,

he was not without interest in the question and did make several forays into quantitative territory. In reference to natural motion in his *On the Heavens* and again in his *Physics*, Aristotle claimed that when two bodies of differing weight descend, the times required to cover a given distance will be inversely proportional to the weights. (A body twice as heavy will require half the time). In the same chapter of the *Physics*, Aristotle introduced resistance into the analysis of natural motion, arguing that if bodies of equal weight move through media of different densities, the times required to traverse a given distance are proportional to the densities of the respective media; that is, the greater the resistance the slower the body moves. Finally, Aristotle also dealt with forced motion in his *Physics*, claiming that if a given force moves a given weight (against its nature) for a given distance in a given time, the same force will move half that weight twice the distance in that same time (or the same distance in half that time); alternatively, half the force will move half the weight the same distance in the same time.

From such statements, some of Aristotle's successors have made a determined 36
effort to extract a general law. This law is customarily stated as:

$$v \propto F/R.$$

That is, velocity (v) is proportional to the motive force (F) and inversely proportional to the resistance (R). For the special case of the natural descent of a heavy body, the motive force is the weight (W) of the body, and the relationship then becomes:

$$v \propto W/R.$$

Such relationships probably do no great violence to Aristotle's intent for most cases of motion; however, giving them mathematical form, as we have done, suggests that they hold for all values of v, F (or W), and R—a claim that Aristotle would certainly have denied. He stated explicitly, for example, that a resistance equal to the motive force will prevent motion altogether, whereas the formula above offers no such result. Moreover, the appearance of velocity in these relationships seriously misrepresents Aristotle's conceptual framework, which contained no concept of velocity as a quantifiable measure of motion, but described motion only in terms of distances and times. Velocity as a technical scientific term to which numerical values might be assigned was a contribution of the Middle Ages (see below, chap. 12).

37 Aristotle has been severely criticized for this theory of motion, on the assumption that any sensible person should have recognized its fatal flaws. Is such criticism justified? In the first place, our goal is to understand the behavior, beliefs, and achievements of historical actors against the background of the culture in which they lived, rather than to assess credit or blame according to the degree to which those historical actors resemble us. In short, historians must always contextualize their subjects. Second, some of the criticisms of Aristotle's theories of motion apply only to theories foisted onto Aristotle by followers and critics, rather than to his own. Third, the theory in its genuinely Aristotelian (and properly contextualized) version makes quite good sense today and would surely have made good sense in the fourth century B.C. For example, various surveys have shown that the majority of modern, university-educated people are prepared to assent to many of the basics of Aristotle's theory of motion. Fourth, the relatively modest level of quantitative content in Aristotle's theory is easily explained as the outcome of his larger philosophy of nature. His primary goal was to understand essential natures, not to explore quantitative relationships between such incidental factors as the space-time (or place-time) coordinates applicable to a moving body; even an exhaustive investigation of the latter gives us no useful information about the former. You may criticize Aristotle, if you like, for not being interested in whatever interests modern scientists, but we do not thereby learn anything significant about Aristotle.

38 Motion in the celestial sphere is an altogether different sort of phenomenon. The heavens, composed of the incorruptible quintessence, possess no contraries and are therefore incapable of qualitative change. It might seem fitting for such a region to be absolutely motionless, but this hypothesis is defeated by the most casual observation of the heavens. Aristotle therefore assigned to the heavens the most perfect of motions—continuous uniform circular motion. Besides being the most perfect of motions, uniform circular motion appears to have the capability of explaining the observed celestial cycles.

39 By Aristotle's day, these cycles had been an object of study for centuries in the Greek world and for millennia in its predecessor civilizations. It was understood that the "fixed" stars move with perfect uniformity, as though fixed to a uniformly rotating sphere, with a period of rotation of approximately one day. But there

were seven stars, the wandering stars or planets, that displayed a more intricate motion, apparently crawling around on the stellar sphere as it went through its daily rotation. These seven were the Sun, Moon, Mercury, Venus, Mars, Jupiter, and Saturn. The sun crawls slowly (about 1°/day), west to east with small variations in speed, through the sphere of fixed stars along a path called the ecliptic, which passes through the center of the zodiac [. . .]. The moon follows approximately the same course, but at the more rapid rate of about 12°/day. The remaining planets also move along the ecliptic (or in its vicinity) with variable speed and with an occasional reversal of direction.

Are such complex motions compatible with the requirement of uniform 40
circular motion in the heavens? Eudoxus, a generation before Aristotle, had already shown that they are. I will return to this subject in chap. 5; for the moment, it will be sufficient to point out that Eudoxus treated each complex planetary motion as a composite of a series of simple uniform circular movements. He did this by assigning to each planet a set of concentric spheres, and to each sphere one component of the complex planetary motion. Aristotle took over this scheme, with various modifications. When he was finished, he had produced an intricate piece of celestial machinery, consisting of fifty-five planetary spheres plus the sphere of the fixed stars.

What is the cause of movement in the heavens? Aristotle's natural philosophy 41
would not allow such a question to go unasked. The celestial spheres are composed, of course, of the quintessence; their motion, being eternal, must be natural rather than forced. The cause of this eternal motion must itself be unmoved, for if we do not postulate an unmoved mover, we quickly find ourselves trapped in an infinite regress: a moving mover must have acquired its motion from yet another moving mover, and so on. Aristotle identified the unmoved mover for the planetary spheres as the "Prime Mover," a living deity representing the highest good, wholly actualized, totally absorbed in self-contemplation, nonspatial, separated from the spheres it (or he or she) moves, and not at all like the traditional anthropomorphic Greek gods. How, then, does the Prime Mover or Unmoved Mover cause motion in the heavens? Not as efficient cause, for that would require contact between the mover and the moved, but as final cause. That is, the Prime Mover is the object of desire for the celestial spheres, which endeavor to imitate its changeless perfection by assuming eternal, uniform circular motions. Any reader who has followed this much of Aristotle's discussion would be justified in assuming that there is

a single Unmoved Mover for the entire cosmos; it comes as something of a surprise, therefore, when Aristotle announces that, in fact, each of the celestial spheres has its own Unmoved Mover, the object of its affection and the final cause of its motion.

[. . .]

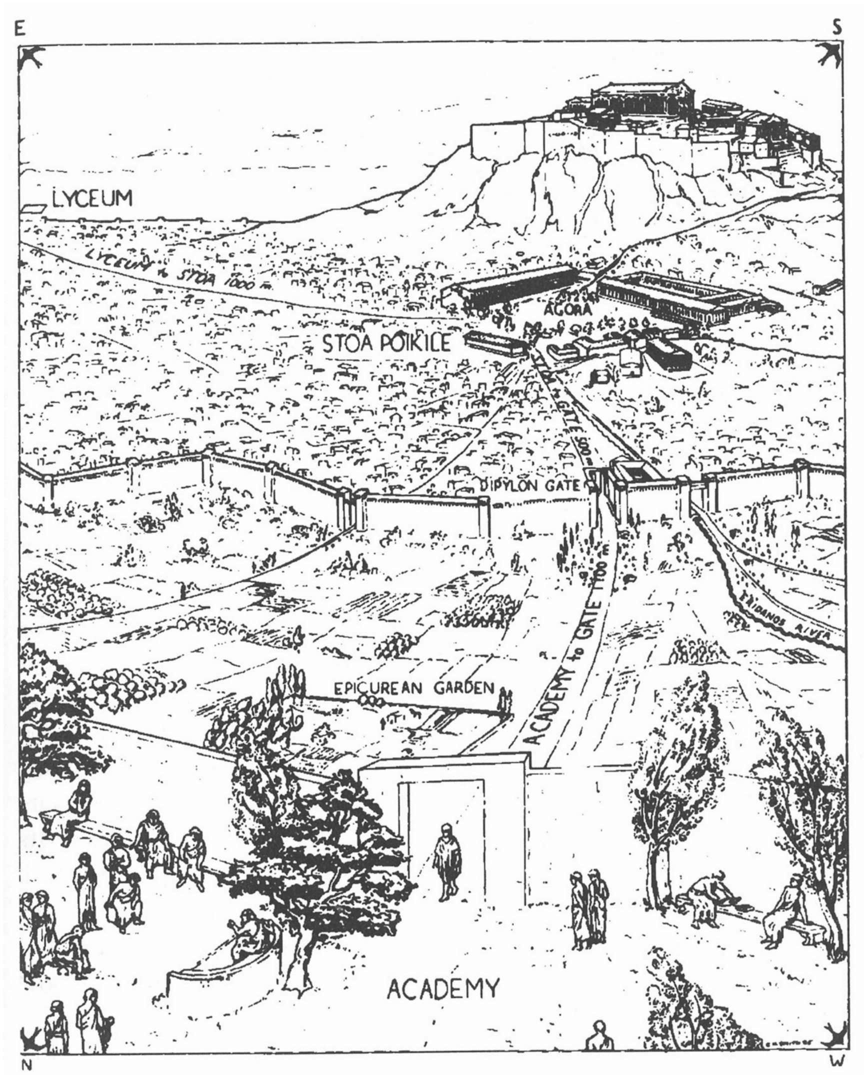


Fig. 4.1. The schools of Hellenistics Athens. © Candace H. Smith.

[. . .]

CHAPTER TWELVE
THE PHYSICS OF THE SUBLUNAR REGION

Medieval physics was not a primitive version of modern physics and cannot be legitimately judged by comparison with its modern namesake. Certainly there is overlap between the two, but medieval physics, deeply engaged with Aristotelian metaphysics and natural philosophy devoted much more attention to fundamental issues that we would classify as "metaphysical" or "philosophical"; concerning the fundamental stuff of the universe, the elements and their constituents, the sources of motion and change, and the like. 1

The medieval natural philosopher (as I will generally refer to him) took his starting point from the text of Aristotle's *Physics* and other works, and devoted his career to clarification of ambiguities, disputation about difficult or contentious portions of the text, and original application or extension of Aristotelian principles. But he was emphatically *not* a slave to the Aristotelian text, as a widespread myth would have it; rather, he was typically a gifted reader and interpreter of the texts of Aristotle and his commentators (including critics), eager to display his logical and creative powers in discussion and debate. There were theological corners of which he needed to be wary, but otherwise the medieval professor had the freedom to go where reason and experience led. 2

[. . .]

Mathematical Description of Motion

Today the application of mathematics to motion needs no defense. Theoretical mechanics, the parent discipline of theories of motion, is mathematical by definition, and to anybody with a grasp of modern physics the mathematical way would seem to be the only way. But perhaps it is only by hindsight and from a modern perspective that this conclusion is obvious; it would not have seemed plausible to many who worked within the Aristotelian tradition. We must remember that Aristotle and his medieval followers regarded motion as one of four kinds of change and that their analysis of change was not meant to focus on local motion, but rather to be applicable to all four classes of change. We also need to recognize that there is nothing obviously mathematical about most kinds of change. When we observe 33 sickness yielding to health, virtue replacing vice, and peace emerging from war, no numbers or geometrical magnitudes leap out at us. The generation or corruption of a substance and the alteration of a quality are not obviously mathematical processes, and it is only by heroic efforts over the centuries that scholars have found ways of placing a mathematical handle on a few kinds of change, including local motion. Let us investigate the early stages of this process in the Middle Ages.

34 The mathematization of nature, of course, had ancient proponents, including the Pythagoreans, Plato, and Archimedes; and early success was achieved in the sciences of astronomy, optics, and the balance (see chap. 5, above). It was inevitable that the success of these efforts would provide encouragement for those interested in mathematizing other subjects. Indeed, Aristotle himself was responsible for a primitive beginning of the mathematical analysis of motion in his *Physics*, where distance and time, both quantifiable, were employed as measures of motion. Aristotle argued that the quicker of two moving objects covers a greater distance in the same time or the same distance in less time, while two objects moving with equal quickness traverse equal distances in equal times. A generation after Aristotle, the mathematician Autolycus of Pitane (fl. 300 B.C.), took a further step, defining a uniform motion as one in which equal distances are traversed in equal times. It is important to note that in these ancient discussions distance and time were taken as the critical measures of motion, to which a numerical value might be assigned, while "quickness" or speed never achieved that status, remaining a vague, unquantified conception.

35 The impact of this mathematical analysis in medieval Europe can first be seen in the work of Gerard of Brussels, a mathematician who may have taught at the University of Paris in the first half of the thirteenth century. For our purposes the most important thing about Gerard's brief *Book on Motion* is the restriction of its contents to what we now call "kinematics"—the purely mathematical description of motion—as opposed to "dynamics," which is concerned with causes. This is an important distinction (bearing a resemblance to the distinction between "instrumentalism" and "realism" in astronomy), which will serve as one of the organizing principles for the remainder of our discussion of medieval theories of motion. For the moment, Gerard of Brussels is important as a harbinger of the kinematic tradition that was to develop in the Latin West.

This tradition flowered among a group of distinguished fourteenth-century 36
logicians and mathematicians affiliated with Merton College, Oxford, between about 1325 and 1350. This group included Thomas Bradwardine (d. 1349), subsequently appointed archbishop of Canterbury; William Heytesbury (fl. 1335); John of Dumbleton (d. ca. 1349); and Richard Swineshead (fl. 1340–55). To begin with, members of the Merton group made explicit the distinction between kinematics and dynamics that was implicitly present in Gerard's *Book on Motion*, noting that motion can be examined from the standpoint either of cause (dynamics) or of effect (kinematics). The Merton scholars proceeded to develop a conceptual framework and a technical vocabulary for dealing with motion kinematically. Included in this conceptual framework were the ideas of "velocity" and "*instantaneous* velocity," both treated as mathematical concepts to which magnitude could be assigned. The Mertonians distinguished between uniform motion (motion at constant velocity) and nonuniform (or accelerated) motion. They also devised a precise definition of uniformly accelerated motion identical to our own: a motion is uniformly accelerated if its velocity increases by equal increments in equal units of time. Finally, the Merton scholars developed a variety of kinematic theorems, several of which we will examine below.

Before we do that, we must consider the philosophical underpinnings of this 37
kinematic achievement. The emergence of velocity as a new measure of motion, to go along with the ancient measures (distance and time), is a development that needs to be explained. Velocity, after all, is quite an abstract conception, which did not force itself on the observer of moving bodies but had to be invented by natural philosophers and imposed on the phenomena. How did this come about? The answer is found in the philosophical analysis of qualities and their strength or intensity. The fundamental idea was that qualities or forms can exist in various degrees or intensities: there is not just a single degree of warmth or cold, but a range of intensities or degrees running from very cold to very hot. Moreover, it was acknowledged that forms or qualities can vary within this range; that is, they can be strengthened and weakened, or, to employ the technical medieval terminology, undergo intensification and remission. Now when this general discussion of qualities and their intensification and remission was transferred to the particular case of local motion (motion being conceived as a quality or something closely

analogous to a quality), the idea of velocity quickly emerged. The intensity of the quality of motion—that which measured its strength or degree—could be none other than swiftness or (to employ the technical medieval Latin term) “velocitas.” Intensification and remission of the quality of motion must then refer to variations in velocity.

38 Reflection about qualities, their intensity, and their intensification thus led the Mertonians to a new distinction: between the *intensity* of a quality (defined above) and its *quantity* (how much of it there is). An example will help us to understand this distinction: it is obvious enough in the case of heat that one hot object can be hotter than another; this is a reference to the intensity of the quality, what we call “temperature.” But we also have a conception of the quantity of heat—how much of it there is. If we have two objects at the same temperature, one of them twice as big, that larger object clearly has twice the “quantity” of this quality of heat. For fourteenth-century mathematicians, it followed that all qualities should submit to a similar analysis, possessing both a quantity (how much of the quality) and an intensity (the degree or strength of the quality). For heat, we have temperature (intensity) and calories (quantity); for weight, heaviness (quantity) and density or specific gravity (intensity); and so on. Could the same analysis be successfully applied to motion? Yes, it could, as we shall see.

39 News of the Merton College achievements in the analysis of qualities was transmitted quickly to other European intellectual centers. In the process, the analysis was enriched and clarified by the development of geometrical representation. The original analysis of qualities at Merton College was carried out verbally, in much the same way as we have been analyzing it. However, the advantages of geometrical analysis were recognized, and fairly elaborate systems of geometrical representation were eventually worked out. One of the first to develop such a system was Giovanni di Casali, a Franciscan from Bologna (who had also spent time in Cambridge), writing about 1351; a far more elaborate geometrical analysis was formulated by Nicole Oresme (d. 1382) at the University of Paris later in the same decade. An examination of Oresme’s scheme may prove as illuminating for us as it no doubt did for his medieval readers.

40 The first step was to represent the intensity of a quality by means of a line segment—a relatively easy step for medieval scholars brought up on Aristotle (who employed lines to represent time) and Euclid (who used lines to represent numerical magnitudes). If line segment AB (fig. 12.2) represents a given intensity

of some quality, then line segment AC represents twice that intensity. This is fine, but it has not yet gotten us very far. The critical next step was to employ this line to represent the intensity of the quality at any point of the subject. Take a rod AE (fig. 12.3), heated differentially, so that the heat increases uniformly from one end to the other. At point A and at whatever intervals you choose, erect a vertical line representing the intensity of heat at that point. If (as we have postulated) the temperature increases uniformly from A to E, then the figure will reveal a uniform lengthening of the vertical lines. Now Oresme made the system a good deal more abstract by substituting a horizontal line for the drawing of the rod (fig. 12.4). This has the effect of creating a generalized system of representation in which the horizontal line (called the "subject line" or the "extension") represents the subject, whatever it might be, while vertical lines represent the intensity of any quality we choose at the points of the subject where they are erected.

Fig. 12.2. The use of a line segment to represent the intensity of a quality.

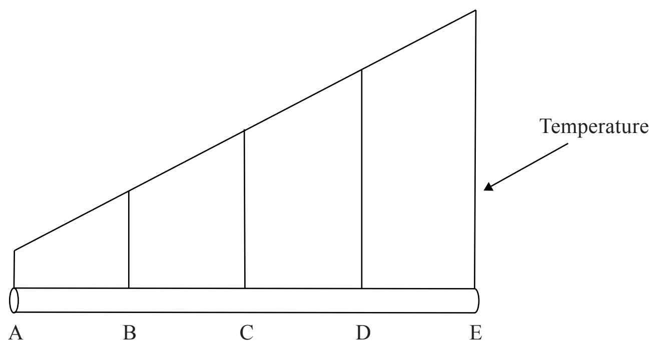


Fig. 12.3. The distribution of temperatures in a rod.

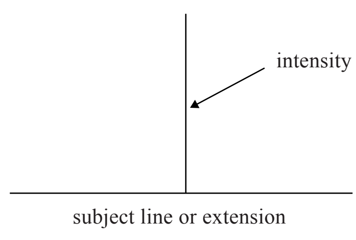


Fig. 12.4. Nicole Oresme's system for representing the distribution of any quality in a subject.

41 What Oresme has produced is a form of geometrical representation—an obvious forerunner of modern graphing techniques—in which the shape of the figure (as in fig. 12.3) informs us about variations in the intensity of a quality over its subject. But how do we make the transition from qualities in general to motion in particular? One way is to consider a body, the different parts of which move with different velocities; a rod held by a pin at one end and rotated about that pin would be a good example. In such a case, we can draw the rod horizontally and erect a perpendicular at any point, indicating the angular velocity of that point. The result will be a distribution of velocities in a subject, as in figure 12.5.

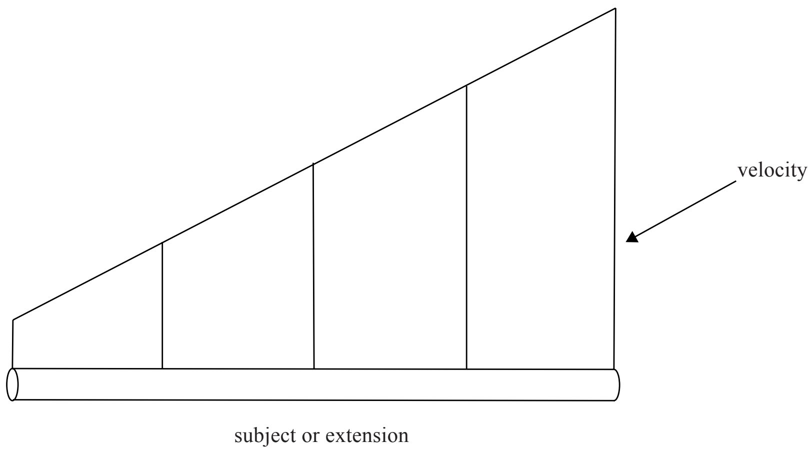


Fig. 12.5. The distribution of velocities in a rod rotating about one end.

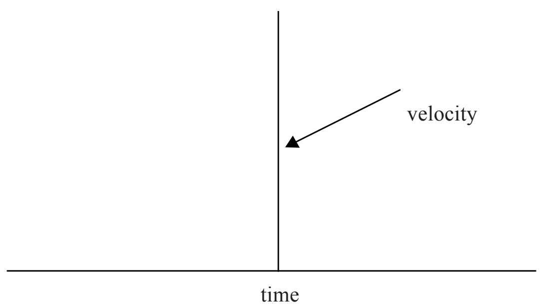


Fig. 12.6. Velocity as a function of time.

42 But there is another case, more difficult because it requires more abstract treatment. Suppose we have a body that moves as a unit, all of its parts having

the same velocity, but a velocity that varies over time. The way to understand this, Oresme explained, is to see that here the subject line is not the extension of a corporeal object, as in the examples above, but the duration of a local motion. Time becomes the subject, represented by the horizontal line. This gives us a primitive coordinate system in which velocity can be plotted as a function of time (see fig. 12.6). Oresme proceeded to discuss various configurations of velocity with respect to time. Uniform velocity will be represented by a figure in which all the vertical lines are of equal length—that is, a rectangle. Nonuniform velocity requires verticals of variable length. Within this category of nonuniform motion, we have uniformly nonuniform velocity (our uniformly accelerated motion), represented by a triangle, and nonuniformly nonuniform motion (nonuniformly accelerated motion), represented by a variety of other figures, the shapes of which are determined by the specific pattern of nonuniformity (see fig. 12.7). Finally, how did Oresme deal with that other feature of qualities noted above—their total quantity? He identified the total quantity of motion with the distance traversed; and this, he argued must be represented by the area of the figure.

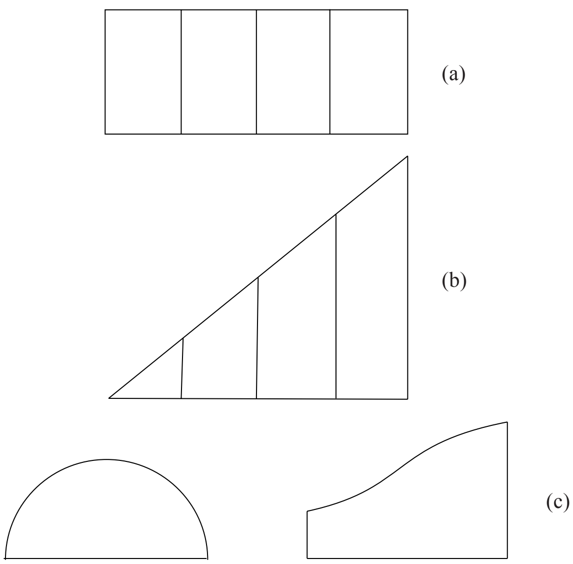


Fig. 12.7. The representation of various motions.
(a) Uniform motion.
(b) Uniformly nonuniform motion (= uniformly accelerated motion).
(c) Nonuniformly nonuniform motions (= nonuniformly accelerated motions).

43 Oresme has very cleverly put geometry to work on behalf of the representation of motions of all varieties. He and those who followed him were not content with having created the geometrical tools. They proceeded to use them to illustrate and prove kinematic theorems applicable to uniform or uniformly accelerated motion. The most important case was the latter, represented in figure 12.7(b).This case was of special interest in the fourteenth century, not because it was identified with any particular motion in the real world but because it offered a substantial mathematical challenge. Let us examine two important theorems applicable to uniformly accelerated motion that emerged from these efforts.

44 The first had already been stated, without geometrical proof or illustration, by the Merton scholars; it is now known as the "Merton rule" or the "mean-speed theorem." This theorem seeks to find a measure for uniformly accelerated motion by comparing it with uniform motion. The theorem claims that a body moving with a uniformly accelerated motion covers the same distance in a given time as if it were to move for the same duration with a uniform speed equal to its mean (or average) speed. Expressed in numerical terms, the claim is that a body accelerating uniformly from a velocity of 10 to a velocity of 30 traverses the same distance as a body moving uniformly for that same period of time with a velocity of 20. Now Oresme provided a simple but elegant geometrical proof of this theorem (fig. 12.8). The uniformly accelerated motion can be represented by triangle ACG and its mean speed by line BE. The uniform motion that is to be compared with the uniformly accelerated motion must therefore be represented by rectangle ACDF (the altitude of which is BE, the mean speed of the uniformly accelerated motion). The Merton rule claims simply that the distance traversed by the accelerated motion is equal to the distance traversed by the uniform motion. Since, in Oresme's diagrams, distance traversed is measured by the area of the figure, we can prove the theorem by showing that the area of triangle ACG equals the area of rectangle ACDF. A glance at the two figures will reveal that this is so.

45 The second theorem, like the first, aimed to elucidate the mathematical properties of uniformly accelerated motion by means of a comparison involving distances traversed. In this case, the distance covered in the first half of a uniformly accelerated motion (beginning from rest) is compared to the distance covered in the second half of the same motion; the claim was that the latter is three times the former. To prove this theorem geometrically, we need merely show that the

area of quadrangle BCGE (fig. 12.8), which represents the distance covered in the second half of the time, BC, is three times the area of triangle ABE, representing the distance traversed in the first half of the time, AB. Once again, inspection will establish that this is true.

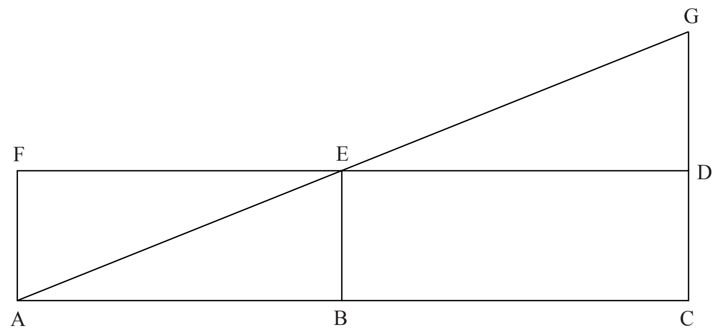


Fig. 12.8. Nicole Oresme's geometrical proof of the Merton rule.

Finally, two general points: First, we must remind ourselves that medieval kinematics was a totally abstract endeavor—much like modern mathematics. It was claimed, for example, that *if* a uniformly accelerated motion were to exist, *then* the Merton rule would apply to it. Never did a medieval scholar identify an instance of such motion in the real world. Is there a satisfactory explanation for such seemingly odd behavior? Yes, there is. Given the technology available in the Middle Ages (particularly for the measurement of time), demonstrating that a particular motion is uniformly accelerated would have been a considerable feat. Even in the twenty-first century, imagine the challenge of proving with precision that a motion is uniformly accelerated, using any or all of the resources available in your local hardware store. But perhaps more importantly, the medieval scholars who developed this kinematic analysis were mathematicians and logicians; and no more than modern mathematicians and logicians would they have thought of moving their place of labor from the study to the workshop. 46

Second, out of this purely intellectual labor came a new conceptual framework for kinematics and a variety of theorems (the Merton rule, for example) that figured prominently in the kinematics developed in the seventeenth century by Galileo—through whom they entered the mainstream of modern mechanics. Proposition 1, theorem 1 of Galileo's analysis of uniformly accelerated motion in his *Two New* 47

Sciences is the Merton rule (or mean-speed theorem). It is implausible to suppose that Galileo was ignorant of his fourteenth-century forerunners.

The Dynamics of Local Motion

48 Having dealt at length with medieval kinematics—the effort to describe motion mathematically—I conclude this discussion of medieval mechanics with a brief account of contributions to the *causal* analysis of motion. The starting point of all dynamical thought in the Middle Ages was the Aristotelian principle that moved things are always moved by a mover. We must first get clear on what this principle was taken to mean in the Middle Ages. We will then look at attempts to identify the mover in several particularly difficult cases of motion. And finally we will examine attempts to quantify the relationship between the force or power of a mover and the resulting velocity of the moved body.

49 Aristotle, readers will recall, divided motion into two categories: natural and forced. A natural motion, by which an object moves toward its natural place, apparently arises from an internal cause or principle: the nature of the body. A motion in any other direction must be a forced motion, produced by the application of an external force in continuous contact with the moved body. This seems clear enough in broad outline, but problems arose when medieval scholars attempted to identify the mover in natural motion and in one particularly troublesome case of forced motion.

50 In his *Physics*, where he gave an account of the mover for natural motion, Aristotle vacillated, suggesting first that natural motion may result from an internal cause, the nature of the body, but arguing later that the nature of the body cannot be the whole story and that the participation of an external mover is also required. Aristotle's ambivalence posed an obvious problem for his medieval followers, who felt compelled to inquire whether or not it is sufficient to affirm that the body is moved by its own nature. Avicenna and Averroes considered this explanation unacceptable on the grounds that it did not distinguish sufficiently between that which is moved (the body) and that which moves it (the nature of the body). They discovered what seemed to them an adequate alternative in the form-matter distinction, proposing that the form of the body is the mover, while its matter is the thing moved. In the West, Thomas Aquinas repudiated this solution, reminding

his readers that matter and form are inseparable and cannot be treated as distinct things. Aquinas argued instead (reviving one of Aristotle's proposals) that the mover in the case of natural motion is whatever generated the body outside its natural place to begin with; thereafter the body requires no mover but simply does what comes naturally: moving toward its natural place. The debate over this issue continued through the later Middle Ages, with no clear victor.

The particular case of forced motion that proved troublesome was that of 51
projectiles; the problem was to explain their continued motion after they lose contact with the original projector (e.g., the hand that threw the rock). Aristotle had assigned causation to the medium, arguing that the projector simultaneously projects the projectile and endows the surrounding medium with the power to produce motion; this power is transmitted from part to part in such a way that the projectile is always surrounded by a portion of the medium capable of moving it. It was clear, according to this account, that an external force, continuously in contact with the projectile, is required.

The first major opposition to Aristotle's explanation came in the commentary 52
on Aristotle's *Physics* by the sixth-century Alexandrian Neoplatonist philosopher John Philoponus (d. after 575), to whom it seemed that the medium serves as resistance rather than mover and who doubted that it could serve both functions simultaneously. As a Neoplatonist and a dedicated anti-Aristotelian, Philoponus launched a broad attack on Aristotelian natural philosophy, including the notion that forced motions require external movers. He proposed, rather, that all motions, natural and forced alike, are the result of internal movers. Therefore, when a projectile is hurled, the projector impresses on the projectile an "incorporeal motive force," and this internal force is responsible for its motion. If this seems an improbable answer, consider the motion of living things, which are apparently moved by internal, rather than external, forces.

Although Philoponus's impressed motive force had radically anti-Aristotelian 53
origins, it was eventually absorbed into the medieval Aristotelian tradition. Philoponus's commentary on Aristotle's *Physics* had an influential career in Arabic translation and seems to have had an indirect impact on medieval Latin thought, although the details of transmission remain to be fully traced. In the thirteenth century, theories bearing a close resemblance to that of Philoponus were discussed and rejected by Roger Bacon and Thomas Aquinas. In the fourteenth century,

the theory of impressed force was defended, first by the Franciscan theologian Franciscus de Marchia (fl. 1320), subsequently by John Buridan (ca. 1295–ca. 1358) and others. Let us examine Buridan's version of the theory, often considered its most advanced form.

54 Buridan employed a new term, "impetus," to denote this *internal* impressed motive force—terminology that remained standard down to the time of Galileo. Buridan described impetus as an internal quality whose nature it is to move the body in which it is impressed, and took pains to distinguish this quality from the motion it produces: "Impetus is a thing of permanent nature distinct from the local motion with which the projectile is moved . . . . And it is probable that impetus is a quality naturally present and predisposed for moving a body in which it is impressed." In defense of his impetus theory, Buridan pointed to the analogous case of a magnet, which is able to impress in iron a quality capable of moving that iron toward the magnet. Like any quality, impetus is corrupted by the presence of opposition or resistance, but otherwise retains its original strength. Buridan took a first step toward quantifying impetus by declaring its strength to be measured by the velocity and the quantity of matter of the body in which it inheres. He also extended the explanatory range of the impetus theory beyond simple projectile motion, arguing that motion in the heavens might plausibly be explained by God's imposition of an impetus on the celestial spheres at the moment of creation; because the heavens offer no resistance, this impetus would not be corrupted, and the celestial spheres would be moved (as observation reveals they are) with an eternally unchanging motion. Finally, he explained the acceleration of a falling body by the assumption that as the body falls its weight continually generates additional impetus in the body; as the impetus increases, so does the velocity of the falling body.

55 The theory of impetus became the dominant explanation of projectile motion until the seventeenth century, when a new theory of motion, which denied that force (either internal or external) is required for the continuation of unresisted motion, gradually won acceptance. There have been many attempts to view the theory of impetus as an important step in the direction of modern dynamics; for example, attention has often been called to the quantitative resemblance between Buridan's impetus (velocity × quantity of matter) and the modern concept of momentum (velocity × mass). No doubt there are connections, but we must note that Buridan's impetus was the *cause* of the continuation of projectile motion, whereas momentum

is the *measure* of a motion that requires no cause for its continuation so long as no resistance is encountered. In short, Buridan was still working within a conceptual framework that was fundamentally Aristotelian; and this meant that he was a world (or worldview) away from those natural philosophers in the seventeenth century who formulated a new mechanics on the basis of a new conception of motion and inertia.

[. . .]

Text 3a

from

The Birth of a New Physics

by I. Bernard Cohen

CHAPTER 7
THE GRAND DESIGN — A NEW PHYSICS

1 The publication of Isaac Newton's *Principia* in 1687 was one of the most notable events in the whole history of physical science. In it one may find the culmination of thousands of years of striving to comprehend the system of the world, the principles of force and of motion, and the physics of bodies moving in different media. It is no small testimony to the vitality of Newton's scientific genius that although the physics of the *Principia* has been altered, improved, and challenged ever since, we still set about solving most problems of celestial mechanics and the physics of gross bodies by proceeding essentially as Newton did some 300 years ago. Newtonian principles of celestial mechanics guide our artificial satellites, our space shuttles, and every spacecraft we launch to explore the vast reaches of our solar system. And if this is not enough to satisfy the canons of greatness, Newton was equally great as a pure mathematician. He invented the differential and integral calculus (produced simultaneously and independently by the German philosopher Gottfried Wilhelm Leibniz), which is the language of physics; he developed the binomial theorem and various properties of infinite series; and he laid the foundations for the calculus

of variations. In optics, Newton began the experimental study of the analysis and composition of light, showing that white light is a mixture of light of many colors, each having a characteristic index of refraction. Upon these researches have risen the science of spectroscopy and the methods of color analysis. Newton invented a reflecting telescope and so showed astronomers how to transcend the limitations of telescopes built of lenses. All in all, his was a fantastic scientific achievement—of a kind that has never been equaled and may never be equaled again.

2 In this book we shall deal exclusively with Newton's system of dynamics and gravitation, the central problems for which the preceding chapters have been a preparation. If you have read them carefully, you have in mind all but one of the major ingredients requisite to an understanding of the Newtonian system of the world. But even if that one were to be given—the analysis of uniform circular motion—the guiding hand of Newton would still be required to put the ingredients together. It took genius to supply the new concept of universal gravitation. Let us see what Newton actually did.

3 First of all, it must be understood that Galileo himself never attempted to display any scheme of forces that would account for the movement of the planets, or of their satellites. As for Copernicus, the *De revolutionibus* contains no important insight into a celestial mechanics. Kepler had tried to supply a celestial mechanism, but the result was never a very happy one. He held that the *anima motrix* emanating from the sun would cause planets to revolve about the sun in circles. He further supposed that magnetic interactions of the sun and a planet would shift the planet during an otherwise circular revolution into an elliptical orbit. Others who contemplated the problems of planetary motion proposed systems of mechanics containing certain features that were later to appear in Newtonian dynamics. One of these was Robert Hooke, who quite understandably thought that Newton should have given him more credit than a mere passing reference for having anticipated parts of the laws of dynamics and gravitation.

Newtonian Anticipations

4 The climactic chapter in the discovery of the mechanics of the universe starts with a pretty story. By the third quarter of the seventeenth century, a group of men had become so eager to advance the new mathematical experimental sciences that

they banded together to perform experiments in concert, to present problems for solution to one another, and to report on their own researches and on those of others as revealed by correspondence, books, and pamphlets. Thus it came about that Robert Hooke, Edmond Halley, and Sir Christopher Wren, England's foremost architect, met to discuss the question, Under what law of force would a planet follow an elliptical orbit? From Kepler's laws—especially the third or harmonic law, but also the second or law of areas—it was clear that the sun somehow or other must control or at least affect the motion of a planet in accordance with the relative proximity of the planet to the sun. Even if the particular mechanisms proposed by Kepler (an *anima motrix* and a magnetic force) had to be rejected, there could be no doubt that some kind of planet-sun interaction keeps the planets in their courses. Furthermore, a more acute intuition than Kepler's would sense that any force emanating from the sun must spread out in all directions from that body, presumably diminishing according to the inverse of the square of its distance from the sun—as the intensity of light diminishes in relation to distance. But to say this much is a very different thing from *proving* it mathematically. For to prove it would require a complete physics with mathematical methods for solving all the attendant and consequent problems. When Newton declined to credit authors who tossed off general statements without being able to prove them mathematically or fit them into a valid framework of dynamics, he was quite justified in saying, as he did of Hooke's claims: "Now is not this very fine? Mathematicians that find out, settle, and do all the business must content themselves with being nothing but dry calculators and drudges; and another, that does nothing but pretend and grasp at all things, must carry away all the invention, as well of those that were to follow him as of those that went before." (See, further, Supplement 11).

In any event, by January 1684 Halley had concluded that the force acting on 5
planets to keep them in their orbits "decreased in the proportion of the squares of the distances reciprocally,"

$$F \propto \frac{1}{D^2}\ ,$$

but he was not able to deduce from that hypothesis the observed motions of the celestial bodies. When Wren and Hooke met later in the month, they agreed with Halley's supposition of a solar force. Hooke boasted "that upon that principle all the laws of the celestial motions were to be [i.e., could be] demonstrated, and that he himself had done it." But despite repeated urgings and Wren's offer of a considerable monetary prize, Hooke did not—and presumably could not—produce a solution. Six months later, in August 1684, Halley decided to go to Cambridge to consult Isaac Newton. On his arrival he learned the "good news" that Newton "had brought this demonstration to perfection." Here is DeMoivre's almost contemporaneous account of that visit:

> After they had been some time together, the Dr. [Halley] asked him what he thought the curve would be that would be described by the planets supposing the force of attraction towards the sun to be reciprocal to the square of their distance from it. Sir Isaac replied immediately that it would be an ellipsis. The Doctor, struck with joy and amazement, asked him how he knew it. Why, saith he, I have calculated it. Whereupon Dr. Halley asked him for his calculation without any further delay. Sir Isaac looked among his papers but could not find it, but he promised him to renew it and then to send it him. Sir Isaac, in order to make good his promise, fell to work again, but he could not come to that conclusion which he thought he had before examined with care. However, he attempted a new way which, though longer than the first, brought him again to his former conclusion. Then he examined carefully what might be the reason why the calculation he had undertaken before did not prove right, and he found that, having drawn an ellipsis coarsely with his own hand, he had drawn the two axes of the curve, instead of drawing two diameters somewhat inclined to one another, whereby he might have fixed his imagination to any two conjugate diameters, which was requisite he should do. That being perceived, he made both his calculations agree together.

6 Spurred on by Halley's visit, Newton resumed work on a subject that had commanded his attention in his twenties when he had laid the foundations of his other great scientific discoveries: the nature of white light and color and the differential and integral calculus. He now put his investigations in order, made great progress, and in the fall term of the year, discussed his research in a series of lectures on dynamics that he gave at Cambridge University, as required by his

professorship. Eventually, with Halley's encouragement, a draft of these lectures, *De motu corporum*, grew into one of the greatest and most influential books any man has yet conceived. Many a scientist has echoed the sentiment that Halley expressed in the ode he wrote as a preface to Newton's *Principia* (or, to give Newton's masterpiece its full title, *Philosophiae naturalis principia mathematica, Mathematical Principles of Natural Philosophy*, London, 1687):

> *Then ye who now on heavenly nectar fare,*
> *Come celebrate with me in song the name*
> *Of Newton, to the Muses dear; for he*
> *Unlocked the hidden treasuries of Truth:*
> *So richly through his mind had Phoebus cast*
> *The radiance of his own divinity.*
> *Nearer the gods no mortal may approach.*

The *Principia*

The *Principia* is divided into three parts or books; we shall concentrate on the first 7
and third. In Book One Newton develops the general principles of the dynamics of moving bodies, and in Book Three he applies the principles to the mechanism of the universe. Book Two deals with a facet of fluid mechanics, the theory of waves, and other aspects of physics.

In Book One, following the preface, a set of definitions, and a discussion of 8
the nature of time and space, Newton presented the "axioms, or laws of motion":

Law I

> Every body perseveres in its state of being at rest or of moving uniformly straight forward, except insofar as it is compelled to change its state by forces impressed upon it.

Law II

> A change in motion is proportional to the motive force impressed and takes place in the direction of the straight line along which that force is impressed. [See Suppl. Note on p. 184.]

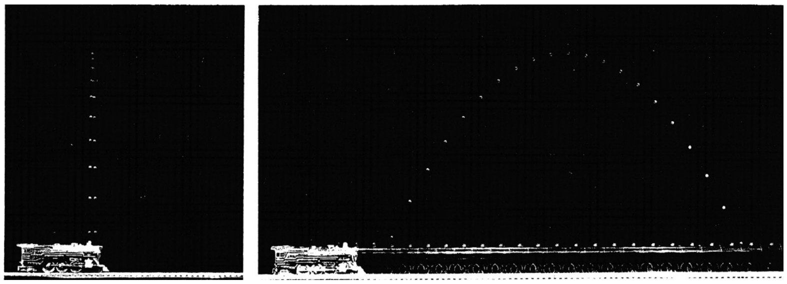


PLATE VI. A ball fired with a spring gun in the smokestack of a moving toy locomotive describes a parabola and lands on the locomotive instead of going straight up and down as it does when the locomotive is standing still. These stroboscopic pictures, with exposures at intervals of one-thirtieth of a second, vividly illustrate one of Galileo's arguments on the behavior of falling bodies and settle the ancient debate about bodies dropped from the masts of moving ships. If the speed of the locomotive were absolutely uniform and if the ball met no air resistance, it would land in the smokestack. (In fact, even under the imperfect conditions of the experiment, the ball hits the smokestack more often than not.) Note that the ball attains the same height whether the locomotive is at rest or moving. Notice, too, that in the picture where the locomotive is standing still, the distances traveled by the ball in the intervals of exposure correspond almost exactly. On the ascent, gravity slows it down; on the descent, gravity speeds it up. Photographs by Berenice Abbott.

9 Observe that if a body is in uniform motion in a straight line, a force at right angles to the direction of motion of the body will not affect the forward motion. This follows from the fact that the acceleration is always in the same direction as the force producing it, so that the acceleration in this case is at right angles to the direction of motion. Thus in the toy train experiment of Chapter 5, the chief force acting is the downward force of gravity, producing a vertical acceleration. The ball, whether moving forward or at rest, is thus caused to slow down in its upward motion until it comes to rest, and then be speeded up or accelerated on the way down.

10 A comparison of the two sets of photographs (p. 83) shows that the upward and downward motions are exactly the same whether the train is at rest or in uniform motion. In the forward direction there is no effect of weight or gravity, since this acts only in a downward direction. The only force in the forward or horizontal direction is the small amount of air friction, which is almost negligible;

so one may say that in the horizontal direction there is no force acting. According to Newton's first law of motion, the ball will continue to move in the forward direction with uniform motion in a straight line just as the train does— a fact you can check by inspecting the photograph. The ball remains above the locomotive whether the train is at rest or in uniform motion in a straight line. This law of motion is sometimes called the *principle of inertia*, and the property that material bodies have of continuing in a state of rest or of uniform motion in a straight line is sometimes known as the bodies' *inertia*.[1]

Newton illustrated Law I by reference to projectiles that continue in their 11
forward motions "so far as they are not retarded by the resistance of the air, or impelled downward by the force of gravity," and he referred also to "the greater bodies of planets and comets." (On the inertial aspect of the motion of "greater bodies" such as "planets and comets," see Supplement 12.) At this one stroke Newton postulated the opposite view of Aristotelian physics. In the latter, no celestial body could move uniformly in a straight line in the absence of a force, because this would be a "violent" motion and so contrary to its nature. Nor could a terrestrial object, as we have seen, move along its "natural" straight line without an external mover or an internal motive force. Newton, presenting a physics that applies simultaneously to both terrestrial and celestial objects, stated that in the absence of a force bodies do not necessarily stand still or come to rest as Aristotle supposed, but they may move at constant rectilinear speed. This "indifference" of all sorts of bodies to rest or uniform straight-line motion in the absence of a force clearly is an advanced form of Galileo's statement in his book on sunspots (p. 88), the difference being that in that work Galileo was writing about uniform motion along a great spherical surface concentric with the earth.

Newton said of the laws of motion that they were "such principles as 12
have been received by mathematicians, and . . . confirmed by [an] abundance of

1 The earliest known statement of this law was made by René Descartes in a book that he did not publish. It appeared in print for the first time in a work by Pierre Gassendi. But prior to Newton's *Principia* there was no completely developed inertial physics. It is not without significance that this early book of Descartes was based on the Copernican point of view; Descartes suppressed it on learning of the condemnation of Galileo. Gassendi likewise was a Copernican. He actually made experiments with objects let fall from moving ships and moving carriages to test Galileo's conclusions about inertial motion. Descartes first published his version of the law of inertia in his *Principles of Philosophy* (1644); the earlier statement, in Descartes's *The World*, was published after Descartes's death in 1650. See Suppl. 8.

experiments. By the first two Laws and the first two Corollaries, Galileo discovered that the descent of bodies varies as the square of the time and that the motion of projectiles is in the curve of a parabola, experience agreeing with both, unless so far as these motions are a little retarded by the resistance of the air." The "two Corollaries" deal with methods used by Galileo and many of his predecessors to combine two different forces or two independent motions. Fifty years after the publication of Galileo's *Two New Sciences* it was difficult for Newton, who had already established an inertial physics, to conceive that Galileo could have come as close as he had to the concept of inertia without having taken full leave of circularity and having known the true principle of linear inertia.

13 Newton was being very generous to Galileo because, however it may be argued that Galileo "really did" have the law of inertia or Newton's Law I, a great stretch of the imagination is required to assign any credit to Galileo for Law II. This law has two parts. In the second half of Newton's statement of Law II, the "change in motion" produced by an "impressed" or "motive" force—whether that is a change in the speed with which a body moves or a change in the direction in which it is moving—is said to be "in the direction of the straight line along which that force is impressed." This much is certainly implied in Galileo's analysis of projectile motion because Galileo assumed that in the forward direction there is no acceleration because there is no horizontal force, except the negligible action of air friction; but in the vertical direction there is an acceleration or continual increase of downward speed, because of the downward-acting weight force. But the first part of Law II—that the change in the magnitude of the motion is related to the motive force—is something else again; only a Newton could have seen it in Galileo's studies of falling bodies. This part of the law says that if an object were to be acted on first by one force F_1 and then by some other force F_2 , the accelerations or changes in speed produced, A_1 and A_2, would be proportional to the forces, or that

$$\frac{F_1}{F_2} = \frac{A_1}{A_2}\text{, or}$$

$$\frac{F_1}{A_1} = \frac{F_2}{A_2}$$

But in analyzing falling, Galileo was dealing with a situation in which only one force acted on each body, its weight W, and the acceleration it produced was g the acceleration of a freely falling body. (For the two forms of Newton's Law II, see p. 184.)

Where Aristotle had said that a given force gives an object a certain 14
characteristic speed, Newton now said that a given force always produces in that body a definite acceleration A. To find the speed V, we must know how long a time T the force has acted, or how long the object has been accelerated, so that Galileo's law

$$V = AT$$

may be applied.

At this point let us try a thought-experiment, in which we assume we have 15
two cubes of aluminum, one just twice the volume of the other. (Incidentally, to "duplicate" a cube—or make a cube having exactly twice the volume as some given cube—is as impossible within the framework of Euclidean geometry as to trisect an angle or to square a circle.) We now subject the smaller cube to a series of forces $F_1, F_2, F_3, \ldots$ and determine the corresponding accelerations $A_1, A_2, A_3, \ldots$ In accordance with Law II, we would find that there is a certain constant value of the ratio of force to acceleration

$$\frac{F_1}{A_1} = \frac{F_2}{A_2} = \frac{F_3}{A_3} = \ldots. = m_s$$

which for this object we may call m_s. We now repeat the operations with the larger cube and find that the same set of forces F_1, F_2, F_3, … respectively produces *another* set of accelerations $a_1, a_2, a_3, \ldots$ In accordance with Newton's second law, the force-acceleration ratio is again a constant which for this object we may call m_l

$$\frac{F_1}{a_1} = \frac{F_2}{a_2} = \frac{F_3}{a_3} = \ldots. = m_l$$

For the larger object the constant proves to be just twice as large as the 16
constant obtained for the smaller one and, in general, so long as we deal with a single variety of matter like pure aluminum, *this constant* is proportional to the volume and so *is a measure of the amount of aluminum in any sample*. This particular constant

is a measure of an object's resistance to acceleration, or a measure of the tendency of that object to stay as it is—either at rest, or in motion in a straight line. For observe that m_l was twice m_s; to give both objects the same acceleration or change in motion the force required for the larger object is just twice what it must be for the smaller. The tendency of any object to continue in its state of motion (at constant speed in a straight line) or its state of rest is called its *inertia*; hence, Newton's Law I is also called the principle of inertia. The constant determined by finding the constant force-acceleration ratio for any given body may thus be called the body's inertia. But for our aluminum blocks this same constant is also a measure of the "quantity of matter" in the object, which is called its *mass*. We now make precise the condition that two objects of different material—say one of brass and the other of wood—shall have the same "quantity of matter": it is that they have the *same mass* as determined by the force-acceleration ratio, or the *same inertia*.

17 In ordinary life, we do not compare the "quantity of matter" in objects in terms of their inertias, but in terms of their weight. Newtonian physics makes it clear why we can, and through its clarification we are able to understand why at any place on the earth two unequal weights in a vacuum fall at the same rate. But we may observe that in at least one common situation we always compare the inertias of objects rather than their weights. This happens when a person hefts two objects to find which is heavier, or has the greater mass. He does not hold them out to see which pulls down more on his arm; instead, he moves them up and down to find which is easier to move. In this way he determines which has the greater resistance to a change in its state of motion in a straight line or of rest—that is, which has the greater inertia. (On Newton's concept of inertia, see Supplement 15.)

Final Formulation of the Law of Inertia

18 At one point in his *Discourses and Demonstrations Concerning Two New Sciences*, Galileo imagined a ball to be rolling along a plane and noted that "equable motion on this plane would be perpetual if the plane were of infinite extent." A plane without limit is all right for a pure mathematician, who is a Platonist in any case. But Galileo was a man who combined just such a Platonism with a concern for applications to the real world of sensory experience. In the *Two New Sciences*, Galileo was not interested only in abstractions as such, but in the analysis of real

motions on or near the earth. So we understand that having talked about a plane without limit, he does not continue with such a fancy, but asks what would happen on such a plane if it were a real earthly plane, which for him means that it is "ended, and [situated] on high." The ball, in the real world of physics, falls off the plane and begins to fall to the ground. In this case,

> the movable (which I conceive of as being endowed with heaviness), driven to the end of this plane and going on further, adds on to its previous equable and indelible motion that downward tendency which it has from its own heaviness. Thus there emerges a certain motion, compounded from equable horizontal and from naturally accelerated downward [motion], which I call "projection."

Unlike Galileo, Newton made a clear separation between the world of abstract mathematics and the world of physics, which he still called philosophy. Thus the *Principia* included both "mathematical principles" as such and those that could be applied in "natural philosophy," but Galileo's *Two New Sciences* included only those mathematical conditions exemplified in nature. For instance, Newton plainly knew that the attractive force exerted by the sun on a planet varies as the inverse-square of the distance 19

$$F \propto \frac{1}{D^2}$$

but in Book One of the *Principia* he explored the consequences not only of this particular force but of others with quite different dependence on the distance, including

$$F \propto D$$

"The System of the World"

At the beginning of Book Three, which was devoted to "The System of the World," 20
Newton explained how it differed from the preceding two, which had been dealing with "The Motion of Bodies":

> In the preceding Books I have laid down . . . principles not philosophical [pertaining to physics] but mathematical: such, namely, as we may build our reasonings upon in philosophical inquiries. These principles are laws and conditions of certain motions, and powers or forces, which chiefly have respect to philosophy; but, lest they should have appeared of themselves dry and barren, I have illustrated them here and there with some philosophical scholiums, giving an account of such things as are of a more general nature, and which philosophy seems chiefly to be founded on: such as the density and the resistance of bodies, spaces void of all bodies, and the motion of light and sounds. It remains that, from the same principles, I now demonstrate the structure of the System of the World.

21 I believe it fair to say that it was the freedom to consider problems either in a purely mathematical way or in a "philosophical" (or physical) way that enabled Newton to express the first law and to develop a complete inertial physics. After all, physics as a science may be developed in a mathematical way but it always must rest on experience—and experience never shows us pure inertial motion. Even in the limited examples of linear inertia discussed by Galileo, there was always some air friction and the motion ceased almost at once, as when a projectile strikes the ground. In the whole range of physics explored by Galileo there is no example of a physical object that has even a component of pure inertial motion for more than a very short time. It was perhaps for this reason that Galileo never framed a general law of inertia. He was too much a physicist.

22 But as a mathematician Newton could easily conceive of a body's moving along a straight line at constant speed forever. The concept "forever," which implies an infinite universe, held no terror for him. Observe that his statement of the law of inertia, that it is the natural condition for bodies to move in straight lines at a constant speed, occurs in Book One of the *Principia*, the portion said by him to be mathematical rather than physical. Now, if it is the natural condition of motion for bodies to move uniformly in straight lines, then this kind of inertial motion must characterize the planets. The planets, however, do not move in straight lines, but rather along ellipses. Using a kind of Galilean approach to this single problem, Newton could say that the planets must therefore be subject to two motions: one inertial (along a straight line at constant speed) and one always at right angles

to that straight line drawing each planet toward its orbit. (See, further, Supplements 11 and 12.)

Though not moving in a straight line, each planet nevertheless represents 23
the best example of inertial motion observable in the universe. Were it not for that component of inertial motion, the force that continually draws the planet away from the straight line would draw the planet in toward the sun until the two bodies collided. Newton once used this argument to prove the existence of God. If the planets had not received a push to give them an inertial (or tangential) component of motion, he said, the solar attractive force would not draw them into an orbit but instead would move each planet in a straight line toward the sun itself. Hence the universe could not be explained in terms of matter alone.

For Galileo pure circular motion could still be inertial, as in the example of an 24
object on or near the surface of the earth. But for Newton pure circular motion was not inertial; it was accelerated and required a force for its continuance. Thus it was Newton who finally shattered the bonds of "circularity" which still had held Galileo in thrall. And so we may understand that it was Newton who showed how to build a celestial mechanics based on the laws of motion, since the elliptical (or almost circular) orbital motion of planets is not purely inertial, but requires additionally the constant action of a force, which turns out to be the force of universal gravitation.

Thus Newton, again unlike Galileo, set out to "demonstrate the structure of 25
the System of the World," or—as we would say today—to show how the general laws of terrestrial motion may be applied to the planets and to their satellites.

[. . .]

Isaac Newton's system of mechanics came to symbolize the rational order 62
of the world, functioning under the "rule of nature." Not only could Newtonian science account for present and past phenomena; the principles could be applied to the prediction of future events. In the *Principia* Newton proved that comets are like the planets, moving in great orbits that must (according to Newtonian rules) be conic sections. Some comets move in ellipses and these must return periodically from far out in space to the visible regions of our solar system, whereas others will visit our solar system and never return. Edmond Halley applied these Newtonian results to an analysis of cometary records of the past and found—among others—

a comet with a period of some seventy-five and a half years. He made a bold Newtonian prediction that this comet would reappear in 1758. When it did so, right on schedule, though long after Halley and Newton were dead, men and women everywhere experienced a new feeling of awe for the powers of human reason abetted by mathematics. This new respect for science was expressed by such adjectives as “amazing,” “phenomenal,” or “extraordinary.” This successful prediction of a future event symbolized the force of the new science: the perfection of the mathematical understanding of nature, realized in the ability to make reliable predictions of the future. Not surprisingly, men and women everywhere saw a promise that all of human knowledge and the regulation of human affairs would yield to a similar rational system of deduction and mathematical inference coupled with experiment and critical observation. The eighteenth century not only was the Enlightenment, but became “preeminently the age of faith in science.” Newton became the symbol of successful science, the ideal for all thought—in philosophy, psychology, government, and the science of society.

63 Newton’s genius enables us to see the full significance of both Galilean mechanics and Kepler’s laws of planetary motion as manifested in the development of the inertial principles required for the Copernican-Keplerian universe. A great French mathematician, Joseph Louis Lagrange (1736–1813), best defined Newton’s achievement. There is only one law of the universe, he said, and Newton discovered it. Newton did not develop modern dynamics all by himself but depended heavily on certain of his predecessors; the debt in no way lessens the magnitude of his achievement. It only emphasizes the importance of such men as Galileo and Kepler, and Descartes, Hooke, and Huygens, who were great enough to make significant contributions to the Newtonian enterprise. Above all, we may see in Newton’s work the degree to which science is a collective and a cumulative activity and we may find in it the measure of the influence of an individual genius on the future of a cooperative scientific effort. In Newton’s achievement, we see how science advances by heroic exercises of the imagination, rather than by patient collecting and sorting of myriads of individual facts. Who, after studying Newton’s magnificent contribution to thought, could deny that pure science exemplifies the creative accomplishment of the human spirit at its pinnacle?

Text 3b

from

The Principia: Mathematical Principles of Natural Philosophy

by Isaac Newton

DEFINITIONS.

Definition 1

Quantity of matter is a measure of matter that arises from its density and volume jointly.

If the density of air is doubled in a space that is also doubled, there is four times as much air, and there is six times as much if the space is tripled. The case is the same for snow and powders condensed by compression or liquefaction, and also for all bodies that are condensed in various ways by any causes whatsoever. For the present, I am not taking into account any medium, if there should be any, freely pervading the interstices between the parts of bodies. Furthermore, I mean this quantity whenever I use the term "body" or "mass" in the following pages. It can always be known from a body's weight, for—by making very accurate experiments with pendulums—I have found it to be proportional to the weight, as will be shown below.

Definition 2

Quantity of motion is a measure of motion that arises from the velocity and the quantity of matter jointly.

The motion of a whole is the sum of the motions of the individual parts, and thus if a body is twice as large as another and has equal velocity there is twice as much motion, and if it has twice the velocity there is four times as much motion.

Definition 3

Inherent force of matter is the power of resisting by which every body, so far as it is able, perseveres in its state either of resting or of moving uniformly straight forward.

This force is always proportional to the body and does not differ in any way from the inertia of the mass except in the manner in which it is conceived. Because of the inertia of matter, every body is only with difficulty put out of its state either of resting or of moving. Consequently, inherent force may also be called by the very significant name of force of inertia. Moreover, a body exerts this force only during a change of its state, caused by another force impressed upon it, and this exercise of force is, depending on the viewpoint, both resistance and impetus: resistance insofar as the body, in order to maintain its state, strives against the impressed force, and impetus insofar as the same body, yielding only with difficulty to the force of a resisting obstacle, endeavors to change the state of that obstacle. Resistance is commonly attributed to resting bodies and impetus to moving bodies; but motion and rest, in the popular sense of the terms, are distinguished from each other only by point of view, and bodies commonly regarded as being at rest are not always truly at rest.

Definition 4

Impressed force is the action exerted on a body to change its state either of resting or of moving uniformly straight forward.

This force consists solely in the action and does not remain in a body after the action has ceased. For a body perseveres in any new state solely by the force of inertia. Moreover, there are various sources of impressed force, such as percussion, pressure, or centripetal force.

Definition 5

> ***Centripetal force is the force by which bodies are drawn from all sides, are impelled, or in any way tend, toward some point as to a center.***

One force of this kind is gravity, by which bodies tend toward the center of the earth; another is magnetic force, by which iron seeks a lodestone; and yet another is that force, whatever it may be, by which the planets are continually drawn back from rectilinear motions and compelled to revolve in curved lines.

A stone whirled in a sling endeavors to leave the hand that is whirling it, and by its endeavor it stretches the sling, doing so the more strongly the more swiftly it revolves; and as soon as it is released, it flies away. The force opposed to that endeavor, that is, the force by which the sling continually draws the stone back toward the hand and keeps it in an orbit, I call centripetal, since it is directed toward the hand as toward the center of an orbit. And the same applies to all bodies that are made to move in orbits. They all endeavor to recede from the centers of their orbits, and unless some force opposed to that endeavor is present, restraining them and keeping them in orbits and hence called by me centripetal, they will go off in straight lines with uniform motion. If a projectile were deprived of the force of gravity, it would not be deflected toward the earth but would go off in a straight line into the heavens and do so with uniform motion, provided that the resistance of the air were removed. The projectile, by its gravity, is drawn back from a rectilinear course and continually deflected toward the earth, and this is so to a greater or lesser degree in proportion to its gravity and its velocity of motion. The less its gravity in proportion to its quantity of matter, or the greater the velocity with which it is projected, the less it will deviate from a rectilinear course and the farther it will go. If a lead ball were projected with a given velocity along a horizontal line from the top of some mountain by the force of gunpowder and went in a curved line for a distance of two miles before falling to the earth, then the same ball projected with twice the velocity would go about twice as far and with ten times the velocity about ten times as far, provided that the resistance of the air were removed. And by increasing the velocity, the distance to which it would be projected could be increased at will and the curvature of the line that it would describe could be decreased, in such a way that it would finally fall at a distance of 10 or 30 or 90 degrees or even go around the whole earth or, lastly, go off into the heavens and continue indefinitely in this motion. And in the same way that a projectile could, by the force of gravity, be deflected into an orbit and go around the whole earth, so too the moon, whether by the force of gravity—if it has gravity—or by any other force by which it may be urged toward the earth, can always be drawn back toward the earth from a rectilinear course and deflected into its orbit; and without such a force the moon cannot be kept in its orbit. If this force were too small, it would not deflect the moon sufficiently from a rectilinear course; if it were too great, it would deflect the moon excessively and draw it down from its orbit toward the earth. In fact, it must be of just the right magnitude, and mathematicians have the task of finding the force by which a body can be kept exactly in any given orbit with a given velocity and, alternatively, to find the curvilinear path into which a body leaving any given place with a given velocity is deflected by a given force.

The quantity of centripetal force is of three kinds: absolute, accelerative, and motive.

[. . .]

Further, it is in this same sense that I call attractions and impulses accelerative and motive. Moreover, I use interchangeably and indiscriminately words signifying attraction, impulse, or any sort of propensity toward a center, considering these forces not from a physical but only from a mathematical point of view. Therefore, let the reader beware of thinking that by words of this kind I am anywhere defining a species or mode of action or a physical cause or reason, or that I am attributing forces in a true and physical sense to centers (which are mathematical points) if I happen to say that centers attract or that centers have forces.

AXIOMS, OR THE LAWS OF MOTION

Law 1

Every body perseveres in its state of being at rest or of moving uniformly straight forward, except insofar as it is compelled to change its state by forces impressed.

Projectiles persevere in their motions, except insofar as they are retarded by the resistance of the air and are impelled downward by the force of gravity. A spinning hoop, which has parts that by their cohesion continually draw one another back from rectilinear motions, does not cease to rotate, except insofar as it is retarded by the air. And larger bodies—planets and comets—preserve for a longer time both their progressive and their circular motions, which take place in spaces having less resistance.

Law 2

A change in motion is proportional to the motive force impressed and takes place along the straight line in which that force is impressed.

If some force generates any motion, twice the force will generate twice the motion, and three times the force will generate three times the motion, whether the force is impressed all at once or successively by degrees. And if the body was previously moving, the new motion (since motion is always in the same direction as the generative force) is added to the original motion if that motion was in the same direction or is subtracted from the original motion if it was in the opposite direction or, if it was in an oblique direction, is combined obliquely and compounded with it according to the directions of both motions.

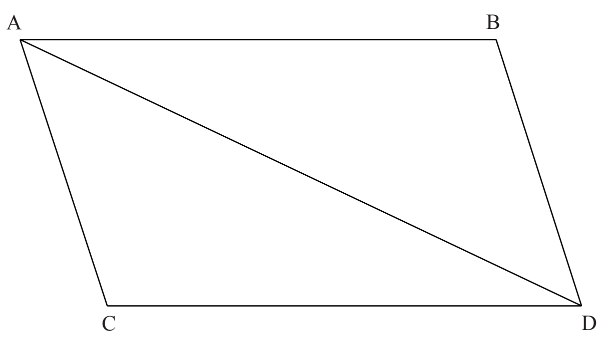
Law 3

To any action there is always an opposite and equal reaction; in other words, the actions of two bodies upon each other are always equal and always opposite in direction.

Whatever presses or draws something else is pressed or drawn just as much by it. If anyone presses a stone with a finger, the finger is also pressed by the stone. If a horse draws a stone tied to a rope, the horse will (so to speak) also be drawn back equally toward the stone, for the rope, stretched out at both ends, will urge the horse toward the stone and the stone toward the horse by one and the same endeavor to go slack and will impede the forward motion of the one as much as it promotes the forward motion of the other. If some body impinging upon another body changes the motion of that body in any way by its own force, then, by the force of the other body (because of the equality of their mutual pressure), it also will in turn undergo the same change in its own motion in the opposite direction. By means of these actions, equal changes occur in the motions, not in the velocities—that is, of course, if the bodies are not impeded by anything else. For the changes in velocities that likewise occur in opposite directions are inversely proportional to the bodies because the motions are changed equally. This law is valid also for attractions, as will be proved in the next scholium.

Corollary 1

A body acted on by [two] forces acting jointly describes the diagonal of a parallelogram in the same time in which it would describe the sides if the forces were acting separately.



Let a body in a given time, by force M alone impressed in A, be carried with uniform motion from A to B, and, by force N alone impressed in the same place, be carried from A to C; then complete the parallelogram ABDC, and by both forces the body will be carried in the same time along the diagonal from A to D. For, since force N acts along the line AC parallel to BD, this force, by law 2, will make no change at all in the velocity toward the line BD which is generated by the other force. Therefore, the body will reach the line BD in the same time whether force N is impressed or not, and so at the end of that time will be found somewhere on the line BD. By the same argument, at the end of the same time it will be found somewhere on the line CD, and accordingly it is necessarily found at the intersection D of both lines. And, by law 1, it will go with [uniform] rectilinear motion from A to D.

[. . .]

Part II
Human Exploration of the World of Life

Our endeavour to understand the world of life achieved a breakthrough with Charles Darwin's mid-19th century attempt to explain the diversity of life, *On the Origin of Species*. The book is often referred to but seldom read. In Text 4, Darwin suggested that natural selection and emergence of varieties are responsible for the origin of species but left the question of the biological mechanism of emergence of varieties unanswered.

About a century later, in 1953, James Watson, Francis Crick and Maurice Wilkins discovered the structure of DNA. For this achievement, they received the Nobel Prize in Physiology and Medicine. We now know that some chemical processes taking place in the DNA are responsible for the emergence of varieties. The excerpt (Text 5) from Watson's *DNA: The Secret of Life* is the story of the discovery of the structure of DNA and also the author's reflections.

Our life sciences have now advanced to the point where we can routinely manipulate life. Some of these manipulations have been good and others have had unforeseen harmful consequences. Methods for curing diseases have usually been regarded as good, for example, while using chemicals to control agricultural pests has often proved damaging to the environment. Text 6 from Rachel Carson's *Silent Spring* rebukes policy-makers for using herbicides to kill unwanted plants.

from

On the Origin of Species

by Charles Darwin

CHAPTER IV
NATURAL SELECTION

Natural Selection—its power compared with man's selection—its power on characters of trifling importance—its power at all ages and on both sexes—Sexual Selection—On the generality of intercrosses between individuals of the same species—Circumstances favourable and unfavourable to Natural Selection, namely, intercrossing, isolation, number of individuals—Slow action—Extinction caused by Natural Selection—Divergence of Character, related to the diversity of inhabitants of any small area, and to naturalisation—Action of Natural Selection, through Divergence of Character and Extinction, on the descendants from a common parent—Explains the Grouping of all organic beings.

How will the struggle for existence, discussed too briefly in the last chapter, act 1
in regard to variation? Can the principle of selection, which we have seen is so potent in the hands of man, apply in nature? I think we shall see that it can act most effectually. Let it be borne in mind in what an endless number of strange peculiarities our domestic productions, and, in a lesser degree, those under nature, vary; and

From Charles Darwin, *On the Origin of Species by Means of Natural Selection* (First Edition) (1859). Reproduced from the public domain.

how strong the hereditary tendency is. Under domestication, it may be truly said that the whole organisation becomes in some degree plastic. Let it be borne in mind how infinitely complex and close-fitting are the mutual relations of all organic beings to each other and to their physical conditions of life. Can it, then, be thought improbable, seeing that variations useful to man have undoubtedly occurred, that other variations useful in some way to each being in the great and complex battle of life, should sometimes occur in the course of thousands of generations? If such do occur, can we doubt (remembering that many more individuals are born than can possibly survive) that individuals having any advantage, however slight, over others, would have the best chance of surviving and of procreating their kind? On the other hand, we may feel sure that any variation in the least degree injurious would be rigidly destroyed. This preservation of favourable variations and the rejection of injurious variations, I call Natural Selection. Variations neither useful nor injurious would not be affected by natural selection, and would be left a fluctuating element, as perhaps we see in the species called polymorphic.

2 We shall best understand the probable course of natural selection by taking the case of a country undergoing some physical change, for instance, of climate. The proportional numbers of its inhabitants would almost immediately undergo a change, and some species might become extinct. We may conclude, from what we have seen of the intimate and complex manner in which the inhabitants of each country are bound together, that any change in the numerical proportions of some of the inhabitants, independently of the change of climate itself, would most seriously affect many of the others. If the country were open on its borders, new forms would certainly immigrate, and this also would seriously disturb the relations of some of the former inhabitants. Let it be remembered how powerful the influence of a single introduced tree or mammal has been shown to be. But in the case of an island, or of a country partly surrounded by barriers, into which new and better adapted forms could not freely enter, we should then have places in the economy of nature which would assuredly be better filled up, if some of the original inhabitants were in some manner modified; for, had the area been open to immigration, these same places would have been seized on by intruders. In such case, every slight modification, which in the course of ages chanced to arise, and which in any way favoured the individuals of any of the species, by better adapting them to their altered conditions, would tend to be preserved; and natural selection would thus have free scope for the work of improvement.

We have reason to believe, as stated in the first chapter, that a change in the conditions of life, by specially acting on the reproductive system, causes or increases variability; and in the foregoing case the conditions of life are supposed to have undergone a change, and this would manifestly be favourable to natural selection, by giving a better chance of profitable variations occurring; and unless profitable variations do occur, natural selection can do nothing. Not that, as I believe, any extreme amount of variability is necessary; as man can certainly produce great results by adding up in any given direction mere individual differences, so could Nature, but far more easily, from having incomparably longer time at her disposal. Nor do I believe that any great physical change, as of climate, or any unusual degree of isolation to check immigration, is actually necessary to produce new and unoccupied places for natural selection to fill up by modifying and improving some of the varying inhabitants. For as all the inhabitants of each country are struggling together with nicely balanced forces, extremely slight modifications in the structure or habits of one inhabitant would often give it an advantage over others; and still further modifications of the same kind would often still further increase the advantage. No country can be named in which all the native inhabitants are now so perfectly adapted to each other and to the physical conditions under which they live, that none of them could anyhow be improved; for in all countries, the natives have been so far conquered by naturalised productions, that they have allowed foreigners to take firm possession of the land. And as foreigners have thus everywhere beaten some of the natives, we may safely conclude that the natives might have been modified with advantage, so as to have better resisted such intruders. 3

As man can produce and certainly has produced a great result by his methodical and unconscious means of selection, what may not nature effect? Man can act only on external and visible characters: nature cares nothing for appearances, except in so far as they may be useful to any being. She can act on every internal organ, on every shade of constitutional difference, on the whole machinery of life. Man selects only for his own good; Nature only for that of the being which she tends. Every selected character is fully exercised by her; and the being is placed under well-suited conditions of life. Man keeps the natives of many climates in the same country; he seldom exercises each selected character in some peculiar and fitting manner; he feeds a long and a short beaked pigeon on the same food; he does not exercise a long-backed or long-legged quadruped in any peculiar manner; he exposes sheep with long and short wool to the same climate. He does not allow the 4 most vigorous males to struggle for the females. He does not rigidly destroy all inferior animals, but protects during each varying season, as far as lies in his power, all his productions. He often begins his selection by some half-monstrous form; or at least by some modification prominent enough to catch his eye, or to be plainly useful to him. Under nature, the slightest difference of structure or constitution may well turn the nicely-balanced scale in the struggle for life, and so be preserved. How fleeting are the wishes and efforts of man! how short his time! and consequently how poor will his products be, compared with those accumulated by nature during whole geological periods. Can we wonder, then, that nature's productions should be far "truer" in character than man's productions; that they should be infinitely better adapted to the most complex conditions of life, and should plainly bear the stamp of far higher workmanship?

5 It may be said that natural selection is daily and hourly scrutinising, throughout the world, every variation, even the slightest; rejecting that which is bad, preserving and adding up all that is good; silently and insensibly working, whenever and wherever opportunity offers, at the improvement of each organic being in relation to its organic and inorganic conditions of life. We see nothing of these slow changes in progress, until the hand of time has marked the long lapse of ages, and then so imperfect is our view into long past geological ages, that we only see that the forms of life are now different from what they formerly were.

6 Although natural selection can act only through and for the good of each being, yet characters and structures, which we are apt to consider as of very trifling importance, may thus be acted on. When we see leaf-eating insects green, and bark-feeders mottled-grey; the alpine ptarmigan white in winter, the red-grouse the colour of heather, and the black-grouse that of peaty earth, we must believe that these tints are of service to these birds and insects in preserving them from danger. Grouse, if not destroyed at some period of their lives, would increase in countless numbers; they are known to suffer largely from birds of prey; and hawks are guided by eyesight to their prey,—so much so, that on parts of the Continent persons are warned not to keep white pigeons, as being the most liable to destruction. Hence I can see no reason to doubt that natural selection might be most effective in giving the proper colour to each kind of grouse, and in keeping that colour, when once acquired, true and constant. Nor ought we to think that the occasional destruction of an animal of any particular colour would produce little effect: we should remember

how essential it is in a flock of white sheep to destroy every lamb with the faintest trace of black. In plants the down on the fruit and the colour of the flesh are considered by botanists as characters of the most trifling importance: yet we hear from an excellent horticulturist, Downing, that in the United States smooth-skinned fruits suffer far more from a beetle, a curculio, than those with down; that purple plums suffer far more from a certain disease than yellow plums; whereas another disease attacks yellow-fleshed peaches far more than those with other coloured flesh. If, with all the aids of art, these slight differences make a great difference in cultivating the several varieties, assuredly, in a state of nature, where the trees would have to struggle with other trees and with a host of enemies, such differences would effectually settle which variety, whether a smooth or downy, a yellow or purple fleshed fruit, should succeed.

[. . .]

Natural selection will modify the structure of the young in relation to the 9
parent, and of the parent in relation to the young. In social animals it will adapt the structure of each individual for the benefit of the community; if each in consequence profits by the selected change. What natural selection cannot do, is to modify the structure of one species, without giving it any advantage, for the good of another species; and though statements to this effect may be found in works of natural history, I cannot find one case which will bear investigation. A structure used only once in an animal's whole life, if of high importance to it, might be modified to any extent by natural selection; for instance, the great jaws possessed by certain insects, and used exclusively for opening the cocoon—or the hard tip to the beak of nestling birds, used for breaking the egg. It has been asserted, that of the best short-beaked tumbler-pigeons more perish in the egg than are able to get out of it; so that fanciers assist in the act of hatching. Now, if nature had to make the beak of a full-grown pigeon very short for the bird's own advantage, the process of modification would be very slow, and there would be simultaneously the most rigorous selection of the young birds within the egg, which had the most powerful and hardest beaks, for all with weak beaks would inevitably perish: or, more delicate and more easily broken shells might be selected, the thickness of the shell being known to vary like every other structure.

10 *Sexual Selection.*—Inasmuch as peculiarities often appear under domestication in one sex and become hereditarily attached to that sex, the same fact probably occurs under nature, and if so, natural selection will be able to modify one sex in its functional relations to the other sex, or in relation to wholly different habits of life in the two sexes, as is sometimes the case with insects. And this leads me to say a few words on what I call Sexual Selection. This depends, not on a struggle for existence, but on a struggle between the males for possession of the females; the result is not death to the unsuccessful competitor, but few or no offspring. Sexual selection is, therefore, less rigorous than natural selection. Generally, the most vigorous males, those which are best fitted for their places in nature, will leave most progeny. But in many cases, victory will depend not on general vigour, but on having special weapons, confined to the male sex. A hornless stag or spurless cock would have a poor chance of leaving offspring. Sexual selection by always allowing the victor to breed might surely give indomitable courage, length to the spur, and strength to the wing to strike in the spurred leg, as well as the brutal cock-fighter, who knows well that he can improve his breed by careful selection of the best cocks. How low in the scale of nature this law of battle descends, I know not; male alligators have been described as fighting, bellowing, and whirling round, like Indians in a war-dance, for the possession of the females; male salmons have been seen fighting all day long; male stag-beetles often bear wounds from the huge mandibles of other males. The war is, perhaps, severest between the males of polygamous animals, and these seem oftenest provided with special weapons. The males of carnivorous animals are already well armed; though to them and to others, special means of defence may be given through means of sexual selection, as the mane to the lion, the shoulder-pad to the boar, and the hooked jaw to the male salmon; for the shield may be as important for victory, as the sword or spear.

11 Amongst birds, the contest is often of a more peaceful character. All those who have attended to the subject, believe that there is the severest rivalry between the males of many species to attract by singing the females. The rock-thrush of Guiana, birds of Paradise, and some others, congregate; and successive males display their gorgeous plumage and perform strange antics before the females, which standing by as spectators, at last choose the most attractive partner. Those who have closely attended to birds in confinement well know that they often take individual preferences and dislikes: thus Sir R. Heron has described how one pied

peacock was eminently attractive to all his hen birds. It may appear childish to attribute any effect to such apparently weak means: I cannot here enter on the details necessary to support this view; but if man can in a short time give elegant carriage and beauty to his bantams, according to his standard of beauty, I can see no good reason to doubt that female birds, by selecting, during thousands of generations, the most melodious or beautiful males, according to their standard of beauty, might produce a marked effect. I strongly suspect that some well-known laws with respect to the plumage of male and female birds, in comparison with the plumage of the young, can be explained on the view of plumage having been chiefly modified by sexual selection, acting when the birds have come to the breeding age or during the breeding season; the modifications thus produced being inherited at corresponding ages or seasons, either by the males alone, or by the males and females; but I have not space here to enter on this subject.

Thus it is, as I believe, that when the males and females of any animal have 12
the same general habits of life, but differ in structure, colour, or ornament, such differences have been mainly caused by sexual selection; that is, individual males have had, in successive generations, some slight advantage over other males, in their weapons, means of defence, or charms; and have transmitted these advantages to their male offspring. Yet, I would not wish to attribute all such sexual differences to this agency: for we see peculiarities arising and becoming attached to the male sex in our domestic animals (as the wattle in male carriers, horn-like protuberances in the cocks of certain fowls, &c.), which we cannot believe to be either useful to the males in battle, or attractive to the females. We see analogous cases under nature, for instance, the tuft of hair on the breast of the turkey-cock, which can hardly be either useful or ornamental to this bird;—indeed, had the tuft appeared under domestication, it would have been called a monstrosity.

Illustrations of the action of Natural Selection.—In order to make it clear 13
how, as I believe, natural selection acts, I must beg permission to give one or two imaginary illustrations. Let us take the case of a wolf, which preys on various animals, securing some by craft, some by strength, and some by fleetness; and let us suppose that the fleetest prey, a deer for instance, had from any change in the country increased in numbers, or that other prey had decreased in numbers, during that season of the year when the wolf is hardest pressed for food. I can under such circumstances see no reason to doubt that the swiftest and slimmest

wolves would have the best chance of surviving, and so be preserved or selected,—provided always that they retained strength to master their prey at this or at some other period of the year, when they might be compelled to prey on other animals. I can see no more reason to doubt this, than that man can improve the fleetness of his greyhounds by careful and methodical selection, or by that unconscious selection which results from each man trying to keep the best dogs without any thought of modifying the breed.

14 Even without any change in the proportional numbers of the animals on which our wolf preyed, a cub might be born with an innate tendency to pursue certain kinds of prey. Nor can this be thought very improbable; for we often observe great differences in the natural tendencies of our domestic animals; one cat, for instance, taking to catch rats, another mice; one cat, according to Mr. St. John, bringing home winged game, another hares or rabbits, and another hunting on marshy ground and almost nightly catching woodcocks or snipes. The tendency to catch rats rather than mice is known to be inherited. Now, if any slight innate change of habit or of structure benefited an individual wolf, it would have the best chance of surviving and of leaving offspring. Some of its young would probably inherit the same habits or structure, and by the repetition of this process, a new variety might be formed which would either supplant or coexist with the parent-form of wolf. Or, again, the wolves inhabiting a mountainous district, and those frequenting the lowlands, would naturally be forced to hunt different prey; and from the continued preservation of the individuals best fitted for the two sites, two varieties might slowly be formed. These varieties would cross and blend where they met; but to this subject of intercrossing we shall soon have to return. I may add, that, according to Mr. Pierce, there are two varieties of the wolf inhabiting the Catskill Mountains in the United States, one with a light greyhound-like form, which pursues deer, and the other more bulky, with shorter legs, which more frequently attacks the shepherd's flocks.

15 Let us now take a more complex case. Certain plants excrete a sweet juice, apparently for the sake of eliminating something injurious from their sap: this is effected by glands at the base of the stipules in some Leguminosae, and at the back of the leaf of the common laurel. This juice, though small in quantity, is greedily sought by insects. Let us now suppose a little sweet juice or nectar to be excreted by the inner bases of the petals of a flower. In this case insects in seeking

the nectar would get dusted with pollen, and would certainly often transport the pollen from one flower to the stigma of another flower. The flowers of two distinct individuals of the same species would thus get crossed; and the act of crossing, we have good reason to believe (as will hereafter be more fully alluded to), would produce very vigorous seedlings, which consequently would have the best chance of flourishing and surviving. Some of these seedlings would probably inherit the nectar-excreting power. Those individual flowers which had the largest glands or nectaries, and which excreted most nectar, would be oftenest visited by insects, and would be oftenest crossed; and so in the long-run would gain the upper hand. Those flowers, also, which had their stamens and pistils placed, in relation to the size and habits of the particular insects which visited them, so as to favour in any degree the transportal of their pollen from flower to flower, would likewise be favoured or selected. We might have taken the case of insects visiting flowers for the sake of collecting pollen instead of nectar; and as pollen is formed for the sole object of fertilisation, its destruction appears a simple loss to the plant; yet if a little pollen were carried, at first occasionally and then habitually, by the pollen-devouring insects from flower to flower, and a cross thus effected, although nine-tenths of the pollen were destroyed, it might still be a great gain to the plant; and those individuals which produced more and more pollen, and had larger and larger anthers, would be selected.

When our plant, by this process of the continued preservation or natural 16
selection of more and more attractive flowers, had been rendered highly attractive to insects, they would, unintentionally on their part, regularly carry pollen from flower to flower; and that they can most effectually do this, I could easily show by many striking instances. I will give only one—not as a very striking case, but as likewise illustrating one step in the separation of the sexes of plants, presently to be alluded to. Some holly-trees bear only male flowers, which have four stamens producing rather a small quantity of pollen, and a rudimentary pistil; other holly-trees bear only female flowers; these have a full-sized pistil, and four stamens with shrivelled anthers, in which not a grain of pollen can be detected. Having found a female tree exactly sixty yards from a male tree, I put the stigmas of twenty flowers, taken from different branches, under the microscope, and on all, without exception, there were pollen-grains, and on some a profusion of pollen. As the wind had set for several days from the female to the male tree, the pollen could not thus have been

carried. The weather had been cold and boisterous, and therefore not favourable to bees, nevertheless every female flower which I examined had been effectually fertilised by the bees, accidentally dusted with pollen, having flown from tree to tree in search of nectar. But to return to our imaginary case: as soon as the plant had been rendered so highly attractive to insects that pollen was regularly carried from flower to flower, another process might commence. No naturalist doubts the advantage of what has been called the "physiological division of labour;" hence we may believe that it would be advantageous to a plant to produce stamens alone in one flower or on one whole plant, and pistils alone in another flower or on another plant. In plants under culture and placed under new conditions of life, sometimes the male organs and sometimes the female organs become more or less impotent; now if we suppose this to occur in ever so slight a degree under nature, then as pollen is already carried regularly from flower to flower, and as a more complete separation of the sexes of our plant would be advantageous on the principle of the division of labour, individuals with this tendency more and more increased, would be continually favoured or selected, until at last a complete separation of the sexes would be effected.

17 Let us now turn to the nectar-feeding insects in our imaginary case: we may suppose the plant of which we have been slowly increasing the nectar by continued selection, to be a common plant; and that certain insects depended in main part on its nectar for food. I could give many facts, showing how anxious bees are to save time; for instance, their habit of cutting holes and sucking the nectar at the bases of certain flowers, which they can, with a very little more trouble, enter by the mouth. Bearing such facts in mind, I can see no reason to doubt that an accidental deviation in the size and form of the body, or in the curvature and length of the proboscis, &c., far too slight to be appreciated by us, might profit a bee or other insect, so that an individual so characterised would be able to obtain its food more quickly, and so have a better chance of living and leaving descendants. Its descendants would probably inherit a tendency to a similar slight deviation of structure. The tubes of the corollas of the common red and incarnate clovers (Trifolium pratense and incarnatum) do not on a hasty glance appear to differ in length; yet the hive-bee can easily suck the nectar out of the incarnate clover, but not out of the common red clover, which is visited by humble-bees alone; so that whole fields of the red clover offer in vain an abundant supply of precious nectar to the hive-bee. Thus it

might be a great advantage to the hive-bee to have a slightly longer or differently constructed proboscis. On the other hand, I have found by experiment that the fertility of clover greatly depends on bees visiting and moving parts of the corolla, so as to push the pollen on to the stigmatic surface. Hence, again, if humble-bees were to become rare in any country, it might be a great advantage to the red clover to have a shorter or more deeply divided tube to its corolla, so that the hive-bee could visit its flowers. Thus I can understand how a flower and a bee might slowly become, either simultaneously or one after the other, modified and adapted in the most perfect manner to each other, by the continued preservation of individuals presenting mutual and slightly favourable deviations of structure.

I am well aware that this doctrine of natural selection, exemplified in the 18
above imaginary instances, is open to the same objections which were at first urged against Sir Charles Lyell's noble views on "the modern changes of the earth, as illustrative of geology;" but we now very seldom hear the action, for instance, of the coast-waves, called a trifling and insignificant cause, when applied to the excavation of gigantic valleys or to the formation of the longest lines of inland cliffs. Natural selection can act only by the preservation and accumulation of infinitesimally small inherited modifications, each profitable to the preserved being; and as modern geology has almost banished such views as the excavation of a great valley by a single diluvial wave, so will natural selection, if it be a true principle, banish the belief of the continued creation of new organic beings, or of any great and sudden modification in their structure.

[. . .]

Extinction.—This subject will be more fully discussed in our chapter 39
on Geology; but it must be here alluded to from being intimately connected with natural selection. Natural selection acts solely through the preservation of variations in some way advantageous, which consequently endure. But as from the high geometrical powers of increase of all organic beings, each area is already fully stocked with inhabitants, it follows that as each selected and favoured form increases in number, so will the less favoured forms decrease and become rare. Rarity, as geology tells us, is the precursor to extinction. We can, also, see that any form represented by few individuals will, during fluctuations in the seasons or

in the number of its enemies, run a good chance of utter extinction. But we may go further than this; for as new forms are continually and slowly being produced, unless we believe that the number of specific forms goes on perpetually and almost indefinitely increasing, numbers inevitably must become extinct. That the number of specific forms has not indefinitely increased, geology shows us plainly; and indeed we can see reason why they should not have thus increased, for the number of places in the polity of nature is not indefinitely great,—not that we have any means of knowing that any one region has as yet got its maximum of species. Probably no region is as yet fully stocked, for at the Cape of Good Hope, where more species of plants are crowded together than in any other quarter of the world, some foreign plants have become naturalised, without causing, as far as we know, the extinction of any natives.

40 Furthermore, the species which are most numerous in individuals will have the best chance of producing within any given period favourable variations. We have evidence of this, in the facts given in the second chapter, showing that it is the common species which afford the greatest number of recorded varieties, or incipient species. Hence, rare species will be less quickly modified or improved within any given period, and they will consequently be beaten in the race for life by the modified descendants of the commoner species.

41 From these several considerations I think it inevitably follows, that as new species in the course of time are formed through natural selection, others will become rarer and rarer, and finally extinct. The forms which stand in closest competition with those undergoing modification and improvement, will naturally suffer most. And we have seen in the chapter on the Struggle for Existence that it is the most closely-allied forms,—varieties of the same species, and species of the same genus or of related genera,—which, from having nearly the same structure, constitution, and habits, generally come into the severest competition with each other. Consequently, each new variety or species, during the progress of its formation, will generally press hardest on its nearest kindred, and tend to exterminate them. We see the same process of extermination amongst our domesticated productions, through the selection of improved forms by man. Many curious instances could be given showing how quickly new breeds of cattle, sheep, and other animals, and varieties of flowers, take the place of older and inferior kinds. In Yorkshire, it is historically known that the ancient black cattle were displaced by the long-horns, and that these

"were swept away by the short-horns" (I quote the words of an agricultural writer) "as if by some murderous pestilence."

Divergence of Character.—The principle, which I have designated by this 42
term, is of high importance on my theory, and explains, as I believe, several important facts. In the first place, varieties, even strongly-marked ones, though having somewhat of the character of species—as is shown by the hopeless doubts in many cases how to rank them—yet certainly differ from each other far less than do good and distinct species. Nevertheless, according to my view, varieties are species in the process of formation, or are, as I have called them, incipient species. How, then, does the lesser difference between varieties become augmented into the greater difference between species? That this does habitually happen, we must infer from most of the innumerable species throughout nature presenting well-marked differences; whereas varieties, the supposed prototypes and parents of future well-marked species, present slight and ill-defined differences. Mere chance, as we may call it, might cause one variety to differ in some character from its parents, and the offspring of this variety again to differ from its parent in the very same character and in a greater degree; but this alone would never account for so habitual and large an amount of difference as that between varieties of the same species and species of the same genus.

As has always been my practice, let us seek light on this head from our 43
domestic productions. We shall here find something analogous. A fancier is struck by a pigeon having a slightly shorter beak; another fancier is struck by a pigeon having a rather longer beak; and on the acknowledged principle that "fanciers do not and will not admire a medium standard, but like extremes," they both go on (as has actually occurred with tumbler-pigeons) choosing and breeding from birds with longer and longer beaks, or with shorter and shorter beaks. Again, we may suppose that at an early period one man preferred swifter horses; another stronger and more bulky horses. The early differences would be very slight; in the course of time, from the continued selection of swifter horses by some breeders, and of stronger ones by others, the differences would become greater, and would be noted as forming two sub-breeds; finally, after the lapse of centuries, the sub-breeds would become converted into two well-established and distinct breeds. As the differences slowly become greater, the inferior animals with intermediate characters, being neither very swift nor very strong, will have been neglected, and will have tended

to disappear. Here, then, we see in man's productions the action of what may be called the principle of divergence, causing differences, at first barely appreciable, steadily to increase, and the breeds to diverge in character both from each other and from their common parent.

44 But how, it may be asked, can any analogous principle apply in nature? I believe it can and does apply most efficiently, from the simple circumstance that the more diversified the descendants from any one species become in structure, constitution, and habits, by so much will they be better enabled to seize on many and widely diversified places in the polity of nature, and so be enabled to increase in numbers.

45 We can clearly see this in the case of animals with simple habits. Take the case of a carnivorous quadruped, of which the number that can be supported in any country has long ago arrived at its full average. If its natural powers of increase be allowed to act, it can succeed in increasing (the country not undergoing any change in its conditions) only by its varying descendants seizing on places at present occupied by other animals: some of them, for instance, being enabled to feed on new kinds of prey, either dead or alive; some inhabiting new stations, climbing trees, frequenting water, and some perhaps becoming less carnivorous. The more diversified in habits and structure the descendants of our carnivorous animal became, the more places they would be enabled to occupy. What applies to one animal will apply throughout all time to all animals—that is, if they vary—for otherwise natural selection can do nothing. So it will be with plants. It has been experimentally proved, that if a plot of ground be sown with one species of grass, and a similar plot be sown with several distinct genera of grasses, a greater number of plants and a greater weight of dry herbage can thus be raised. The same has been found to hold good when first one variety and then several mixed varieties of wheat have been sown on equal spaces of ground. Hence, if any one species of grass were to go on varying, and those varieties were continually selected which differed from each other in at all the same manner as distinct species and genera of grasses differ from each other, a greater number of individual plants of this species of grass, including its modified descendants, would succeed in living on the same piece of ground. And we well know that each species and each variety of grass is annually sowing almost countless seeds; and thus, as it may be said, is striving its utmost to increase its numbers. Consequently, I cannot doubt that in the course of many thousands of generations, the most distinct varieties of any one species of grass

would always have the best chance of succeeding and of increasing in numbers, and thus of supplanting the less distinct varieties; and varieties, when rendered very distinct from each other, take the rank of species.

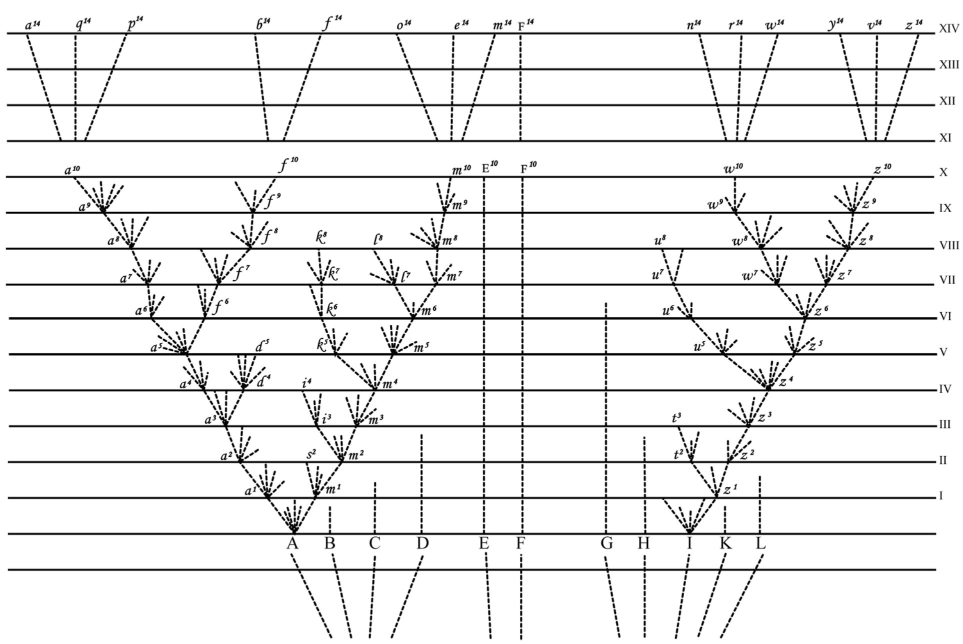
The truth of the principle, that the greatest amount of life can be supported by 46
great diversification of structure, is seen under many natural circumstances. In an extremely small area, especially if freely open to immigration, and where the contest between individual and individual must be severe, we always find great diversity in its inhabitants. For instance, I found that a piece of turf, three feet by four in size, which had been exposed for many years to exactly the same conditions, supported twenty species of plants, and these belonged to eighteen genera and to eight orders, which shows how much these plants differed from each other. So it is with the plants and insects on small and uniform islets; and so in small ponds of fresh water. Farmers find that they can raise most food by a rotation of plants belonging to the most different orders: nature follows what may be called a simultaneous rotation. Most of the animals and plants which live close round any small piece of ground, could live on it (supposing it not to be in any way peculiar in its nature), and may be said to be striving to the utmost to live there; but, it is seen, that where they come into the closest competition with each other, the advantages of diversification of structure, with the accompanying differences of habit and constitution, determine that the inhabitants, which thus jostle each other most closely, shall, as a general rule, belong to what we call different genera and orders.

[. . .]

After the foregoing discussion, which ought to have been much amplified, we 50
may, I think, assume that the modified descendants of any one species will succeed by so much the better as they become more diversified in structure, and are thus enabled to encroach on places occupied by other beings. Now let us see how this principle of great benefit being derived from divergence of character, combined with the principles of natural selection and of extinction, will tend to act.

The accompanying diagram will aid us in understanding this rather perplexing 51
subject. Let A to L represent the species of a genus large in its own country; these species are supposed to resemble each other in unequal degrees, as is so generally the case in nature, and as is represented in the diagram by the letters standing at unequal distances. I have said a large genus, because we have seen in the second

chapter, that on an average more of the species of large genera vary than of small genera; and the varying species of the large genera present a greater number of varieties. We have, also, seen that the species, which are the commonest and the most widely-diffused, vary more than rare species with restricted ranges. Let (A) be a common, widely-diffused, and varying species, belonging to a genus large in its own country. The little fan of diverging dotted lines of unequal lengths proceeding from (A), may represent its varying offspring. The variations are supposed to be extremely slight, but of the most diversified nature; they are not supposed all to appear simultaneously, but often after long intervals of time; nor are they all supposed to endure for equal periods. Only those variations which are in some way profitable will be preserved or naturally selected. And here the importance of the principle of benefit being derived from divergence of character comes in; for this will generally lead to the most different or divergent variations (represented by the outer dotted lines) being preserved and accumulated by natural selection. When a dotted line reaches one of the horizontal lines, and is there marked by a small numbered letter, a sufficient amount of variation is supposed to have been accumulated to have formed a fairly well-marked variety, such as would be thought worthy of record in a systematic work.



The intervals between the horizontal lines in the diagram, may represent each a thousand generations; but it would have been better if each had represented ten thousand generations. After a thousand generations, species (A) is supposed to have produced two fairly well-marked varieties, namely a^1 and m^1. These two varieties will generally continue to be exposed to the same conditions which made their parents variable, and the tendency to variability is in itself hereditary, consequently they will tend to vary, and generally to vary in nearly the same manner as their parents varied. Moreover, these two varieties, being only slightly modified forms, will tend to inherit those advantages which made their common parent (A) more numerous than most of the other inhabitants of the same country; they will likewise partake of those more general advantages which made the genus to which the parent-species belonged, a large genus in its own country. And these circumstances we know to be favourable to the production of new varieties. 52

If, then, these two varieties be variable, the most divergent of their variations will generally be preserved during the next thousand generations. And after this interval, variety a^1 is supposed in the diagram to have produced variety a^2, which will, owing to the principle of divergence, differ more from (A) than did variety a^1. Variety m^1 is supposed to have produced two varieties, namely m^2 and s^2, differing from each other, and more considerably from their common parent (A). We may continue the process by similar steps for any length of time; some of the varieties, after each thousand generations, producing only a single variety, but in a more and more modified condition, some producing two or three varieties, and some failing to produce any. Thus the varieties or modified descendants, proceeding from the common parent (A), will generally go on increasing in number and diverging in character. In the diagram the process is represented up to the ten-thousandth generation, and under a condensed and simplified form up to the fourteen-thousandth generation. 53

But I must here remark that I do not suppose that the process ever goes on so regularly as is represented in the diagram, though in itself made somewhat irregular. I am far from thinking that the most divergent varieties will invariably prevail and multiply: a medium form may often long endure, and may or may not produce more than one modified descendant; for natural selection will always act according to the nature of the places which are either unoccupied or not perfectly occupied by other beings; and this will depend on infinitely complex relations. But as a general 54 rule, the more diversified in structure the descendants from any one species can be rendered, the more places they will be enabled to seize on, and the more their modified progeny will be increased. In our diagram the line of succession is broken at regular intervals by small numbered letters marking the successive forms which have become sufficiently distinct to be recorded as varieties. But these breaks are imaginary, and might have been inserted anywhere, after intervals long enough to have allowed the accumulation of a considerable amount of divergent variation.

55 As all the modified descendants from a common and widely-diffused species, belonging to a large genus, will tend to partake of the same advantages which made their parent successful in life, they will generally go on multiplying in number as well as diverging in character: this is represented in the diagram by the several divergent branches proceeding from (A). The modified offspring from the later and more highly improved branches in the lines of descent, will, it is probable, often take the place of, and so destroy, the earlier and less improved branches: this is represented in the diagram by some of the lower branches not reaching to the upper horizontal lines. In some cases I do not doubt that the process of modification will be confined to a single line of descent, and the number of the descendants will not be increased; although the amount of divergent modification may have been increased in the successive generations. This case would be represented in the diagram, if all the lines proceeding from (A) were removed, excepting that from a^1 to a^{10}. In the same way, for instance, the English race-horse and English pointer have apparently both gone on slowly diverging in character from their original stocks, without either having given off any fresh branches or races.

56 After ten thousand generations, species (A) is supposed to have produced three forms, a^{10}, f^{10}, and m^{10}, which, from having diverged in character during the successive generations, will have come to differ largely, but perhaps unequally, from each other and from their common parent. If we suppose the amount of change between each horizontal line in our diagram to be excessively small, these three forms may still be only well-marked varieties; or they may have arrived at the doubtful category of sub-species; but we have only to suppose the steps in the process of modification to be more numerous or greater in amount, to convert these three forms into well-defined species: thus the diagram illustrates the steps by which the small differences distinguishing varieties are increased into the larger differences distinguishing species. By continuing the same process for a greater number of

generations (as shown in the diagram in a condensed and simplified manner), we get eight species, marked by the letters between a^{14} and m^{14}, all descended from (A). Thus, as I believe, species are multiplied and genera are formed.

In a large genus it is probable that more than one species would vary. In the 57
diagram I have assumed that a second species (I) has produced, by analogous steps, after ten thousand generations, either two well-marked varieties (w^{10} and z^{10}) or two species, according to the amount of change supposed to be represented between the horizontal lines. After fourteen thousand generations, six new species, marked by the letters n^{14} to z^{14}, are supposed to have been produced. In each genus, the species, which are already extremely different in character, will generally tend to produce the greatest number of modified descendants; for these will have the best chance of filling new and widely different places in the polity of nature: hence in the diagram I have chosen the extreme species (A), and the nearly extreme species (I), as those which have largely varied, and have given rise to new varieties and species. The other nine species (marked by capital letters) of our original genus, may for a long period continue transmitting unaltered descendants; and this is shown in the diagram by the dotted lines not prolonged far upwards from want of space.

But during the process of modification, represented in the diagram, another 58
of our principles, namely that of extinction, will have played an important part. As in each fully stocked country natural selection necessarily acts by the selected form having some advantage in the struggle for life over other forms, there will be a constant tendency in the improved descendants of any one species to supplant and exterminate in each stage of descent their predecessors and their original parent. For it should be remembered that the competition will generally be most severe between those forms which are most nearly related to each other in habits, constitution, and structure. Hence all the intermediate forms between the earlier and later states, that is between the less and more improved state of a species, as well as the original parent-species itself, will generally tend to become extinct. So it probably will be with many whole collateral lines of descent, which will be conquered by later and improved lines of descent. If, however, the modified offspring of a species get into some distinct country, or become quickly adapted to some quite new station, in which child and parent do not come into competition, both may continue to exist.

If then our diagram be assumed to represent a considerable amount of 59
modification, species (A) and all the earlier varieties will have become extinct,

having been replaced by eight new species (a^{14} to m^{14}); and (I) will have been replaced by six (n^{14} to z^{14}) new species.

60 But we may go further than this. The original species of our genus were supposed to resemble each other in unequal degrees, as is so generally the case in nature; species (A) being more nearly related to B, C, and D, than to the other species; and species (I) more to G, H, K, L, than to the others. These two species (A) and (I), were also supposed to be very common and widely diffused species, so that they must originally have had some advantage over most of the other species of the genus. Their modified descendants, fourteen in number at the fourteen-thousandth generation, will probably have inherited some of the same advantages: they have also been modified and improved in a diversified manner at each stage of descent, so as to have become adapted to many related places in the natural economy of their country. It seems, therefore, to me extremely probable that they will have taken the places of, and thus exterminated, not only their parents (A) and (I), but likewise some of the original species which were most nearly related to their parents. Hence very few of the original species will have transmitted offspring to the fourteen-thousandth generation. We may suppose that only one (F), of the two species which were least closely related to the other nine original species, has transmitted descendants to this late stage of descent.

61 The new species in our diagram descended from the original eleven species, will now be fifteen in number. Owing to the divergent tendency of natural selection, the extreme amount of difference in character between species a^{14} and z^{14} will be much greater than that between the most different of the original eleven species. The new species, moreover, will be allied to each other in a widely different manner. Of the eight descendants from (A) the three marked a^{14}, q^{14}, p^{14}, will be nearly related from having recently branched off from a^{10}; b^{14} and f^{14}, from having diverged at an earlier period from a^{5}, will be in some degree distinct from the three first-named species; and lastly, o^{14}, e^{14}, and m^{14}, will be nearly related one to the other, but from having diverged at the first commencement of the process of modification, will be widely different from the other five species, and may constitute a sub-genus or even a distinct genus.

62 The six descendants from (I) will form two sub-genera or even genera. But as the original species (I) differed largely from (A), standing nearly at the extreme points of the original genus, the six descendants from (I) will, owing to inheritance,

differ considerably from the eight descendants from (A); the two groups, moreover, are supposed to have gone on diverging in different directions. The intermediate species, also (and this is a very important consideration), which connected the original species (A) and (I), have all become, excepting (F), extinct, and have left no descendants. Hence the six new species descended from (I), and the eight descended from (A), will have to be ranked as very distinct genera, or even as distinct sub-families.

Thus it is, as I believe, that two or more genera are produced by descent, with 63
modification, from two or more species of the same genus. And the two or more parent-species are supposed to have descended from some one species of an earlier genus. In our diagram, this is indicated by the broken lines, beneath the capital letters, converging in sub-branches downwards towards a single point; this point representing a single species, the supposed single parent of our several new sub-genera and genera.

[. . .]

Summary of Chapter.—If during the long course of ages and under 68
varying conditions of life, organic beings vary at all in the several parts of their organisation, and I think this cannot be disputed; if there be, owing to the high geometrical powers of increase of each species, at some age, season, or year, a severe struggle for life, and this certainly cannot be disputed; then, considering the infinite complexity of the relations of all organic beings to each other and to their conditions of existence, causing an infinite diversity in structure, constitution, and habits, to be advantageous to them, I think it would be a most extraordinary fact if no variation ever had occurred useful to each being's own welfare, in the same way as so many variations have occurred useful to man. But if variations useful to any organic being do occur, assuredly individuals thus characterised will have the best chance of being preserved in the struggle for life; and from the strong principle of inheritance they will tend to produce offspring similarly characterised. This principle of preservation, I have called, for the sake of brevity, Natural Selection. Natural selection, on the principle of qualities being inherited at corresponding ages, can modify the egg, seed, or young, as easily as the adult. Amongst many animals, sexual selection will give its aid to ordinary selection, by assuring to the

most vigorous and best adapted males the greatest number of offspring. Sexual selection will also give characters useful to the males alone, in their struggles with other males.

69 Whether natural selection has really thus acted in nature, in modifying and adapting the various forms of life to their several conditions and stations, must be judged of by the general tenour and balance of evidence given in the following chapters. But we already see how it entails extinction; and how largely extinction has acted in the world's history, geology plainly declares. Natural selection, also, leads to divergence of character; for more living beings can be supported on the same area the more they diverge in structure, habits, and constitution, of which we see proof by looking at the inhabitants of any small spot or at naturalised productions. Therefore during the modification of the descendants of any one species, and during the incessant struggle of all species to increase in numbers, the more diversified these descendants become, the better will be their chance of succeeding in the battle of life. Thus the small differences distinguishing varieties of the same species, will steadily tend to increase till they come to equal the greater differences between species of the same genus, or even of distinct genera.

70 We have seen that it is the common, the widely-diffused, and widely-ranging species, belonging to the larger genera, which vary most; and these will tend to transmit to their modified offspring that superiority which now makes them dominant in their own countries. Natural selection, as has just been remarked, leads to divergence of character and to much extinction of the less improved and intermediate forms of life. On these principles, I believe, the nature of the affinities of all organic beings may be explained. It is a truly wonderful fact—the wonder of which we are apt to overlook from familiarity—that all animals and all plants throughout all time and space should be related to each other in group subordinate to group, in the manner which we everywhere behold—namely, varieties of the same species most closely related together, species of the same genus less closely and unequally related together, forming sections and sub-genera, species of distinct genera much less closely related, and genera related in different degrees, forming sub-families, families, orders, sub-classes, and classes. The several subordinate groups in any class cannot be ranked in a single file, but seem rather to be clustered round points, and these round other points, and so on in almost endless cycles. On

the view that each species has been independently created, I can see no explanation of this great fact in the classification of all organic beings; but, to the best of my judgment, it is explained through inheritance and the complex action of natural selection, entailing extinction and divergence of character, as we have seen illustrated in the diagram.

The affinities of all the beings of the same class have sometimes been 71
represented by a great tree. I believe this simile largely speaks the truth. The green and budding twigs may represent existing species; and those produced during each former year may represent the long succession of extinct species. At each period of growth all the growing twigs have tried to branch out on all sides, and to overtop and kill the surrounding twigs and branches, in the same manner as species and groups of species have tried to overmaster other species in the great battle for life. The limbs divided into great branches, and these into lesser and lesser branches, were themselves once, when the tree was small, budding twigs; and this connexion of the former and present buds by ramifying branches may well represent the classification of all extinct and living species in groups subordinate to groups. Of the many twigs which flourished when the tree was a mere bush, only two or three, now grown into great branches, yet survive and bear all the other branches; so with the species which lived during long-past geological periods, very few now have living and modified descendants. From the first growth of the tree, many a limb and branch has decayed and dropped off; and these lost branches of various sizes may represent those whole orders, families, and genera which have now no living representatives, and which are known to us only from having been found in a fossil state. As we here and there see a thin straggling branch springing from a fork low down in a tree, and which by some chance has been favoured and is still alive on its summit, so we occasionally see an animal like the Ornithorhynchus or Lepidosiren, which in some small degree connects by its affinities two large branches of life, and which has apparently been saved from fatal competition by having inhabited a protected station. As buds give rise by growth to fresh buds, and these, if vigorous, branch out and overtop on all sides many a feebler branch, so by generation I believe it has been with the great Tree of Life, which fills with its dead and broken branches the crust of the earth, and covers the surface with its ever branching and beautiful ramifications.

Text 5

from

DNA: The Secret of Life

by James D. Watson

CHAPTER ONE
BEGINNINGS OF GENETICS: FROM MENDEL TO HITLER

My mother, Bonnie Jean, believed in genes. She was proud of her father's Scottish 1
origins, and saw in him the traditional Scottish virtues of honesty, hard work, and thriftiness. She, too, possessed these qualities and felt that they must have been passed down to her from him. His tragic early death meant that her only nongenetic legacy was a set of tiny little girl's kilts he had ordered for her from Glasgow. Perhaps therefore it is not surprising that she valued her father's biological legacy over his material one.

Growing up, I had endless arguments with Mother about the relative roles 2
played by nature and nurture in shaping us. By choosing nurture over nature, I was effectively subscribing to the belief that I could make myself into whatever I wanted to be. I did not want to accept that my genes mattered that much, preferring to attribute my Watson grandmother's extreme fatness to her having overeaten. If her shape was the product of her genes, then I too might have a hefty future. However, even as a teenager, I would not have disputed the evident basics of inheritance, that like begets like. My arguments with my mother concerned complex characteristics

like aspects of personality, not the simple attributes that, even as an obstinate adolescent, I could see were passed down over the generations, resulting in "family likeness." My nose is my mother's and now belongs to my son Duncan.

3 Sometimes characteristics come and go within a few generations, but sometimes they persist over many. One of the most famous examples of a long-lived trait is known as the "Hapsburg Lip." This distinctive elongation of the jaw and droopiness to the lower lip—which made the Hapsburg rulers of Europe such a nightmare assignment for generations of court portrait painters—was passed down intact over at least twenty-three generations.

At age eleven, with my sister Elizabeth and my father, James

4 The Hapsburgs added to their genetic woes by intermarrying. Arranging marriages between different branches of the Hapsburg clan and often among close relatives may have made political sense as a way of building alliances and ensuring dynastic succession, but it was anything but astute in genetic terms. Inbreeding of this kind can result in genetic disease, as the Hapsburgs found out to their cost. Charles II, the last of the Hapsburg monarchs in Spain, not only boasted a prize-worthy example of the family lip—he could not even chew his own food—but was also a complete invalid, and incapable, despite two marriages, of producing children.

5 Genetic disease has long stalked humanity. In some cases, such as Charles II's, it has had a direct impact on history. Retrospective diagnosis has suggested that

George III, the English king whose principal claim to fame is to have lost the American colonies in the Revolutionary War, suffered from an inherited disease, porphyria, which causes periodic bouts of madness. Some historians—mainly British ones—have argued that it was the distraction caused by George's illness that permitted the Americans' against-the-odds military success. While most hereditary diseases have no such geopolitical impact, they nevertheless have brutal and often tragic consequences for the afflicted families, sometimes for many generations. Understanding genetics is not just about understanding why we look like our parents. It is also about coming to grips with some of humankind's oldest enemies: the flaws in our genes that cause genetic disease.

Our ancestors must have wondered about the workings of heredity as soon as evolution endowed them with brains capable of formulating the right kind of question. And the readily observable principle that close relatives tend to be similar can carry you a long way if, like our ancestors, your concern with the application of genetics is limited to practical matters like improving domesticated animals (for, say, milk yield in cattle) and plants (for, say, the size of fruit). Generations of careful selection—breeding initially to domesticate appropriate species, and then breeding only from the most productive cows and from the trees with the largest fruit—resulted in animals and plants tailor-made for human purposes. Underlying this enormous unrecorded effort is that simple rule of thumb: that the most productive cows will produce highly productive offspring and from the seeds of trees with large fruit large-fruited trees will grow. Thus, despite the extraordinary advances of the past hundred years or so, the twentieth and twenty-first centuries by no means have a monopoly on genetic insight. Although it wasn't until 1909 that the British biologist William Bateson gave the science of inheritance a name, genetics, and although the DNA revolution has opened up new and extraordinary vistas of potential progress, in fact the single greatest application of genetics to human well-being was carried out eons ago by anonymous ancient farmers. Almost everything we eat—cereals, fruit, meat, dairy products—is the legacy of that earliest and most far-reaching application of genetic manipulations to human problems. 6

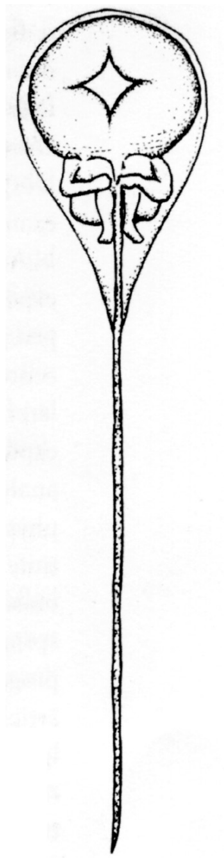
An understanding of the actual mechanics of genetics proved a tougher nut to crack. Gregor Mendel (1822–1884) published his famous paper on the subject in 1866 (and it was ignored by the scientific community for another thirty-four years). Why did it take so long? After all, heredity is a major aspect of the natural world, and, more important, it is readily, and universally, observable: a dog owner sees 7 how a cross between a brown and black dog turns out, and all parents consciously or subconsciously track the appearance of their own characteristics in their children. One simple reason is that genetic mechanisms turn out to be complicated. Mendel's solution to the problem is not intuitively obvious: children are not, after all, simply a *blend* of their parents' characteristics. Perhaps most important was the failure by early biologists to distinguish between two fundamentally different processes, heredity and development. Today we understand that a fertilized egg contains the genetic information, contributed by both parents, that determines whether someone will be afflicted with, say, porphyria. That is heredity. The subsequent process, the *development* of a new individual from that humble starting point of a single cell, the fertilized egg, involves implementing that information. Broken down in terms of academic disciplines, genetics focuses on the information and developmental biology focuses on the use of that information. Lumping heredity and development together into a single phenomenon, early scientists never asked the questions that might have steered them toward the secret of heredity. Nevertheless, the effort had been under way in some form since the dawn of Western history.

8 The Greeks, including Hippocrates, pondered heredity. They devised a theory of "pangenesis," which claimed that sex involved the transfer of miniaturized body parts: "Hairs, nails, veins, arteries, tendons and their bones, albeit invisible as their particles are so small. While growing, they gradually separate from each other." This idea enjoyed a brief renaissance when Charles Darwin, desperate to support his theory of evolution by natural selection with a viable hypothesis of inheritance, put forward a modified version of pangenesis in the second half of the nineteenth century. In Darwin's scheme, each organ—eyes, kidneys, bones—contributed circulating "gemmules" that accumulated in the sex organs, and were ultimately exchanged in the course of sexual reproduction. Because these gemmules were produced throughout an organism's lifetime, Darwin argued any change that occurred in the individual after birth, like the stretch of a giraffe's neck imparted by craning for the highest foliage, could be passed on to the next generation. Ironically, then, to buttress his theory of natural selection Darwin came to champion aspects of Jean-Baptiste Lamarck's theory of inheritance of acquired characteristics—the very theory that his evolutionary ideas did so much to discredit. Darwin was invoking

only Lamarck's theory of inheritance; he continued to believe that natural selection was the driving force behind evolution, but supposed that natural selection operated on the variation produced by pangenesis. Had Darwin known about Mendel's work (although Mendel published his results shortly after *The Origin of Species* appeared, Darwin was never aware of them), he might have been spared the embarrassment of this late-career endorsement of some of Lamarck's ideas.

Whereas pangenesis supposed that embryos were assembled from a set of 9
minuscule components, another approach, "preformationism," avoided the assembly step altogether: either the egg or the sperm (exactly which was a contentious issue) contained a complete *preformed* individual called a homunculus. Development was therefore merely a matter of enlarging this into a fully formed being. In the days of preformationism, what we now recognize as genetic disease was variously interpreted: sometimes as a manifestation of the wrath of God or the mischief of demons and devils; sometimes as evidence of either an excess of or a deficit of the father's "seed"; sometimes as the result of "wicked thoughts" on the part of the mother during pregnancy. On the premise that fetal malformation can result when a pregnant mother's desires are thwarted, leaving her feeling stressed and frustrated, Napoleon passed a law permitting expectant mothers to shoplift. None of these notions, needless to say, did much to advance our understanding of genetic disease.

By the early nineteenth century, better microscopes had defeated preforma- 10
tionism. Look as hard as you like, you will never see a tiny homunculus curled up inside a sperm or egg cell. Pangenesis, though an earlier misconception, lasted rather longer—the argument would persist that the gemmules were simply too small to visualize—but was eventually laid to rest by August Weismann, who argued that inheritance depended on the continuity of germ plasm between generations and thus changes to the body over an individual's lifetime could *not* be transmitted to subsequent generations. His simple experiment involved cutting the tails off several generations of mice. According to Darwin's pangenesis, tailless mice would produce gemmules signifying "no tail" and so their offspring should develop a severely stunted hind appendage or none at all. When Weismann showed that the tail kept appearing after many generations of amputees, pangenesis bit the dust.



Genetics before Mendel: a homunculus, a preformed miniature person imagined to exist in the head of a sperm cell

11 Gregor Mendel was the one who got it right. By any standards, however, he was an unlikely candidate for scientific superstardom. Born to a farming family in what is now the Czech Republic, he excelled at the village school and, at twenty-one, entered the Augustinian monastery at Brünn. After proving a disaster as a parish priest—his response to the ministry was a nervous breakdown—he tried his hand at teaching. By all accounts he was a good teacher, but in order to qualify to teach a full range of subjects, he had to take an exam. He failed it. Mendel's father superior, Abbot Napp, then dispatched him to the University of Vienna, where he was to bone up full-time for the retesting. Despite apparently doing well in physics at Vienna, Mendel again failed the exam, and so never rose above the rank of substitute teacher.

Around 1856, at Abbot Napp's suggestion, Mendel undertook some scientific experiments on heredity. He chose to study a number of characteristics of the pea plants he grew in his own patch of the monastery garden. In 1865 he presented his results to the local natural history society in two lectures, and, a year later, published them in the society's journal. The work was a tour de force: the experiments were brilliantly designed and painstakingly executed, and his analysis of the results was insightful and deft. It seems that his training in physics contributed to his breakthrough because, unlike other biologists of that time, he approached the problem quantitatively. Rather than simply noting that crossbreeding of red and white flowers resulted in some red and some white offspring, Mendel actually counted them, realizing that the ratios of red to white progeny might be significant—as indeed they are. Despite sending copies of his article to various prominent scientists, Mendel found himself completely ignored by the scientific community. His attempt to draw attention to his results merely backfired. He wrote to his one contact among the ranking scientists of the day, botanist Karl Nägeli in Munich, asking him to replicate the experiments, and he duly sent off 140 carefully labeled packets of seeds. He should not have bothered. Nägeli believed that the obscure monk should be of service to him, rather than the other way around, so he sent Mendel seeds of his own favorite plant, hawkweed, challenging the monk to re-create his results with a different species. Sad to say, for various reasons, hawkweed is not well-suited to breeding experiments such as those Mendel had performed on the peas. The entire exercise was a waste of his time. 12

Mendel's low-profile existence as monk-teacher-researcher ended abruptly in 1868 when, on Napp's death, he was elected abbot of the monastery. Although he continued his research—increasingly on bees and the weather—administrative duties were a burden, especially as the monastery became embroiled in a messy dispute over back taxes. Other factors, too, hampered him as a scientist. Portliness eventually curtailed his fieldwork: as he wrote, hill climbing had become "very difficult for me in a world where universal gravitation prevails." His doctors prescribed tobacco to keep his weight in check, and he obliged them by smoking twenty cigars a day, as many as Winston Churchill. It was not his lungs, however, that let him down: in 1884, at the age of sixty-one, Mendel succumbed to a combination of heart and kidney disease. 13

Not only were Mendel's results buried in an obscure journal, but they would have been unintelligible to most scientists of the era. He was far ahead of his time 14 with his combination of careful experiment and sophisticated quantitative analysis. Little wonder, perhaps, that it was not until 1900 that the scientific community caught up with him. The rediscovery of Mendel's work, by three plant geneticists interested in similar problems, provoked a revolution in biology. At last the scientific world was ready for the monk's peas.

15 Mendel realized that there are specific factors—later to be called "genes"—that are passed from parent to offspring. He worked out that these factors come in pairs and that the offspring receives one from each parent.

16 Noticing that peas came in two distinct colors, green and yellow, he deduced that there were two versions of the pea-color gene. A pea has to have two copies of the G version if it is to become green, in which case we say that it is GG for the pea-color gene. It must therefore have received a G pea-color gene from both of its parents. However, yellow peas can result both from YY and YG combinations. Having only one copy of the Y version is sufficient to produce yellow peas. Y trumps G. Because in the YG case the Y signal dominates the G signal, we call Y "dominant." The subordinate G version of the pea-color gene is called "recessive."

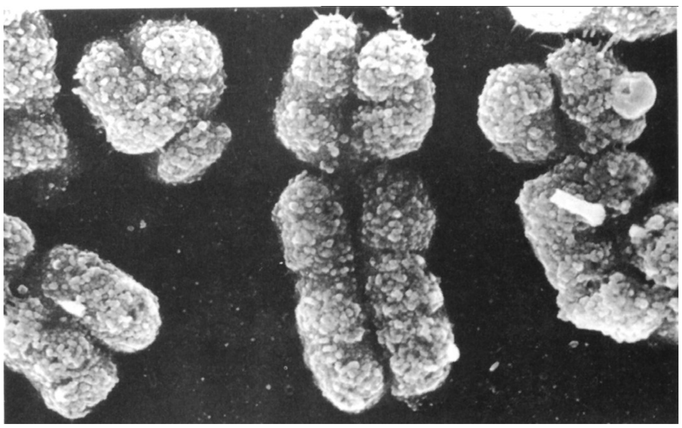
17 Each parent pea plant has two copies of the pea-color gene, yet it contributes only one copy to each offspring; the other copy is furnished by the other parent. In plants, pollen grains contain sperm cells—the male contribution to the next generation—and each sperm cell contains just one copy of the pea-color gene. A parent pea plant with a YG combination will produce sperm that contain either a Y version or a G one. Mendel discovered that the process is random: 50 percent of the sperm produced by that plant will have a Y and 50 percent will have a G.

18 Suddenly many of the mysteries of heredity made sense. Characteristics, like the Hapsburg Lip, that are transmitted with a high probability (actually 50 percent) from generation to generation are dominant. Other characteristics that appear in family trees much more sporadically, often skipping generations, may be recessive. When a gene is recessive an individual has to have two copies of it for the corresponding trait to be expressed. Those with one copy of the gene are carriers: they don't themselves exhibit the characteristic, but they can pass the gene on. Albinism, in which the body fails to produce pigment so the skin and hair are strikingly white, is an example of a recessive characteristic that is transmitted in this way. Therefore, to be albino you have to have two copies of the gene, one from each parent. (This was the case with the Reverend Dr. William Archibald

Spooner, who was also—perhaps only by coincidence—prone to a peculiar form of linguistic confusion whereby, for example, "a well-oiled bicycle" might become "a well-boiled icicle." Such reversals would come to be termed "spoonerisms" in his honor.) Your parents, meanwhile, may have shown no sign of the gene at all. If, as is often the case, each has only one copy, then they are both carriers. The trait has skipped at least one generation.

Mendel's results implied that *things*—material objects—were transmitted from generation to generation. But what was the nature of these things? 19

At about the time of Mendel's death in 1884, scientists using ever-improving optics to study the minute architecture of cells coined the term "chromosome" to describe the long stringy bodies in the cell nucleus. But it was not until 1902 that Mendel and chromosomes came together. 20

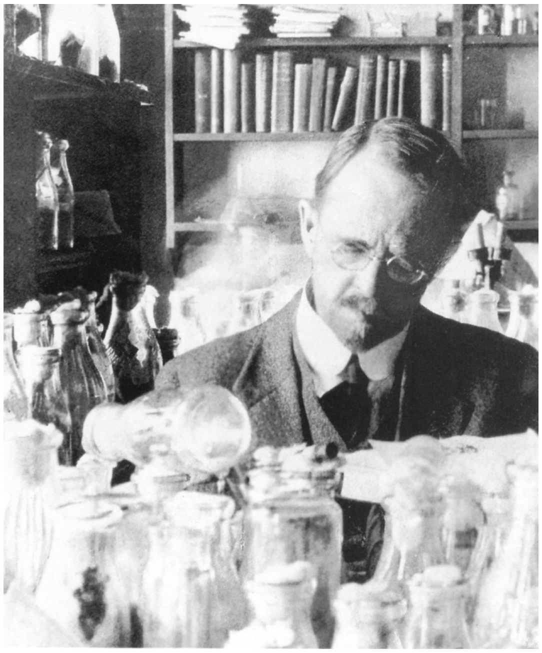


The human X chromosome, as seen with an electron microscope

A medical student at Columbia University, Walter Sutton, realized that chromosomes had a lot in common with Mendel's mysterious factors. Studying grasshopper chromosomes, Sutton noticed that most of the time they are doubled up—just like Mendel's paired factors. But Sutton also identified one type of cell in which chromosomes were not paired: the sex cells. Grasshopper sperm have only a single set of chromosomes, not a double set. This was exactly what Mendel had described: his pea plant sperm cells also only carried a single copy of each of his factors. It was clear that Mendel's factors, now called genes, must be on the chromosomes. 21

22 In Germany Theodor Boveri independently came to the same conclusions as Sutton, and so the biological revolution their work had precipitated came to be called the Sutton-Boveri chromosome theory of inheritance. Suddenly genes were real. They were on chromosomes, and you could actually see chromosomes through the microscope.

23 Not everyone bought the Sutton-Boveri theory. One skeptic was Thomas Hunt Morgan, also at Columbia. Looking down the microscope at those stringy chromosomes, he could not see how they could account for all the changes that occur from one generation to the next. If all the genes were arranged along chromosomes, and all chromosomes were transmitted intact from one generation to the next, then surely many characteristics would be inherited together. But since empirical evidence showed this not to be the case, the chromosomal theory seemed insufficient to explain the variation observed in nature. Being an astute experimentalist, however, Morgan had an idea how he might resolve such discrepancies. He turned to the fruit fly, *Drosophila melanogaster*, the drab little beast that, ever since Morgan, has been so beloved by geneticists.



Notoriously camera shy T. H. Morgan was photographed surreptitiously while at work in the fly room.

In fact, Morgan was not the first to use the fruit fly in breeding experiments—that distinction belonged to a lab at Harvard that first put the critter to work in 1901—but it was Morgan's work that put the fly on the scientific map. *Drosophila* is a good choice for genetic experiments. It is easy to find (as anyone who has left out a bunch of overripe bananas during the summer well knows); it is easy to raise (bananas will do as feed); and you can accommodate hundreds of flies in a single milk bottle (Morgan's students had no difficulty acquiring milk bottles, pinching them at dawn from doorsteps in their Manhattan neighborhood); and it breeds and breeds and breeds (a whole generation takes about ten days, and each female lays several hundred eggs). Starting in 1907 in a famously squalid, cockroach-infested, banana-stinking lab that came to be known affectionately as the "fly room," Morgan and his students ("Morgan's boys" as they were called) set to work on fruit flies. 24

Unlike Mendel, who could rely on the variant strains isolated over the years by farmers and gardeners—yellow peas as opposed to green ones, wrinkled skin as opposed to smooth—Morgan had no menu of established genetic differences in the fruit fly to draw upon. And you cannot do genetics until you have isolated some distinct characteristics to track through the generations. Morgan's first goal therefore was to find "mutants," the fruit fly equivalents of yellow or wrinkled peas. He was looking for genetic novelties, random variations that somehow simply appeared in the population. 25

One of the first mutants Morgan observed turned out to be one of the most instructive. While normal fruit flies have red eyes, these had white ones. And he noticed that the white-eyed flies were typically male. It was known that the sex of a fruit fly—or, for that matter, the sex of a human—is determined chromosomally: females have two copies of the X chromosome, whereas males have one copy of the X and one copy of the much smaller Y. In light of this information, the white-eye result suddenly made sense: the eye-color gene is located on the X chromosome and the white-eye mutation, W, is recessive. Because males have only a single X chromosome, even recessive genes, in the absence of a dominant counterpart to suppress them, are automatically expressed. White-eyed females were relatively rare because they typically had only one copy of W so they expressed the dominant red eye color. By correlating a gene—the one for eye color—with a chromosome, the X, Morgan, despite his initial reservations, had effectively proved the Sutton-Boveri theory. He had also found an example of "sex-linkage," in which a particular characteristic is disproportionately represented in one sex. 26

27 Like Morgan's fruit flies, Queen Victoria provides a famous example of sex-linkage. On one of her X chromosomes, she had a mutated gene for hemophilia, the "bleeding disease" in whose victims proper blood clotting fails to occur. Because her other copy was normal, and the hemophilia gene is recessive, she herself did not have the disease. But she was a carrier. Her daughters did not have the disease either; evidently each possessed at least one copy of the normal version. But Victoria's sons were not all so lucky. Like all males (fruit fly males included), each had only one X chromosome; this was necessarily derived from Victoria (a Y chromosome could have come only from Prince Albert, Victoria's husband). Because Victoria had one mutated copy and one normal copy, each of her sons had a 50-50 chance of having the disease. Prince Leopold drew the short straw: he developed hemophilia, and died at thirty-one, bleeding to death after a minor fall. Two of Victoria's daughters, Princesses Alice and Beatrice, were carriers, having inherited the mutated gene from their mother. They each produced carrier daughters and sons with hemophilia. Alice's grandson Alexis, heir to the Russian throne, had hemophilia, and would doubtless have died young had the Bolsheviks not gotten to him first.

28 Morgan's fruit flies had other secrets to reveal. In the course of studying genes located on the same chromosome, Morgan and his students found that chromosomes actually break apart and re-form during the production of sperm and egg cells. This meant that Morgan's original objections to the Sutton-Boveri theory were unwarranted: the breaking and re-forming—"recombination," in modern genetic parlance—shuffles gene copies between members of a chromosome pair. This means that, say, the copy of chromosome 12 I got from my mother (the other, of course, comes from my father) is in fact a mix of my mother's two copies of chromosome 12, one of which came from her mother and one from her father. Her two 12s recombined—exchanged material—during the production of the egg cell that eventually turned into me. Thus my maternally derived chromosome 12 can be viewed as a mosaic of my grandparents' 12s. Of course, my mother's maternally derived 12 was itself a mosaic of her grandparents' 12s, and so on.

29 Recombination permitted Morgan and his students to map out the positions of particular genes along a given chromosome. Recombination involves breaking (and re-forming) chromosomes. Because genes are arranged like beads along a chromosome string, a break is statistically much more likely to occur between

two genes that are far apart (with more potential break points intervening) on the chromosome than between two genes that are close together. If, therefore, we see a lot of reshuffling for any two genes on a single chromosome, we can conclude that they are a long way apart; the rarer the reshuffling, the closer the genes likely are. This basic and immensely powerful principle underlies all of genetic mapping. One of the primary tools of scientists involved in the Human Genome Project and of researchers at the forefront of the battle against genetic disease was thus developed all those years ago in the filthy, cluttered Columbia fly room. Each new headline in the science section of the newspaper these days along the lines of "Gene for Something Located" is a tribute to the pioneering work of Morgan and his boys.

The rediscovery of Mendel's work, and the breakthroughs that followed it, 30
sparked a surge of interest in the social significance of genetics. While scientists had been grappling with the precise mechanisms of heredity through the eighteenth and nineteenth centuries, public concern had been mounting about the burden placed on society by what came to be called the "degenerate classes"—the inhabitants of poorhouses, workhouses, and insane asylums. What could be done with these people? It remained a matter of controversy whether they should be treated charitably—which, the less charitably inclined claimed, ensured such folk would never exert themselves and would therefore remain forever dependent on the largesse of the state or of private institutions—or whether they should be simply ignored, which, according to the charitably inclined, would result only in perpetuating the inability of the unfortunate to extricate themselves from their blighted circumstances.

The publication of Darwin's *Origin of Species* in 1859 brought these issues 31
into sharp focus. Although Darwin carefully omitted to mention human evolution, fearing that to do so would only further inflame an already raging controversy, it required no great leap of imagination to apply his idea of natural selection to humans. Natural selection is the force that determines the fate of all genetic variations in nature—mutations like the one Morgan found in the fruit fly eye-color gene, but also perhaps differences in the abilities of human individuals to fend for themselves.

Natural populations have an enormous reproductive potential. Take fruit flies, 32
with their generation time of just ten days, and females that produce some three hundred eggs apiece (half of which will be female): starting with a single fruit fly

couple, after a month (i.e., three generations later), you will have 150 × 150 × 150 fruit flies on your hands—that's more than 3 million flies, all of them derived from just one pair in just one month. Darwin made the point by choosing a species from the other end of the reproductive spectrum:

> The elephant is reckoned to be the slowest breeder of all known animals, and I have taken some pains to estimate its probable minimum rate of natural increase: it will be under the mark to assume that it breeds when thirty years old, and goes on breeding till ninety years old, bringing forth three pairs of young in this interval; if this be so, at the end of the fifth century there would be alive fifteen million elephants, descended from the first pair.

33 All these calculations assume that all the baby fruit flies and all the baby elephants make it successfully to adulthood. In theory, therefore, there must be an infinitely large supply of food and water to sustain this kind of reproductive overdrive. In reality, of course, those resources are limited, and not all baby fruit flies or baby elephants make it. There is competition among individuals within a species for those resources. What determines who wins the struggle for access to the resources? Darwin pointed out genetic variation means that some individuals have advantages in what he called "the struggle for existence." To take the famous example of Darwin's finches from the Galápagos Islands, those individuals with genetic advantages—like the right size of beak for eating the most abundant seeds—are more likely to survive and reproduce. So the advantageous genetic variant—having a bill the right size—tends to be passed on to the next generation. The result is that natural selection enriches the next generation with the beneficial mutation so that eventually, over enough generations, every member of the species ends up with that characteristic.

34 The Victorians applied the same logic to humans. They looked around and were alarmed by what they saw. The decent, moral, hardworking middle classes were being massively outreproduced by the dirty, immoral, lazy lower classes. The Victorians assumed that the virtues of decency, morality, and hard work ran in families just as the vices of filth, wantonness, and indolence did. Such characteristics must then be hereditary; thus, to the Victorians, morality and immorality were merely

two of Darwin's genetic variants. And if the great unwashed were outreproducing the respectable classes, then the "bad" genes would be increasing in the human population. The species was doomed! Humans would gradually become more and more depraved as the "immorality" gene became more and more common.

Francis Galton had good reason to pay special attention to Darwin's book, 35
as the author was his cousin and friend. Darwin, some thirteen years older, had provided guidance during Galton's rather rocky college experience. But it was *The Origin of Species* that would inspire Galton to start a social and genetic crusade that would ultimately have disastrous consequences. In 1883, a year after his cousin's death, Galton gave the movement a name: eugenics.

Eugenics was only one of Galton's many interests; Galton enthusiasts refer to 36
him as a polymath, detractors as a dilettante. In fact, he made significant contributions to geography, anthropology, psychology, genetics, meteorology, statistics, and, by setting fingerprint analysis on a sound scientific footing, to criminology. Born in 1822 into a prosperous family, his education—partly in medicine and partly in mathematics—was mostly a chronicle of defeated expectations. The death of his father when he was twenty-one simultaneously freed him from paternal restraint and yielded a handsome inheritance; the young man duly took advantage of both. After a full six years of being what might be described today as a trust-fund dropout, however, Galton settled down to become a productive member of the Victorian establishment. He made his name leading an expedition to a then little known region of southwest Africa in 1850–52. In his account of his explorations, we encounter the first instance of the one strand that connects his many varied interests: he counted and measured everything. Galton was only happy when he could reduce a phenomenon to a set of numbers.

At a missionary station he encountered a striking specimen of steatopygia— 37
a condition of particularly protuberant buttocks, common among the indigenous Nama women of the region—and realized that this woman was naturally endowed with the figure that was then fashionable in Europe. The only difference was that it required enormous (and costly) ingenuity on the part of European dressmakers to create the desired "look" for their clients.

> I profess to be a scientific man, and was exceedingly anxious to obtain accurate measurements of her shape; but there was a difficulty in doing

this. I did not know a word of Hottentot [the Dutch name for the Nama], and could never therefore have explained to the lady what the object of my footrule could be; and I really dared not ask my worthy missionary host to interpret for me. I therefore felt in a dilemma as I gazed at her form, that gift of bounteous nature to this favoured race, which no mantua-maker, with all her crinoline and stuffing, can do otherwise than humbly imitate. The object of my admiration stood under a tree, and was turning herself about to all points of the compass, as ladies who wish to be admired usually do. Of a sudden my eye fell upon my sextant; the bright thought struck me, and I took a series of observations upon her figure in every direction, up and down, crossways, diagonally, and so forth, and I registered them carefully upon an outline drawing for fear of any mistake; this being done, I boldly pulled out my measuring tape, and measured the distance from where I was to the place she stood, and having thus obtained both base and angles, I worked out the results by trigonometry and logarithms.

A nineteenth-century exaggerated view of a Nama woman

38 Galton's passion for quantification resulted in his developing many of the fundamental principles of modern statistics. It also yielded some clever observations. For example, he tested the efficacy of prayer. He figured that if prayer worked,

those most prayed for should be at an advantage; to test the hypothesis he studied the longevity of British monarchs. Every Sunday, congregations in the Church of England following the *Book of Common Prayer* beseeched God to "Endue the king/queen plenteously with heavenly gifts; Grant him/her in health and wealth long to live." Surely, Galton reasoned, the cumulative effect of all those prayers should be beneficial. In fact, prayer seemed ineffectual: he found that on average the monarchs died somewhat younger than other members of the British aristocracy.

Because of the Darwin connection—their common grandfather, Erasmus 39
Darwin, too was one of the intellectual giants of his day—Galton was especially sensitive to the way in which certain lineages seemed to spawn disproportionately large numbers of prominent and successful people. In 1869 he published what would become the underpinning of all his ideas on eugenics, a treatise called *Hereditary Genius: An Inquiry into Its Laws and Consequences*. In it he purported to show that talent, like simple genetic traits such as the Hapsburg Lip, does indeed run in families; he recounted, for example, how some families had produced generation after generation of judges. His analysis largely neglected to take into account the effect of the environment: the son of a prominent judge is, after all, rather more likely to become a judge—by virtue of his father's connections, if nothing else—than the son of a peasant farmer. Galton did not, however, completely overlook the effect of the environment, and it was he who first referred to the "nature/nurture" dichotomy, possibly in reference to Shakespeare's irredeemable villain, Caliban, "a devil, a born devil, on whose nature/Nurture can never stick."

The results of his analysis, however, left no doubt in Galton's mind. 40

> I have no patience with the hypothesis occasionally expressed, and often implied, especially in tales written to teach children to be good, that babies are born pretty much alike, and that the sole agencies in creating differences between boy and boy, and man and man, are steady application and moral effort. It is in the most unqualified manner that I object to pretensions of natural equality.

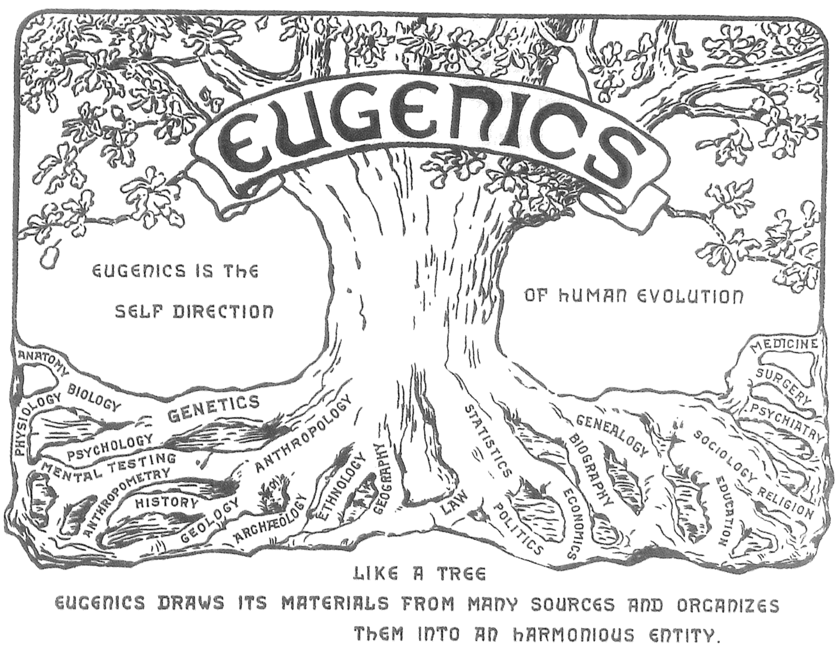
A corollary of his conviction that these traits are genetically determined, he 41
argued, was that it would be possible to "improve" the human stock by preferentially breeding gifted individuals, and preventing the less gifted from reproducing.

> It is easy . . . to obtain by careful selection a permanent breed of dogs or horses gifted with peculiar powers of running, or of doing anything else, so it would be quite practicable to produce a highly-gifted race of men by judicious marriages during several consecutive generations.

42 Galton introduced the terms *eugenics* (literally "good in birth") to describe this application of the basic principle of agricultural breeding to humans. In time, eugenics came to refer to "self-directed human evolution": by making conscious choices about who should have children, eugenicists believed that they could head off the "eugenic crisis" precipitated in the Victorian imagination by the high rates of reproduction of inferior stock coupled with the typically small families of the superior middle classes.

[. . .]

D N A



Eugenics as it was perceived during the first part of the twentieth century: an opportunity for humans to control their own evolutionary destiny

CHAPTER TWO
THE DOUBLE HELIX: THIS IS LIFE

I got hooked on the gene during my third year at the University of Chicago. Until then, I had planned to be a naturalist and looked forward to a career far removed from the urban bustle of Chicago's South Side, where I grew up. My change of heart was inspired not by an unforgettable teacher but a little book that appeared in 1944, *What Is Life?*, by the Austrian-born father of wave mechanics, Erwin Schrödinger. It grew out of several lectures he had given the year before at the Institute for Advanced Study in Dublin. That a great physicist had taken the time to write about biology caught my fancy. In those days, like most people, I considered chemistry and physics to be the "real" sciences, and theoretical physicists were science's top dogs. 1

Schrödinger argued that life could be thought of in terms of storing and passing on biological information. Chromosomes were thus simply information bearers. Because so much information had to be packed into every cell, it must be compressed into what Schrödinger called a "hereditary code-script" embedded in the molecular fabric of chromosomes. To understand life, then, we would have to identify these molecules, and crack their code. He even speculated that understanding life—which would involve finding the gene—might take us beyond the laws of physics as we then understood them. Schrödinger's book was tremendously influential. Many of those who would become major players in Act 1 of molecular biology's great drama, including Francis Crick (a former physicist himself), had, like me, read *What Is Life?* and been impressed. 2

In my own case, Schrödinger struck a chord because I too was intrigued by the essence of life. A small minority of scientists still thought life depended upon a vital force emanating from an all-powerful god. But like most of my teachers, I disdained the very idea of vitalism. If such a "vital" force were calling the shots in nature's game, there was little hope life would ever be understood through the methods of science. On the other hand, the notion that life might be perpetuated by means of an instruction book inscribed in a secret code appealed to me. What sort of molecular code could be so elaborate as to convey all the multitudinous wonder of the living world? And what sort of molecular trick could ensure that the code is exactly copied every time a chromosome duplicates? 3

4 At the time of Schrödinger's Dublin lectures, most biologists supposed that proteins would eventually be identified as the primary bearers of genetic instruction. Proteins are molecular chains built up from twenty different building blocks, the amino acids. Because permutations in the order of amino acids along the chain are virtually infinite, proteins could, in principle, readily encode the information underpinning life's extraordinary diversity. DNA then was not considered a serious candidate for the bearer of code-scripts, even though it was exclusively located on chromosomes and had been known about for some seventy-five years. In 1869, Friedrich Miescher, a Swiss biochemist working in Germany, had isolated from pus-soaked bandages supplied by a local hospital a substance he called "nuclein." Because pus consists largely of white blood cells, which, unlike red blood cells, have nuclei and therefore DNA-containing chromosomes, Miescher had stumbled on a good source of DNA. When he later discovered that "nuclein" was to be found in chromosomes alone, Miescher understood that his discovery was indeed a big one. In 1893, he wrote: "Inheritance insures a continuity in form from generation to generation that lies even deeper than the chemical molecule. It lies in the structuring atomic groups. In this sense, I am a supporter of the chemical heredity theory."

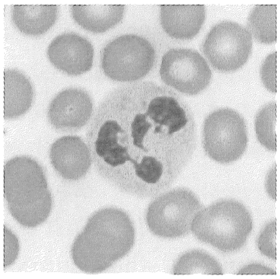
The physicist Erwin Schrödinger, whose book What Is Life? turned me on to the gene

5 Nevertheless, for decades afterward, chemistry would remain unequal to the task of analyzing the immense size and complexity of the DNA molecule. Only

in the 1930s was DNA shown to be a long molecule containing four different chemical bases: adenine (A), guanine (G), thymine (T), and cytosine (C). But at the time of Schrödinger's lectures, it was still unclear just how the subunits (called deoxynucleotides) of the molecule were chemically linked. Nor was it known whether DNA molecules might vary in their sequences of the four different bases. If DNA were indeed Schrödinger's code-script, then the molecule would have to be capable of existing in an immense number of different forms. But back then it was still considered a possibility that one simple sequence like AGTC might be repeated over and over along the entire length of DNA chains.

DNA did not move into the genetic limelight until 1944, when Oswald Avery's 6
lab at the Rockefeller Institute in New York City reported that the composition of the surface coats of pneumonia bacteria could be changed. This was not the result he and his junior colleagues, Colin MacLeod and Maclyn McCarty, expected.

For more than a decade Avery's group had been following up on another most 7
unexpected observation made in 1928 by Fred Griffith, a scientist in the British Ministry of Health. Griffith was interested in pneumonia and studied its bacterial agent, *Pneumococcus*. It was known that there were two strains, designated "smooth" (S) and "rough" (R) according to their appearance under the microscope. These strains differed not only visually but also in their virulence. Inject S bacteria into a mouse, and within a few days the mouse dies; inject R bacteria and the mouse remains healthy. It turns out that S bacterial cells have a coating that prevents the mouse's immune system from recognizing the invader. The R cells have no such coating and are therefore readily attacked by the mouse's immune defenses.



A view through the microscope of blood cells treated with a chemical that stains DNA. In order to maximize their oxygen-transporting capacity, red blood cells have no nucleus and therefore no DNA. But white blood cells, which patrol the bloodstream in search of intruders, have a nucleus containing chromosomes.

8 Through his involvement with public health, Griffith knew that multiple strains had sometimes been isolated from a single patient, and so he was curious about how different strains might interact in his unfortunate mice. With one combination, he made a remarkable discovery: when he injected heatkilled S bacteria (harmless) *and* normal R bacteria (also harmless), the mouse died. How could two harmless forms of bacteria conspire to become lethal? The clue came when he isolated the *Pneumococcus* bacteria retrieved from the dead mice and discovered living S bacteria. It appeared the living innocuous R bacteria had acquired something from the dead S variant; whatever it was, that something had allowed the R in the presence of the heat-killed S bacteria to transform itself into a living killer S strain. Griffith confirmed that this change was for real by culturing the S bacteria from the dead mouse over several generations: the bacteria bred true for the S type, just as any regular S strain would. A *genetic* change had indeed occurred to the R bacteria injected into the mouse.

9 Though this transformation phenomenon seemed to defy all understanding, Griffith's observations at first created little stir in the scientific world. This was partly because Griffith was intensely private and so averse to large gatherings that he seldom attended scientific conferences. Once, he had to be virtually forced to give a lecture. Bundled into a taxi and escorted to the hall by colleagues, he discoursed in a mumbled monotone, emphasizing an obscure corner of his microbiological work but making no mention of bacterial transformation. Luckily, however, not everyone overlooked Griffith's breakthrough.

10 Oswald Avery was also interested in the sugarlike coats of the *Pneumococcus*. He set out to duplicate Griffith's experiment in order to isolate and characterize whatever it was that had caused those R cells to change to the S type. In 1944 Avery, MacLeod, and McCarty published their results: an exquisite set of experiments showing unequivocally that DNA was the transforming principle. Culturing the bacteria in the test tube rather than in mice made it much easier to search for the chemical identity of the transforming factor in the heat-killed S cells. Methodically destroying one by one the biochemical components of the heat-treated S cells, Avery and his group looked to see whether transformation was prevented. First they degraded the sugarlike coat of the S bacteria. Transformation still occurred: the coat was not the transforming principle. Next they used a mixture of two protein-destroying enzymes, trypsin and chymotrypsin, to degrade virtually all the proteins

in the S cells. To their surprise, transformation was again unaffected. Next they tried an enzyme (RNase) that breaks down RNA (ribonucleic acid), a second class of nucleic acids similar to DNA and possibly involved in protein synthesis. Again transformation occurred. Finally, they came to DNA, exposing the S bacterial extracts to the DNA-destroying enzyme, DNase. This time they hit a home run. All S-inducing activity ceased completely. The transforming factor was DNA.

In part because of its bombshell implications, the resulting February 1944 11
paper by Avery, MacLeod, and McCarty met with a mixed response. Many geneticists accepted their conclusions. After all, DNA was found on every chromosome; why shouldn't it be the genetic material? By contrast, however, most biochemists expressed doubt that DNA was a complex enough molecule to act as the repository of such a vast quantity of biological information. They continued to believe that proteins, the other component of chromosomes, would prove to be the hereditary substance. In principle, as the biochemists rightly noted, it would be much easier to encode a vast body of complex information using the twenty-letter amino-acid alphabet of proteins than the four-letter nucleotide alphabet of DNA. Particularly vitriolic in his rejection of DNA as the genetic substance was Avery's own colleague at the Rockefeller Institute, the protein chemist Alfred Mirsky. By then, however, Avery was no longer scientifically active. The Rockefeller Institute had mandatorily retired him at age sixty-five.

Avery missed out on more than the opportunity to defend his work against 12
the attacks of his colleagues: He was never awarded the Nobel Prize, which was certainly his due, for identifying DNA as the transforming principle. Because the Nobel committee makes its records public fifty years following each award, we now know that Avery's candidacy was blocked by the Swedish physical chemist Einar Hammarsten. Though Hammarsten's reputation was based largely on his having produced DNA samples of unprecedented high quality, he still believed genes to be an undiscovered class of proteins. In fact, even after the double helix was found, Hammarsten continued to insist that Avery should not receive the prize until after the mechanism of DNA transformation had been completely worked out. Avery died in 1955; had he lived only a few more years, he would almost certainly have gotten the prize.

When I arrived at Indiana University in the fall of 1947 with plans to pursue 13
the gene for my Ph.D. thesis, Avery's paper came up over and over in conversations.

By then, no one doubted the reproducibility of his results, and more recent work coming out of the Rockefeller Institute made it all the less likely that proteins would prove to be the genetic actors in bacterial transformation. DNA had at last become an important objective for chemists setting their sights on the next breakthrough. In Cambridge, England, the canny Scottish chemist Alexander Todd rose to the challenge of identifying the chemical bonds that linked together nucleotides in DNA. By early 1951, his lab had proved that these links were always the same, such that the backbone of the DNA molecule was very regular. During the same period, the Austrian-born refugee Erwin Chargaff, at the College of Physicians and Surgeons of Columbia University, used the new technique of paper chromatography to measure the relative amounts of the four DNA bases in DNA samples extracted from a variety of vertebrates and bacteria. While some species had DNA in which adenine and thymine predominated, others had DNA with more guanine and cytosine. The possibility thus presented itself that no two DNA molecules had the same composition.

14 At Indiana I joined a small group of visionary scientists, mostly physicists and chemists, studying the reproductive process of the viruses that attack bacteria (bacteriophages—"phages" for short). The Phage Group was born when my Ph.D. supervisor, the Italian-trained medic Salvador Luria and his close friend, the German-born theoretical physicist Max Delbrück, teamed up with the American physical chemist Alfred Hershey. During World War II both Luria and Delbrück were considered enemy aliens, and thus ineligible to serve in the war effort of American science, even though Luria, a Jew, had been forced to leave France for New York City and Delbrück had fled Germany as an objector to Nazism. Thus excluded, they continued to work in their respective university labs—Luria at Indiana and Delbrück at Vanderbilt—and collaborated on phage experiments during successive summers at Cold Spring Harbor. In 1943, they joined forces with the brilliant but taciturn Hershey, then doing phage research of his own at Washington University in St. Louis.

15 The Phage Group's program was based on its belief that phages, like all viruses, were in effect naked genes. This concept had first been proposed in 1922 by the imaginative American geneticist Herman J. Muller, who three years later demonstrated that X rays cause mutations. His belated Nobel Prize came in 1946, just after he joined the faculty of Indiana University. It was his presence, in fact, that

led me to Indiana. Having started his career under T.H. Morgan, Muller knew better than anyone else how genetics had evolved during the first half of the twentieth century, and I was enthralled by his lectures during my first term. His work on fruit flies (*Drosophila*), however, seemed to me to belong more to the past than to the future, and I only briefly considered doing thesis research under his supervision. I opted instead for Luria's phages, an even speedier experimental subject than *Drosophila*: genetic crosses of phages done one day could be analyzed the next.

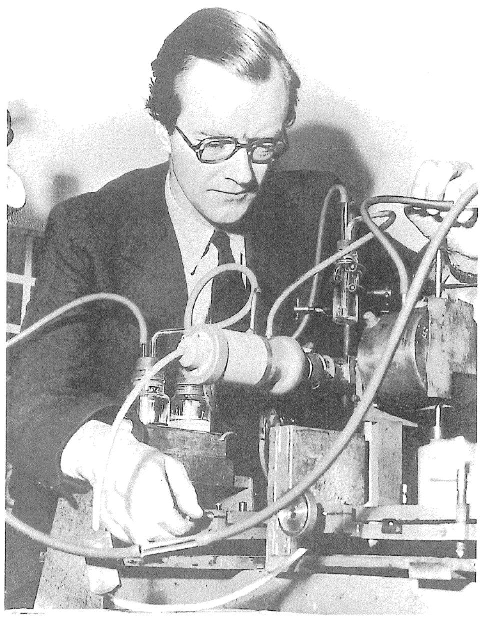
For my Ph.D. thesis research, Luria had me follow in his footsteps by studying how X rays killed phage particles. Initially I had hoped to show that viral death was caused by damage to phage DNA. Reluctantly, however, I eventually had to concede that my experimental approach could never give unambiguous answers at the chemical level. I could draw only biological conclusions. Even though phages were indeed effectively naked genes, I realized that the deep answers the Phage Group was seeking could be arrived at only through advanced chemistry. DNA somehow had to transcend its status as an acronym; it had to be understood as a molecular structure in all its chemical detail. 16

Upon finishing my thesis, I saw no alternative but to move to a lab where I could study DNA chemistry. Unfortunately, however, knowing almost no pure chemistry, I would have been out of my depth in any lab attempting difficult experiments in organic or physical chemistry. I therefore took a postdoctoral fellowship in the Copenhagen lab of the biochemist Herman Kalckar in the fall of 1950. He was studying the synthesis of the small molecules that make up DNA, but I figured out quickly that his biochemical approach would never lead to an understanding of the essence of the gene. Every day spent in his lab would be one more day's delay in learning how DNA carried genetic information. 17

My Copenhagen year nonetheless ended productively. To escape the cold Danish spring, I went to the Zoological Station at Naples during April and May. During my last week there, I attended a small conference on X-ray diffraction methods for determining the 3-D structure of molecules. X-ray diffraction is a way of studying the atomic structure of any molecule that can be crystallized. The crystal is bombarded with X rays, which bounce off its atoms and are scattered. The scatter pattern gives information about the structure of the molecule but, taken alone, is not enough to solve the structure. The additional information needed is the "phase assignment," which deals with the wave properties of the molecule. Solving 18 the phase problem was not easy, and at that time only the most audacious scientists were willing to take it on. Most of the successes of the diffraction method had been achieved with relatively simple molecules.

19 My expectations for the conference were low. I believed that a three-dimensional understanding of protein structure, or for that matter of DNA, was more than a decade away. Disappointing earlier X-ray photos suggested that DNA was particularly unlikely to yield up its secrets via the X-ray approach. These results were not surprising since the exact sequences of DNA were expected to differ from one individual molecule to another. The resulting irregularity of surface configurations would understandably prevent the long thin DNA chains from lying neatly side by side in the regular repeating patterns required for X-ray analysis to be successful.

20 It was therefore a surprise and a delight to hear the last-minute talk on DNA by a thirty-four-year-old Englishman named Maurice Wilkins from the Biophysics Lab of King's College, London. Wilkins was a physicist who during the war had worked on the Manhattan Project. For him, as for many of the other scientists involved, the actual deployment of the bomb on Hiroshima and Nagasaki, supposedly the culmination of all their work, was profoundly disillusioning.



Maurice Wilkins in his lab at King's College, London

He considered forsaking science altogether to become a painter in Paris, but biology intervened. He too had read Schrödinger's book, and was now tackling DNA with X-ray diffraction. 21

He displayed a photograph of an X-ray diffraction pattern he had recently obtained, and its many precise reflections indicated a highly regular crystalline packing. DNA, one had to conclude, must have a regular structure, the elucidation of which might well reveal the nature of the gene. Instantly I saw myself moving to London to help Wilkins find the structure. My attempts to converse with him after his talk, however, went nowhere. All I got for my efforts was a declaration of his conviction that much hard work lay ahead. 22

While I was hitting consecutive dead ends, back in America the world's preeminent chemist, Caltech's Linus Pauling, announced a major triumph: he had found the exact arrangement in which chains of amino acids (called *polypeptides*) fold up in proteins, and called his structure the α-helix (alpha helix). That it was Pauling who made this breakthrough was no surprise: he was a scientific superstar. His book *The Nature of the Chemical Bond* essentially laid the foundation of modern chemistry, and, for chemists of the day, it was the Bible. Pauling had been a precocious child. When he was nine, his father, a druggist in Oregon, wrote to the *Oregonian* newspaper requesting suggestions of reading matter for his bookish son, adding that he had already read the Bible and Darwin's *Origin of Species*. But the early death of Pauling's father, which brought the family to financial ruin, makes it remarkable that the promising young man managed to get an education at all. 23

As soon as I returned to Copenhagen I read about Pauling's α-helix. To my surprise, his model was not based on a deductive leap from experimental X-ray diffraction data. Instead, it was Pauling's long experience as a structural chemist that had emboldened him to infer which type of helical fold would be most compatible with the underlying chemical features of the polypeptide chain. Pauling made scale models of the different parts of the protein molecule, working out plausible schemes in three dimensions. He had reduced the problem to a kind of three-dimensional jigsaw puzzle in a way that was simple yet brilliant. 24

Whether the α-helix was correct—in addition to being pretty—was now the question. Only a week later, I got the answer. Sir Lawrence Bragg, the English inventor of X-ray crystallography and 1915 Nobel laureate in Physics, came to Copenhagen and excitedly reported that his junior colleague, the Austrian-born 25 chemist Max Perutz, had ingeniously used synthetic polypeptides to confirm the correctness of Pauling's α-helix. It was a bittersweet triumph for Bragg's Cavendish Laboratory. The year before, they had completely missed the boat in their paper outlining possible helical folds for polypeptide chains.

Lawrence Bragg (left) with Linus Pauling, who is carrying a model of the α-helix

26 By then Salvador Luria had tentatively arranged for me to take up a research position at the Cavendish. Located at Cambridge University, this was the most famous laboratory in all of science. Here Ernest Rutherford first described the structure of the atom. Now it was Bragg's own domain, and I was to work as apprentice to the English chemist John Kendrew, who was interested in determining the 3-D structure of the protein myoglobin. Luria advised me to visit the Cavendish as soon as possible. With Kendrew in the States, Max Perutz would check me out. Together, Kendrew and Perutz had earlier established the Medical Research Council (MRC) Unit for the Study of the Structure of Biological Systems.

27 A month later in Cambridge, Perutz assured me that I could quickly master the necessary X-ray diffraction theory and should have no difficulty fitting in with the others in their tiny MRC Unit. To my relief, he was not put off by my biology background. Nor was Lawrence Bragg, who briefly came down from his office to look me over.

28 I was twenty-three when I arrived back at the MRC Unit in Cambridge in early October. I found myself sharing space in the biochemistry room with a thirty-five-year-old ex-physicist, Francis Crick, who had spent the war working

on magnetic mines for the Admiralty. When the war ended, Crick had planned to stay on in military research, but, on reading Schrödinger's *What Is Life?*, he had moved toward biology. Now he was at the Cavendish to pursue the 3-D structure of proteins for his Ph.D.

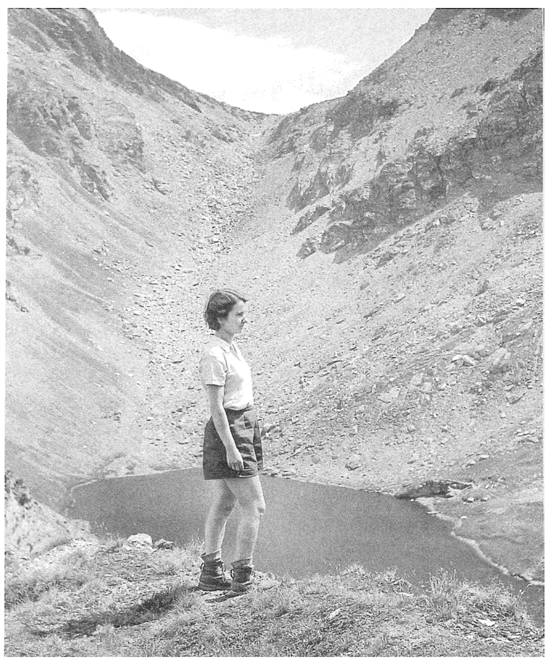
Crick was always fascinated by the intricacies of important problems. His 29
endless questions as a child compelled his weary parents to buy him a children's encyclopedia, hoping that it would satisfy his curiosity. But it only made him insecure: he confided to his mother his fear that everything would have been discovered by the time he grew up, leaving him nothing to do. His mother reassured him (correctly, as it happened) that there would still be a thing or two for him to figure out.

A great talker, Crick was invariably the center of attention in any gathering. 30
His booming laugh was forever echoing down the hallways of the Cavendish. As the MRC Unit's resident theoretician, he used to come up with a novel insight at least once a month, and he would explain his latest idea at great length to anyone willing to listen. The morning we met he lit up when he learned that my objective in coming to Cambridge was to learn enough crystallography to have a go at the DNA structure. Soon I was asking Crick's opinion about using Pauling's model-building approach to go directly for the structure. Would we need many more years of diffraction experimentation before modeling would be practicable? To bring us up to speed on the status of DNA structural studies, Crick invited Maurice Wilkins, a friend since the end of the war, up from London for Sunday lunch. Then we could learn what progress Wilkins had made since his talk in Naples.

Francis Crick with the Cavendish X-ray tube

31 Wilkins expressed his belief that DNA's structure was a helix, formed by several chains of linked nucleotides twisted around each other. All that remained to be settled was the number of chains. At the time, Wilkins favored three on the basis of his density measurements of DNA fibers. He was keen to start model-building, but he had run into a roadblock in the form of a new addition to the King's College Biophysics Unit, Rosalind Franklin.

32 A thirty-one-year-old Cambridge-trained physical chemist, Franklin was an obsessively professional scientist; for her twenty-ninth birthday all she requested was her own subscription to her field's technical journal, *Acta Crystallographica*. Logical and precise, she was impatient with those who acted otherwise. And she was given to strong opinions, once describing her Ph.D. thesis adviser, Ronald Norrish, a future Nobel Laureate, as "stupid, bigoted, deceitful, ill-mannered and tyrannical." Outside the laboratory, she was a determined and gutsy mountaineer, and, coming from the upper echelons of London society, she belonged to a more rarefied social world than most scientists. At the end of a hard day at the bench, she would occasionally change out of her lab coat into an elegant evening gown and disappear into the night.



Rosalind Franklin on one of the mountain hiking vacations she loved

Just back from a four-year X-ray crystallographic investigation of graphite 33
in Paris, Franklin had been assigned to the DNA project while Wilkins was away from King's. Unfortunately, the pair soon proved incompatible. Franklin, direct and data-focused, and Wilkins, retiring and speculative, were destined never to collaborate. Shortly before Wilkins accepted our lunch invitation, the two had had a big blowup in which Franklin had insisted that no model-building could commence before she collected much more extensive diffraction data. Now they effectively didn't communicate, and Wilkins would have no chance to learn of her progress until Franklin presented her lab seminar scheduled for the beginning of November. If we wanted to listen, Crick and I were welcome to go as Wilkins's guests.

Crick was unable to make the seminar, so I attended alone and briefed him 34
later on what I believed to be its key take-home messages on crystalline DNA. In particular, I described from memory Franklin's measurements of the crystallographic repeats and the water content. This prompted Crick to begin sketching helical grids on a sheet of paper, explaining that the new helical X-ray theory he had devised with Bill Cochran and Vladimir Vand would permit even me, a former bird-watcher, to predict correctly the diffraction patterns expected from the molecular models we would soon be building at the Cavendish.

As soon as we got back to Cambridge, I arranged for the Cavendish machine 35
shop to construct the phosphorous atom models needed for short sections of the sugar phosphate backbone found in DNA. Once these became available, we tested different ways the backbones might twist around each other in the center of the DNA molecule. Their regular repeating atomic structure should allow the atoms to come together in a consistent, repeated conformation. Following Wilkins's hunch, we focused on three-chain models. When one of these appeared to be almost plausible, Crick made a phone call to Wilkins to announce we had a model we thought might be DNA.

The next day both Wilkins and Franklin came up to see what we had done. 36
The threat of unanticipated competition briefly united them in common purpose. Franklin wasted no time in faulting our basic concept. My memory was that she had reported almost no water present in crystalline DNA. In fact, the opposite was true. Being a crystallographic novice, I had confused the terms "unit cell" and "asymmetric unit." Crystalline DNA was in fact water-rich. Consequently, Franklin pointed out, the backbone had to be on the outside and not, as we had it, in the

center, if only to accommodate all the water molecules she had observed in her crystals.

37 That unfortunate November day cast a very long shadow. Franklin's opposition to model-building was reinforced. Doing experiments, not playing with Tinkertoy representations of atoms, was the way she intended to proceed. Even worse, Sir Lawrence Bragg passed down the word that Crick and I should desist from all further attempts at building a DNA model. It was further decreed that DNA research should be left to the King's lab, with Cambridge continuing to focus solely on proteins. There was no sense in two MRC-funded labs competing against each other. With no more bright ideas up our sleeves, Crick and I were reluctantly forced to back off, at least for the time being.

38 It was not a good moment to be condemned to the DNA sidelines. Linus Pauling had written Wilkins to request a copy of the crystalline DNA diffraction pattern. Though Wilkins had declined, saying he wanted more time to interpret it himself, Pauling was hardly obliged to depend upon data from King's. If he wished, he could easily start serious X-ray diffraction studies at Caltech.

39 The following spring, I duly turned away from DNA and set about extending prewar studies on the pencil-shaped tobacco mosaic virus using the Cavendish's powerful new X-ray beam. This light experimental workload gave me plenty of time to wander through various Cambridge libraries. In the zoology building, I read Erwin Chargaff's paper describing his finding that the DNA bases adenine and thymine occurred in roughly equal amounts, as did the bases guanine and cytosine. Hearing of these one-to-one ratios Crick wondered whether, during DNA duplication, adenine residues might be attracted to thymine and vice versa, and whether a corresponding attraction might exist between guanine and cytosine. If so, base sequences on the "parental" chains (e.g., ATGC) would have to be complementary to those on "daughter" strands (yielding in this case TACG).

40 These remained idle thoughts until Erwin Chargaff came through Cambridge in the summer of 1952 on his way to the International Biochemical Congress in Paris. Chargaff expressed annoyance that neither Crick nor I saw the need to know the chemical structures of the four bases. He was even more upset when we told him that we could simply look up the structures in textbooks as the need arose. I was left hoping that Chargaff's data would prove irrelevant. Crick, however, was energized to do several experiments looking for molecular "sandwiches" that might

form when adenine and thymine (or alternatively, guanine and cytosine) were mixed together in solution. But his experiments went nowhere.

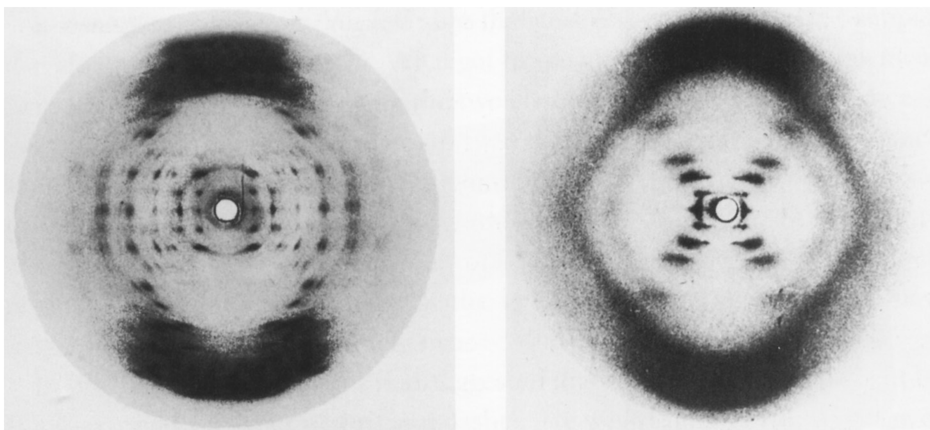
Like Chargaff, Linus Pauling also attended the International Biochemical 41
Congress, where the big news was the latest result from the Phage Group. Alfred Hershey and Martha Chase at Cold Spring Harbor had just confirmed Avery's transforming principle: DNA was the hereditary material! Hershey and Chase proved that only the DNA of the phage virus enters bacterial cells; its protein coat remains on the outside. It was more obvious than ever that DNA must be understood at the molecular level if we were to uncover the essence of the gene. With Hershey and Chase's result the talk of the town, I was sure that Pauling would now bring his formidable intellect and chemical wisdom to bear on the problem of DNA.

Early in 1953, Pauling did indeed publish a paper outlining the structure 42
of DNA. Reading it anxiously I saw that he was proposing a three-chain model with sugar phosphate backbones forming a dense central core. Superficially it was similar to our botched model of fifteen months earlier. But instead of using positively charged atoms (e.g., Mg^{2+}) to stabilize the negatively charged backbones, Pauling made the unorthodox suggestion that the phosphates were held together by hydrogen bonds. But it seemed to me, the biologist, that such hydrogen bonds required extremely acidic conditions never found in cells. With a mad dash to Alexander Todd's nearby organic chemistry lab my belief was confirmed: The impossible had happened. The world's best-known, if not best, chemist had gotten his chemistry wrong. In effect, Pauling had knocked the A off of DNA. Our quarry was deoxyribonucleic acid, but the structure he was proposing was not even acidic.

Hurriedly I took the manuscript to London to inform Wilkins and Franklin 43
they were still in the game. Convinced that DNA was not a helix, Franklin had no wish even to read the article and deal with the distraction of Pauling's helical ideas, even when I offered Crick's arguments for helices. Wilkins, however, was very interested indeed in the news I brought; he was now more certain than ever that DNA was helical. To prove the point, he showed me a photograph obtained more than six months earlier by Franklin's graduate student Raymond Gosling, who had X-rayed the so-called B form of DNA. Until that moment, I didn't know a B form even existed. Franklin had put this picture aside, preferring to concentrate on the A form, which she thought would more likely yield useful data. The X-ray pattern of this B form was a distinct cross. Since Crick and others had already deduced that

such a pattern of reflections would be created by a helix, this evidence made it clear that DNA had to be a helix! In fact, despite Franklin's reservations, this was no surprise. Geometry itself suggested that a helix was the most logical arrangement for a long string of repeating units such as the nucleotides of DNA. But we still did not know what that helix looked like, nor how many chains it contained.

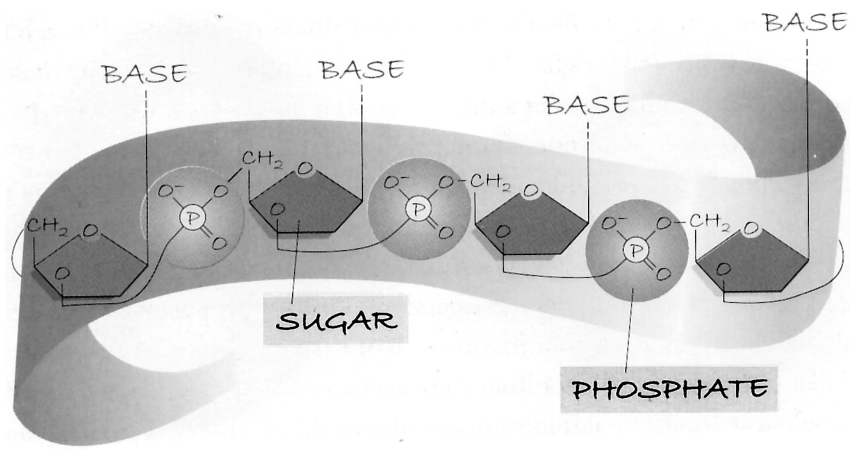
44 The time had come to resume building helical models of DNA. Pauling was bound to realize soon enough that his brainchild was wrong. I urged Wilkins to waste no time. But he wanted to wait until Franklin had completed her scheduled departure for another lab later that spring. She had decided to move on to avoid the unpleasantness at King's. Before leaving, she had been ordered to stop further work with DNA and had already passed on many of her diffraction images to Wilkins.



X-ray photos of the A and B forms of DNA from, respectively, Maurice Wilkins and Rosalind Franklin. The differences in molecular structure are caused by differences in the amount of water associated with each DNA molecule.

45 When I returned to Cambridge and broke the news of the DNA B form, Bragg no longer saw any reason for Crick and me to avoid DNA. He very much wanted the DNA structure to be found on his side of the Atlantic. So we went back to model-building, looking for a way the known basic components of DNA—the backbone of the molecule and the four different bases, adenine, thymine, guanine, and cytosine—could fit together to make a helix. I commissioned the shop at the Cavendish to make us a set of tin bases, but they couldn't produce them fast enough for me: I ended up cutting out rough approximations from stiff cardboard.

By this time I realized the DNA density-measurement evidence actually 46
slightly favored a two-chain, rather than three-chain, model. So I decided to search out plausible double helices. As a biologist, I preferred the idea of a genetic molecule made of two, rather than three, components. After all, chromosomes, like cells, increase in number by duplicating, not triplicating.



The chemical backbone of DNA

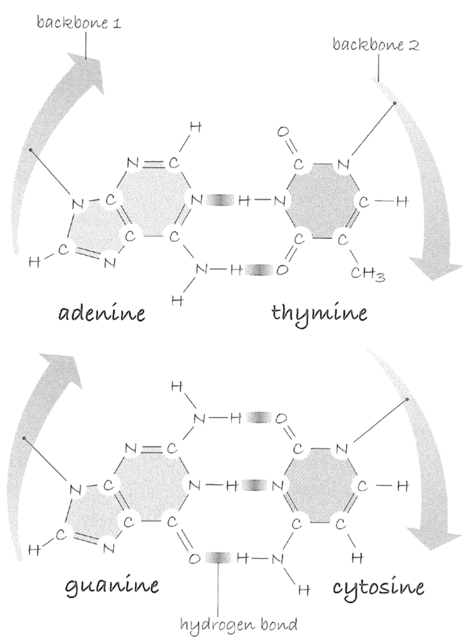
I knew that our previous model with the backbone on the inside and the bases 47
hanging out was wrong. Chemical evidence from the University of Nottingham, which I had too long ignored, indicated that the bases must be hydrogen-bonded to each other. They could only form bonds like this in the regular manner implied by the X-ray diffraction data if they were in the center of the molecule. But how could they come together in pairs? For two weeks I got nowhere, misled by an error in my nucleic acid chemistry textbook. Happily, on February 27, Jerry Donahue, a theoretical chemist visiting the Cavendish from Caltech, pointed out that the textbook was wrong. So I changed the locations of the hydrogen atoms on my cardboard cutouts of the molecules.

The next morning, February 28, 1953, the key features of the DNA model all 48
fell into place. The two chains were held together by strong hydrogen bonds between adenine-thymine and guanine-cytosine base pairs. The inferences Crick had drawn the year before based on Chargaff's research had indeed been correct. Adenine does bond to thymine and guanine does bond to cytosine, but not through flat surfaces to form molecular sandwiches. When Crick arrived, he took it all in rapidly, and gave

my base-pairing scheme his blessing. He realized right away that it would result in the two strands of the double helix running in opposite directions.

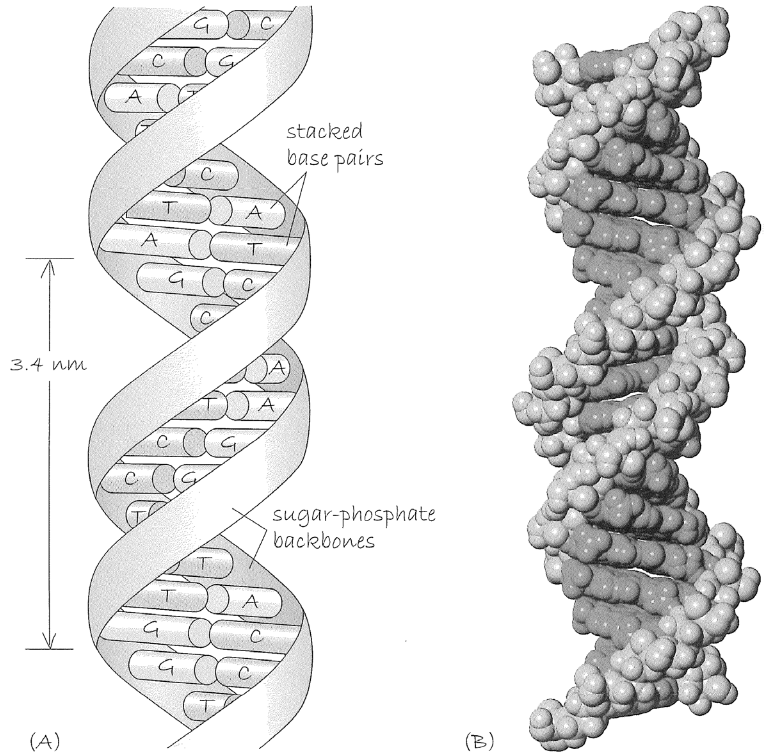
49 It was quite a moment. We felt sure that this was it. Anything that simple, that elegant just had to be right. What got us most excited was the complementarity of the base sequences along the two chains. If you knew the sequence—the order of bases—along one chain, you automatically knew the sequence along the other. It was immediately apparent that this must be how the genetic messages of genes are copied so exactly when chromosomes duplicate prior to cell division. The molecule would "unzip" to form two separate strands. Each separate strand then could serve as the template for the synthesis of a new strand, one double helix becoming two.

50 In *What is Life?* Schrödinger had suggested that the language of life might be like Morse code, a series of dots and dashes. He wasn't far off. The language of DNA is a linear series of As, Ts, Gs, and Cs. And just as transcribing a page out of a book can result in the odd typo, the rare mistake creeps in when all these As, Ts, Gs, and Cs are being copied along a chromosome. These errors are the mutations geneticists had talked about for almost fifty years. Change an "i" to an "a" and "Jim" becomes "Jam" in English; change a T to a C and "ATG" becomes "ACG" in DNA.



The insight that made it all come together: complementary pairing of the bases

DNA



Bases and backbone in place: the double helix. (A) is a schematic showing the system of base-pairing that binds the two strands together. (B) is a "spacefilling" model showing, to scale, the atomic detail of the molecule.

51 The double helix made sense chemically and it made sense biologically. Now there was no need to be concerned about Schrödinger's suggestion that new laws of physics might be necessary for an understanding of how the hereditary code-script is duplicated: genes in fact were no different from the rest of chemistry. Later that day, during lunch at the Eagle, the pub virtually adjacent to the Cavendish Lab, Crick, ever the talker, could not help but tell everyone we had just found the "secret of life." I myself, though no less electrified by the thought, would have waited until we had a pretty three-dimensional model to show off.

52 Among the first to see our demonstration model was the chemist Alexander Todd. That the nature of the gene was so simple both surprised and pleased him.

Later, however, he must have asked himself why his own lab, having established the general chemical structure of DNA chains, had not moved on to asking how the chains folded up in three dimensions. Instead the essence of the molecule was left to be discovered by a two-man team, a biologist and a physicist, neither of whom possessed a detailed command even of undergraduate chemistry. But paradoxically, this was, at least in part, the key to our success: Crick and I arrived at the double helix first precisely because most chemists at that time thought DNA too big a molecule to understand by chemical analysis.

53 At the same time, the only two chemists with the vision to seek DNA's 3-D structure made major tactical mistakes: Rosalind Franklin's was her resistance to model-building; Linus Pauling's was a matter of simply neglecting to read the existing literature on DNA, particularly the data on its base composition published by Chargaff. Ironically, Pauling and Chargaff sailed across the Atlantic on the same ship following the Paris Biochemical Congress in 1952, but failed to hit it off. Pauling was long accustomed to being right. And he believed there was no chemical problem he could not work out from first principles by himself. Usually this confidence was not misplaced. During the Cold War, as a prominent critic of the American nuclear weapons development program, he was questioned by the FBI after giving a talk. How did he know how much plutonium there is in an atomic bomb? Pauling's response was "Nobody told me. I figured it out."

54 Over the next several months Crick and (to a lesser extent) I relished showing off our model to an endless stream of curious scientists. However, the Cambridge biochemists did not invite us to give a formal talk in the biochemistry building. They started to refer to it as the "WC," punning our initials with those used in Britain for the toilet or water closet. That we had found the double helix without doing experiments irked them.

55 The manuscript that we submitted to *Nature* in early April was published just over three weeks later, on April 25, 1953. Accompanying it were two longer papers by Franklin and Wilkins, both supporting the general correctness of our model. In June, I gave the first presentation of our model at the Cold Spring Harbor symposium on viruses. Max Delbrück saw to it that I was offered, at the last minute, an invitation to speak. To this intellectually high-powered meeting I brought a three-dimensional model built in the Cavendish, the adeninethymine base pairs in red and the guanine-cytosine base pairs in green.

No. 4356 April 25, 1953 NATURE 737

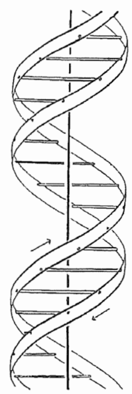
MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three inter-twined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

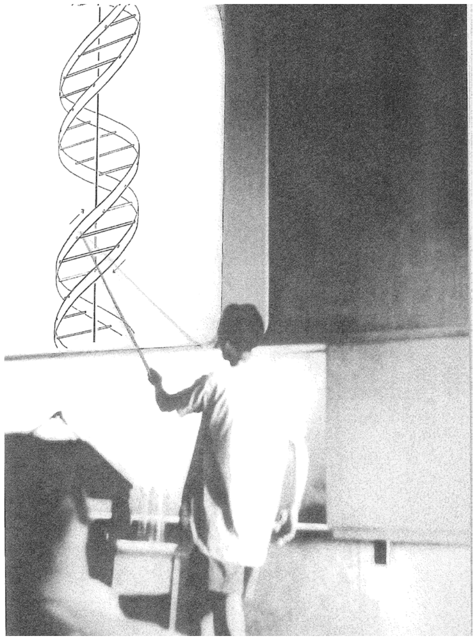
We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on inter-atomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. WATSON
F. H. C. CRICK

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,
Cavendish Laboratory, Cambridge.
April 2.

[1] Pauling, L., and Corey, R. B., *Nature*, 171, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, 39, 84 (1953).
[2] Furberg, S., *Acta Chem. Scand.*, 6, 634 (1952).
[3] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, 9, 402 (1952).
[4] Wyatt, G. R., *J. Gen. Physiol.*, 36, 201 (1952).
[5] Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).
[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, 10, 192 (1953).

Short and sweet: our Nature *paper announcing the discovery. The same issue also carried longer articles by Rosalind Franklin and Maurice Wilkins.*

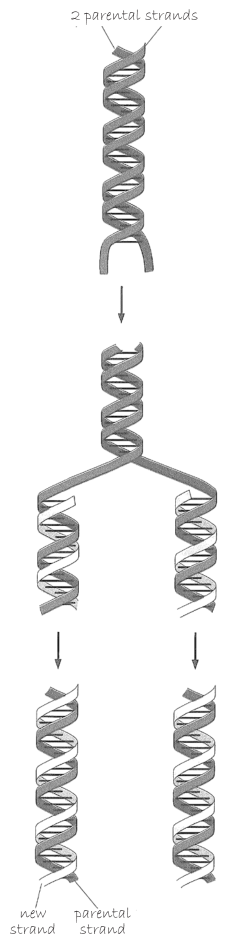


Unveiling the double helix: my lecture at Cold Spring Harbor Laboratory, June 1953

56 In the audience was Seymour Benzer, yet another ex-physicist who had heeded the clarion call of Schrödinger's book. He immediately understood what our breakthrough meant for his studies of mutations in viruses. He realized that he could now do for a short stretch of bacteriophage DNA what Morgan's boys had done forty years earlier for fruit fly chromosomes: he would map mutations—determine their order—along a gene, just as the fruit fly pioneers had mapped genes along a chromosome. Like Morgan, Benzer would have to depend on recombination to generate new genetic combinations, but, whereas Morgan had the advantage of a ready mechanism of recombination—the production of sex cells in a fruit fly—Benzer had to induce recombination by simultaneously infecting a single bacterial host cell with two different strains of bacteriophage, which differed by one or more mutations in the region of interest. Within the bacterial cell, recombination—the exchange of segments of molecules—would occasionally occur between the different viral DNA molecules, producing new permutations of mutations—so-called "recombinants." Within a single astonishingly productive year in his Purdue University lab, Benzer produced a map of a single bacteriophage gene, *rII*, showing

how a series of mutations—all errors in the genetic script—were laid out linearly along the virus DNA. The language was simple and linear, just like a line of text on the written page.

The response of the Hungarian physicist Leo Szilard to my Cold Spring 57
Harbor talk on the double helix was less academic. His question was, “Can you patent it?” At one time Szilard’s main source of income had been a patent that he held with Einstein, and he had later tried unsuccessfully to patent with Enrico Fermi the nuclear reactor they built at the University of Chicago in 1942. But then as now patents were given only for useful inventions and at the time no one could conceive of a practical use for DNA. Perhaps then, Szilard suggested, we should copyright it.

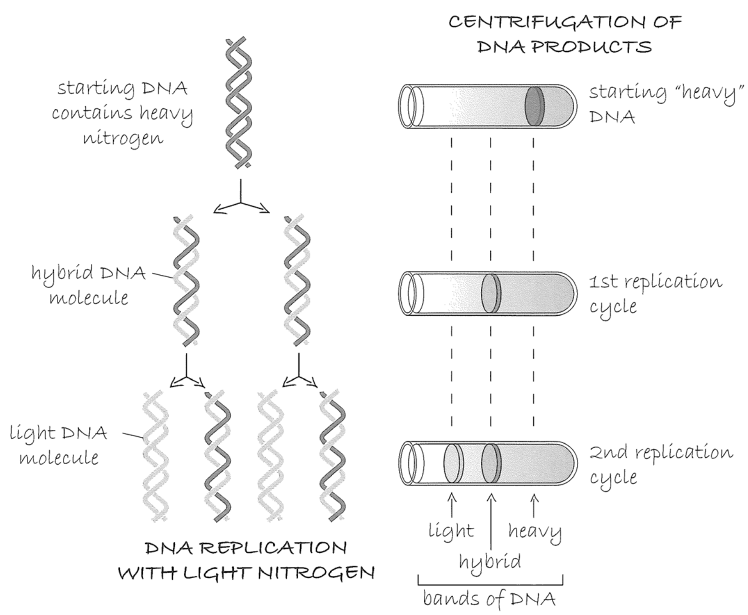


DNA replication: the double helix is unzipped and each strand copied.

58 There remained, however, a single missing piece in the double helical jigsaw puzzle: our unzipping idea for DNA replication had yet to be experimentally verified. Max Delbrück, for example, was unconvinced. Though he liked the double helix as a model, he worried that unzipping it might generate horrible knots. Five years later, a former student of Pauling's, Matt Meselson, and the equally bright young phage worker Frank Stahl put to rest such fears when they published the results of a single elegant experiment.

59 They had met in the summer of 1954 at the Marine Biological Laboratory at Woods Hole, Massachusetts, where I was then lecturing, and agreed—over a good many gin martinis—that they should get together to do some science. The result of their collaboration has been described as "the most beautiful experiment in biology."

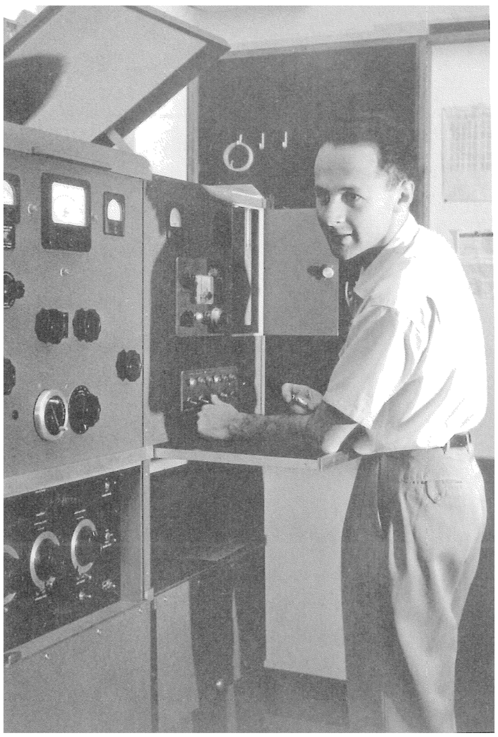
The Double Helix



The Meselson-Stahl experiment

60 They used a centrifugation technique that allowed them to sort molecules according to slight differences in weight; following a centrifugal spin, heavier molecules end up nearer the bottom of the test tube than lighter ones. Because

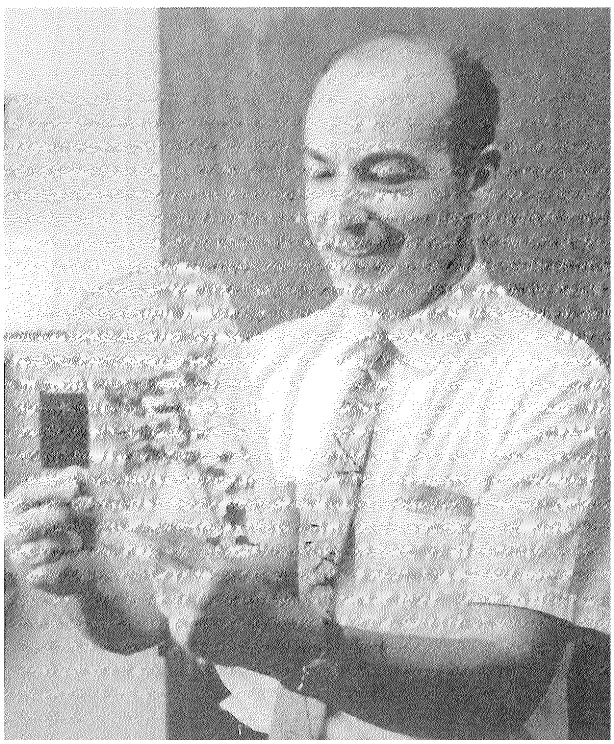
nitrogen atoms (N) are a component of DNA, and because they exist in two distinct forms, one light and one heavy, Meselson and Stahl were able to tag segments of DNA and thereby track the process of its replication in bacteria. Initially all the bacteria were raised in a medium containing heavy N, which was thus incorporated in both strands of the DNA. From this culture they took a sample, transferring it to a medium containing only light N, ensuring that the next round of DNA replication would have to make use of light N. If, as Crick and I had predicted, DNA replication involves unzipping the double helix and copying each strand, the resultant two "daughter" DNA molecules in the experiment would be hybrids, each consisting of one heavy N strand (the template strand derived from the "parent" molecule) and one light N strand (the one newly fabricated from the new medium). Meselson and Stahl's centrifugation procedure bore out these expectations precisely. They found three discrete bands in their centrifuge tubes, with the heavy-then-light sample halfway between the heavy-heavy and light-light samples. DNA replication works just as our model supposed it would.



Matt Meselson beside an ultra-centrifuge, the hardware at the heart of "the most beautiful experiment in biology"

61 The biochemical nuts and bolts of DNA replication were being analyzed at around the same time in Arthur Kornberg's laboratory at Washington University in St. Louis. By developing a new, "cell-free" system for DNA synthesis, Kornberg discovered an enzyme (DNA polymerase) that links the DNA components and makes the chemical bonds of the DNA backbone. Kornberg's enzymatic synthesis of DNA was such an unanticipated and important event that he was awarded the 1959 Nobel Prize in Physiology or Medicine, less than two years after the key experiments. After his prize was announced, Kornberg was photographed holding a copy of the double helix model I had taken to Cold Spring Harbor in 1953.

62 It was not until 1962 that Francis Crick, Maurice Wilkins, and I were to receive our own Nobel Prize in Physiology or Medicine. Four years earlier, Rosalind Franklin had died of ovarian cancer at the tragically young age of thirty-seven. Before then Crick had become a close colleague and a real friend of Franklin's. Following the two operations that would fail to stem the advance of her cancer, Franklin convalesced with Crick and his wife, Odile, in Cambridge.



Arthur Kornberg at the time of winning his Nobel Prize

It was and remains a long-standing rule of the Nobel Committee never to split 63
a single prize more than three ways. Had Franklin lived, the problem would have arisen whether to bestow the award upon her or Maurice Wilkins. The Swedes might have resolved the dilemma by awarding them both the Nobel Prize in Chemistry that year. Instead, it went to Max Perutz and John Kendrew, who had elucidated the three-dimensional structures of hemoglobin and myoglobin respectively.

The discovery of the double helix sounded the death knell for vitalism. Serious 64
scientists, even those religiously inclined, realized that a complete understanding of life would not require the revelation of new laws of nature. Life was just a matter of physics and chemistry, albeit exquisitely organized physics and chemistry. The immediate task ahead would be to figure out how the DNA-encoded script of life went about its work. How does the molecular machinery of cells read the messages of DNA molecules? As the next chapter will reveal, the unexpected complexity of the reading mechanism led to profound insights into how life first came about.

Text 6

from

Silent Spring

by Rachel Carson

CHAPTER 6
EARTH'S GREEN MANTLE

Water, soil, and the earth's green mantle of plants make up the world that supports the animal life of the earth. Although modern man seldom remembers the fact, he could not exist without the plants that harness the sun's energy and manufacture the basic foodstuffs he depends upon for life. Our attitude toward plants is a singularly narrow one. If we see any immediate utility in a plant we foster it. If for any reason we find its presence undesirable or merely a matter of indifference, we may condemn it to destruction forthwith. Besides the various plants that are poisonous to man or his livestock, or crowd out food plants, many are marked for destruction merely because, according to our narrow view, they happen to be in the wrong place at the wrong time. Many others are destroyed merely because they happen to be associates of the unwanted plants. 1

The earth's vegetation is part of a web of life in which there are intimate and essential relations between plants and the earth, between plants and other plants, between plants and animals. Sometimes we have no choice but to disturb these relationships, but we should do so thoughtfully, with full awareness that what we do may have consequences remote in time and place. But no such humility marks 2 the booming "weed killer" business of the present day, in which soaring sales and expanding uses mark the production of plant-killing chemicals.

3 One of the most tragic examples of our unthinking bludgeoning of the landscape is to be seen in the sagebrush lands of the West, where a vast campaign is on to destroy the sage and to substitute grasslands. If ever an enterprise needed to be illuminated with a sense of the history and meaning of the landscape, it is this. For here the natural landscape is eloquent of the interplay of forces that have created it. It is spread before us like the pages of an open book in which we can read why the land is what it is, and why we should preserve its integrity. But the pages lie unread.

4 The land of the sage is the land of the high western plains and the lower slopes of the mountains that rise above them, a land born of the great uplift of the Rocky Mountain system many millions of years ago. It is a place of harsh extremes of climate: of long winters when blizzards drive down from the mountains and snow lies deep on the plains, of summers whose heat is relieved by only scanty rains, with drought biting deep into the soil, and drying winds stealing moisture from leaf and stem.

5 As the landscape evolved, there must have been a long period of trial and error in which plants attempted the colonization of this high and windswept land. One after another must have failed. At last one group of plants evolved which combined all the qualities needed to survive. The sage—low-growing and shrubby—could hold its place on the mountain slopes and on the plains, and within its small gray leaves it could hold moisture enough to defy the thieving winds. It was no accident, but rather the result of long ages of experimentation by nature, that the great plains of the West became the land of the sage.

6 Along with the plants, animal life, too, was evolving in harmony with the searching requirements of the land. In time there were two as perfectly adjusted to their habitat as the sage. One was a mammal, the fleet and graceful pronghorn antelope. The other was a bird, the sage grouse—the "cock of the plains" of Lewis and Clark.

7 The sage and the grouse seem made for each other. The original range of the bird coincided with the range of the sage, and as the sagelands have been reduced, so the populations of grouse have dwindled. The sage is all things to these birds of the plains. The low sage of the foothill ranges shelters their nests and their young;

the denser growths are loafing and roosting areas; at all times the sage provides the staple food of the grouse. Yet it is a two-way relationship. The spectacular courtship displays of the cocks help loosen the soil beneath and around the sage, aiding invasion by grasses which grow in the shelter of sagebrush.

The antelope, too, have adjusted their lives to the sage. They are primarily animals of the plains, and in winter when the first snows come those that have summered in the mountains move down to the lower elevations. There the sage provides the food that tides them over the winter. Where all other plants have shed their leaves, the sage remains evergreen, the gray-green leaves—bitter, aromatic, rich in proteins, fats, and needed minerals—clinging to the stems of the dense and shrubby plants. Though the snows pile up, the tops of the sage remain exposed, or can be reached by the sharp, pawing hoofs of the antelope. Then grouse feed on them too, finding them on bare and windswept ledges or following the antelope to feed where they have scratched away the snow. 8

And other life looks to the sage. Mule deer often feed on it. Sage may mean survival for winter-grazing livestock. Sheep graze many winter ranges where the big sagebrush forms almost pure stands. For half the year it is their principal forage, a plant of higher energy value than even alfalfa hay. 9

The bitter upland plains, the purple wastes of sage, the wild, swift antelope, and the grouse are then a natural system in perfect balance. Are? The verb must be changed—at least in those already vast and growing areas where man is attempting to improve on nature's way. In the name of progress the land management agencies have set about to satisfy the insatiable demands of the cattlemen for more grazing land. By this they mean grassland—grass without sage. So in a land which nature found suited to grass growing mixed with and under the shelter of sage, it is now proposed to eliminate the sage and create unbroken grassland. Few seem to have asked whether grasslands are a stable and desirable goal in this region. Certainly nature's own answer was otherwise. The annual precipitation in this land where the rains seldom fall is not enough to support good sod-forming grass; it favors rather the perennial bunch grass that grows in the shelter of the sage. 10

Yet the program of sage eradication has been under way for a number of years. Several government agencies are active in it; industry has joined with enthusiasm to promote and encourage an enterprise which creates expanded markets not only for grass seed but for a large assortment of machines for cutting and plowing and 11 seeding. The newest addition to the weapons is the use of chemical sprays. Now millions of acres of sagebrush lands are sprayed each year.

12 What are the results? The eventual effects of eliminating sage and seeding with grass are largely conjectural. Men of long experience with the ways of the land say that in this country there is better growth of grass between and under the sage than can possibly be had in pure stands, once the moisture-holding sage is gone.

13 But even if the program succeeds in its immediate objective, it is clear that the whole closely knit fabric of life has been ripped apart. The antelope and the grouse will disappear along with the sage. The deer will suffer, too, and the land will be poorer for the destruction of the wild things that belong to it. Even the livestock which are the intended beneficiaries will suffer; no amount of lush green grass in summer can help the sheep starving in the winter storms for lack of the sage and bitterbrush and other wild vegetation of the plains.

14 These are the first and obvious effects. The second is of a kind that is always associated with the shotgun approach to nature: the spraying also eliminates a great many plants that were not its intended target. Justice William O. Douglas, in his recent book *My Wilderness: East to Katahdin*, has told of an appalling example of ecological destruction wrought by the United States Forest Service in the Bridger National Forest in Wyoming. Some 10,000 acres of sagelands were sprayed by the Service, yielding to pressure of cattlemen for more grasslands. The sage was killed, as intended. But so was the green, life giving ribbon of willows that traced its way across these plains, following the meandering streams. Moose had lived in these willow thickets, for willow is to the moose what sage is to the antelope. Beaver had lived there, too, feeding on the willows, felling them and making a strong dam across the tiny stream. Through the labor of the beavers, a lake backed up. Trout in the mountain streams seldom were more than six inches long; in the lake they thrived so prodigiously that many grew to five pounds. Waterfowl were attracted to the lake, also. Merely because of the presence of the willows and the beavers that depended on them, the region was an attractive recreational area with excellent fishing and hunting.

15 But with the "improvement" instituted by the Forest Service, the willows went the way of the sagebrush, killed by the same impartial spray. When Justice Douglas visited the area in 1959, the year of the spraying, he was shocked to see the shriveled and dying willows—the "vast, incredible damage." What would become

of the moose? Of the beavers and the little world they had constructed? A year later he returned to read the answers in the devastated landscape. The moose were gone and so were the beaver. Their principal dam had gone out for want of attention by its skilled architects, and the lake had drained away. None of the large trout were left. None could live in the tiny creek that remained, threading its way through a bare, hot land where no shade remained. The living world was shattered.

Besides the more than four million acres of rangelands sprayed each year, tremendous areas of other types of land are also potential or actual recipients of chemical treatments for weed control. For example, an area larger than all of New England—some 50 million acres—is under management by utility corporations and much of it is routinely treated for "brush control." In the Southwest an estimated 75 million acres of mesquite lands require management by some means, and chemical spraying is the method most actively pushed. An unknown but very large acreage of timber-producing lands is now aerially sprayed in order to "weed out" the hardwoods from the more spray-resistant conifers. Treatment of agricultural lands with herbicides doubled in the decade following 1949, totaling 53 million acres in 1959. And the combined acreage of private lawns, parks, and golf courses now being treated must reach an astronomical figure. 16

The chemical weed killers are a bright new toy. They work in a spectacular way; they give a giddy sense of power over nature to those who wield them, and as for the long-range and less obvious effects—these are easily brushed aside as the baseless imaginings of pessimists. The "agricultural engineers" speak blithely of "chemical plowing" in a world that is urged to beat its plowshares into spray guns. The town fathers of a thousand communities lend willing ears to the chemical salesman and the eager contractors who will rid the roadsides of "brush"—for a price. It is cheaper than mowing, is the cry. So, perhaps, it appears in the neat rows of figures in the official books; but were the true costs entered, the costs not only in dollars but in the many equally valid debits we shall presently consider, the wholesale broadcasting of chemicals would be seen to be more costly in dollars as well as infinitely damaging to the long-range health of the landscape and to all the varied interests that depend on it. 17

Take, for instance, that commodity prized by every chamber of commerce throughout the land—the good will of vacationing tourists. There is a steadily growing chorus of outraged protest about the disfigurement of once beautiful 18 roadsides by chemical sprays, which substitute a sere expanse of brown, withered vegetation for the beauty of fern and wildflower, of native shrubs adorned with blossom or berry. “We are making a dirty, brown, dying-looking mess along the sides of our roads,” a New England woman wrote angrily to her newspaper. “This is not what the tourists expect, with all the money we are spending advertising the beautiful scenery.”

19 In the summer of 1960 conservationists from many states converged on a peaceful Maine island to witness its presentation to the National Audubon Society by its owner, Millicent Todd Bingham. The focus that day was on the preservation of the natural landscape and of the intricate web of life whose interwoven strands lead from microbes to man. But in the background of all the conversations among the visitors to the island was indignation at the despoiling of the roads they had traveled. Once it had been a joy to follow those roads through the evergreen forests, roads lined with bayberry and sweet fern, alder and huckleberry. Now all was brown desolation. One of the conservationists wrote of that August pilgrimage to a Maine island: “I returned. . . angry at the desecration of the Maine roadsides. Where, in previous years, the highways were bordered with wildflowers and attractive shrubs, there were only the scars of dead vegetation for mile after mile . . . . As an economic proposition, can Maine afford the loss of tourist goodwill that such sights induce?”

20 Maine roadsides are merely one example, though a particularly sad one for those of us who have a deep love for the beauty of that state, of the senseless destruction that is going on in the name of roadside brush control throughout the nation.

21 Botanists at the Connecticut Arboretum declare that the elimination of beautiful native shrubs and wildflowers has reached the proportions of a “roadside crisis,” Azaleas, mountain laurel, blueberries, huckleberries, viburnums, dogwood, bayberry, sweet fern, low shadbush, winterberry, chokecherry, and wild plum are dying before the chemical barrage. So are the daisies, black-eyed Susans, Queen Anne’s lace, goldenrods, and fall asters which lend grace and beauty to the landscape.

22 The spraying is not only improperly planned but studded with abuses such as these. In a southern New England town one contractor finished his work with some chemical remaining in his tank. He discharged this along woodland roadsides where no spraying had been authorized. As a result the community lost the blue

and golden beauty of its autumn roads, where asters and goldenrod would have made a display worth traveling far to see. In another New England community a contractor changed the state specifications for town spraying without the knowledge of the highway department and sprayed roadside vegetation to a height of eight feet instead of the specified maximum of four feet, leaving a broad, disfiguring, brown swath. In a Massachusetts community the town officials purchased a weed killer from a zealous chemical salesman, unaware that it contained arsenic. One result of the subsequent roadside spraying was the death of a dozen cows from arsenic poisoning.

Trees within the Connecticut Arboretum Natural Area were seriously injured 23
when the town of Waterford sprayed the roadsides with chemical weed killers in 1957. Even large trees not directly sprayed were affected. The leaves of the oaks began to curl and turn brown, although it was the season for spring growth. Then new shoots began to be put forth and grew with abnormal rapidity, giving a weeping appearance to the trees. Two seasons later, large branches on these trees had died, others were without leaves, and the deformed, weeping effect of whole trees persisted.

I know well a stretch of road where nature's own landscaping has provided 24
a border of alder, viburnum, sweet fern, and juniper with seasonally changing accents of bright flowers, or of fruits hanging in jeweled clusters in the fall. The road had no heavy load of traffic to support; there were few sharp curves or intersections where brush could obstruct the driver's vision. But the sprayers took over and the miles along that road became something to be traversed quickly, a sight to be endured with one's mind closed to thoughts of the sterile and hideous world we are letting our technicians make. But here and there authority had somehow faltered and by an unaccountable oversight there were oases of beauty in the midst of austere and regimented control—oases that made the desecration of the greater part of the road the more unbearable. In such places my spirit lifted to the sight of the drifts of white clover or the clouds of purple vetch with here and there the flaming cup of a wood lily.

Such plants are "weeds" only to those who make a business of selling and applying 25
chemicals. In a volume of *Proceedings* of one of the weed-control conferences that are now regular institutions, I once read an extraordinary statement of a weed killer's philosophy. The author defended the killing of good plants "simply because they are

in bad company." Those who complain about killing wildflowers along roadsides reminded him, he said, of antivivisectionists "to whom, if one were to judge by their actions, the life of a stray dog is more sacred than the lives of children."

26 To the author of this paper, many of us would unquestionably be suspect, convicted of some deep perversion of character because we prefer the sight of the vetch and the clover and the wood lily in all their delicate and transient beauty to that of roadsides scorched as by fire, the shrubs brown and brittle, the bracken that once lifted high its proud lacework now withered and drooping. We would seem deplorably weak that we can tolerate the sight of such "weeds," that we do not rejoice in their eradication, that we are not filled with exultation that man has once more triumphed over miscreant nature.

27 Justice Douglas tells of attending a meeting of federal field men who were discussing protests by citizens against plans for the spraying of sagebrush that I mentioned earlier in this chapter. These men considered it hilariously funny that an old lady had opposed the plan because the wildflowers would be destroyed. "Yet, was not her right to search out a banded cup or a tiger lily as inalienable as the right of stockmen to search out grass or of a lumberman to claim a tree?" asks this humane and perceptive jurist. "The esthetic values of the wilderness are as much our inheritance as the veins of copper and gold in our hills and the forests in our mountains."

28 There is of course more to the wish to preserve our roadside vegetation than even such esthetic considerations. In the economy of nature the natural vegetation has its essential place. Hedgerows along country roads and bordering fields provide food, cover, and nesting areas for birds and homes for many small animals. Of some 70 species of shrubs and vines that are typical roadside species in the eastern states alone, about 65 are important to wildlife as food.

29 Such vegetation is also the habitat of wild bees and other pollinating insects. Man is more dependent on these wild pollinators than he usually realizes. Even the farmer himself seldom understands the value of wild bees and often participates in the very measures that rob him of their services. Some agricultural crops and many wild plants are partly or wholly dependent on the services of the native pollinating insects. Several hundred species of wild bees take part in the pollination of cultivated crops—100 species visiting the flowers of alfalfa alone. Without insect pollination, most of the soil-holding and soil-enriching plants of uncultivated areas

would die out, with far-reaching consequences to the ecology of the whole region. Many herbs, shrubs, and trees of forests and range depend on native insects for their reproduction; without these plants many wild animals and range stock would find little food. Now clean cultivation and the chemical destruction of hedgerows and weeds are eliminating the last sanctuaries of these pollinating insects and breaking the threads that bind life to life.

These insects, so essential to our agriculture and indeed to our landscape 30
as we know it, deserve something better from us than the senseless destruction of their habitat. Honeybees and wild bees depend heavily on such "weeds" as goldenrod, mustard, and dandelions for pollen that serves as the food of their young. Vetch furnishes essential spring forage for bees before the alfalfa is in bloom, tiding them over this early season so that they are ready to pollinate the alfalfa. In the fall they depend on goldenrod at a season when no other food is available, to stock up for the winter. By the precise and delicate timing that is nature's own, the emergence of one species of wild bees takes place on the very day of the opening of the willow blossoms. There is no dearth of men who understand these things, but these are not the men who order the wholesale drenching of the landscape with chemicals.

And where are the men who supposedly understand the value of proper habitat 31
for the preservation of wildlife? Too many of them are to be found defending herbicides as "harmless" to wildlife because they are thought to be less toxic than insecticides. Therefore, it is said, no harm is done. But as the herbicides rain down on forest and field, on marsh and rangeland, they are bringing about marked changes and even permanent destruction of wildlife habitat. To destroy the homes and the food of wildlife is perhaps worse in the long run than direct killing.

The irony of this all-out chemical assault on roadsides and utility rights-of- 32
way is twofold. It is perpetuating the problem it seeks to correct, for as experience has clearly shown, the blanket application of herbicides does not permanently control roadside "brush" and the spraying has to be repeated year after year. And as a further irony, we persist in doing this despite the fact that a perfectly sound method of *selective* spraying is known, which can achieve long-term vegetational control and eliminate repeated spraying in most types of vegetation.

The object of brush control along roads and rights-of-way is not to sweep the 33
land clear of everything but grass; it is, rather, to eliminate plants ultimately tall

enough to present an obstruction to drivers' vision or interference with wires on rights-of-way. This means, in general, trees. Most shrubs are low enough to present no hazard; so, certainly, are ferns and wildflowers.

34 Selective spraying was developed by Dr. Frank Egler during a period of years at the American Museum of Natural History as director of a Committee for Brush Control Recommendations for Rights-of-Way. It took advantage of the inherent stability of nature, building on the fact that most communities of shrubs are strongly resistant to invasion by trees. By comparison, grasslands are easily invaded by tree seedlings. The object of selective spraying is not to produce grass on roadsides and rights-of-way but to eliminate the tall woody plants by direct treatment and to preserve all other vegetation. One treatment may be sufficient, with a possible follow-up for extremely resistant species; thereafter the shrubs assert control and the trees do not return. The best and cheapest controls for vegetation are not chemicals but other plants.

35 The method has been tested in research areas scattered throughout the eastern United States. Results show that once properly treated, an area becomes stabilized, *requiring no re-spraying for at least 20 years*. The spraying can often be done by men on foot, using knapsack sprayers, and having complete control over their material. Sometimes compressor pumps and material can be mounted on truck chassis, but there is no blanket spraying. Treatment is directed only to trees and any exceptionally tall shrubs that must be eliminated. The integrity of the environment is thereby preserved, the enormous value of the wildlife habitat remains intact, and the beauty of shrub and fern and wildflower has not been sacrificed.

36 Here and there the method of vegetation management by selective spraying has been adopted. For the most part, entrenched custom dies hard and blanket spraying continues to thrive, to exact its heavy annual costs from the taxpayer, and to inflict its damage on the ecological web of life. It thrives, surely, only because the facts are not known. When taxpayers understand that the bill for spraying the town roads should come due only once a generation instead of once a year, they will surely rise up and demand a change of method.

37 Among the many advantages of selective spraying is the fact that it minimizes the amount of chemical applied to the landscape. There is no broadcasting of material but, rather, concentrated application to the base of the trees. The potential harm to wildlife is therefore kept to a minimum.

The most widely used herbicides are 2,4-D, 2,4,5-T, and related compounds. Whether or not these are actually toxic is a matter of controversy. People spraying their lawns with 2,4-D and becoming wet with spray have occasionally developed severe neuritis and even paralysis. Although such incidents are apparently uncommon, medical authorities advise caution in use of such compounds. Other hazards, more obscure, may also attend the use of 2,4-D. It has been shown experimentally to disturb the basic physiological process of respiration in the cell, and to imitate X-rays in damaging the chromosomes. Some very recent work indicates that reproduction of birds may be adversely affected by these and certain other herbicides at levels far below those that cause death. 38

Apart from any directly toxic effects, curious indirect results follow the use of certain herbicides. It has been found that animals, both wild herbivores and livestock, are sometimes strangely attracted to a plant that has been sprayed, even though it is not one of their natural foods. If a highly poisonous herbicide such as arsenic has been used, this intense desire to reach the wilting vegetation inevitably has disastrous results. Fatal results may follow, also, from less toxic herbicides if the plant itself happens to be poisonous or perhaps to possess thorns or burs. Poisonous range weeds, for example, have suddenly become attractive to livestock after spraying, and the animals have died from indulging this unnatural appetite. The literature of veterinary medicine abounds in similar examples: swine eating sprayed cockleburs with consequent severe illness, lambs eating sprayed thistles, bees poisoned by pasturing on mustard sprayed after it came into bloom. Wild cherry, the leaves of which are highly poisonous, has exerted a fatal attraction for cattle once its foliage has been sprayed with 2,4-D. Apparently the wilting that follows spraying (or cutting) makes the plant attractive. Ragwort has provided other examples. Livestock ordinarily avoid this plant unless forced to turn to it in late winter and early spring by lack of other forage. However, the animals eagerly feed on it after its foliage has been sprayed with 2,4-D. 39

The explanation of this peculiar behavior sometimes appears to lie in the changes which the chemical brings about in the metabolism of the plant itself. There is temporarily a marked increase in sugar content, making the plant more attractive to many animals. 40

Another curious effect of 2,4-D has important effects for livestock, wildlife, and apparently for men as well. Experiments carried out about a decade ago showed 41 that after treatment with this chemical there is a sharp increase in the nitrate content of corn and of sugar beets. The same effect was suspected in sorghum, sunflower, spiderwort, lambs quarters, pigweed, and smartweed. Some of these are normally ignored by cattle, but are eaten with relish after treatment with 2,4-D. A number of deaths among cattle have been traced to sprayed weeds, according to some agricultural specialists. The danger lies in the increase in nitrates, for the peculiar physiology of the ruminant at once poses a critical problem. Most such animals have a digestive system of extraordinary complexity, including a stomach divided into four chambers. The digestion of cellulose is accomplished through the action of microorganisms (rumen bacteria) in one of the chambers. When the animal feeds on vegetation containing an abnormally high level of nitrates, the microorganisms in the rumen act on the nitrates to change them into highly toxic nitrites. Thereafter a fatal chain of events ensues: the nitrites act on the blood pigment to form a chocolate-brown substance in which the oxygen is so firmly held that it cannot take part in respiration, hence oxygen is not transferred from the lungs to the tissues. Death occurs within a few hours from anoxia, or lack of oxygen. The various reports of livestock losses after grazing on certain weeds treated with 2,4-D therefore have a logical explanation. The same danger exists for wild animals belonging to the group of ruminants, such as deer, antelope, sheep, and goats.

42 Although various factors (such as exceptionally dry weather) can cause an increase in nitrate content, the effect of the soaring sales and applications of 2,4-D cannot be ignored. The situation was considered important enough by the University of Wisconsin Agricultural Experiment Station to justify a warning in 1957 that "plants killed by 2,4-D may contain large amounts of nitrate." The hazard extends to human beings as well as animals and may help to explain the recent mysterious increase in "silo deaths." When corn, oats, or sorghum containing large amounts of nitrates are ensiled they release poisonous nitrogen oxide gases, creating a deadly hazard to anyone entering the silo. Only a few breaths of one of these gases can cause a diffuse chemical pneumonia. In a series of such cases studied by the University of Minnesota Medical School all but one terminated fatally.

43 "Once again we are walking in nature like an elephant in the china cabinet." So C. J. Briejèr, a Dutch scientist of rare understanding, sums up our use of weed

killers. "In my opinion too much is taken for granted. We do not know whether all weeds in crops are harmful or whether some of them are useful," says Dr. Briejèr.

Seldom is the question asked, What is the relation between the weed and the soil? Perhaps, even from our narrow standpoint of direct self-interest, the relation is a useful one. As we have seen, soil and the living things in and upon it exist in a relation of interdependence and mutual benefit. Presumably the weed is taking something from the soil; perhaps it is also contributing something to it. A practical example was provided recently by the parks in a city in Holland. The roses were doing badly. Soil samples showed heavy infestations by tiny nematode worms. Scientists of the Dutch Plant Protection Service did not recommend chemical sprays or soil treatments; instead, they suggested that marigolds be planted among the roses. This plant, which the purist would doubtless consider a weed in any rosebed, releases an excretion from its roots that kills the soil nematodes. The advice was taken; some beds were planted with marigolds, some left without as controls. The results were striking. With the aid of the marigolds the roses flourished; in the control beds they were sickly and drooping. Marigolds are now used in many places for combating nematodes. 44

In the same way, and perhaps quite unknown to us, other plants that we ruthlessly eradicate may be performing a function that is necessary to the health of the soil. One very useful function of natural plant communities—now pretty generally stigmatized as "weeds"—is to serve as an indicator of the condition of the soil. This useful function is of course lost where chemical weed killers have been used. 45

Those who find an answer to all problems in spraying also overlook a matter of great scientific importance—the need to preserve some natural plant communities. We need these as a standard against which we can measure the changes our own activities bring about. We need them as wild habitats in which original populations of insects and other organisms can be maintained, for, as will be explained in Chapter 16, the development of resistance to insecticides is changing the genetic factors of insects and perhaps other organisms. One scientist has even suggested that some sort of "zoo" should be established to preserve insects, mites, and the like, before their genetic composition is further changed. 46

Some experts warn of subtle but far-reaching vegetational shifts as a result of the growing use of herbicides. The chemical 2,4-D, by killing out the broad-leaved plants, allows the grasses to thrive in the reduced competition—now some of the 47 grasses themselves have become "weeds," presenting a new problem in control and giving the cycle another turn. This strange situation is acknowledged in a recent issue of a journal devoted to crop problems: "With the widespread use of 2,4-D to control broadleaved weeds, grass weeds in particular have increasingly become a threat to corn and soybean yields."

48 Ragweed, the bane of hay fever sufferers, offers an interesting example of the way efforts to control nature sometimes boomerang. Many thousands of gallons of chemicals have been discharged along roadsides in the name of ragweed control. But the unfortunate truth is that blanket spraying is resulting in more ragweed, not less. Ragweed is an annual; its seedlings require open soil to become established each year. Our best protection against this plant is therefore the maintenance of dense shrubs, ferns, and other perennial vegetation. Spraying frequently destroys this protective vegetation and creates open, barren areas which the ragweed hastens to fill. It is probable, moreover, that the pollen content of the atmosphere is not related to roadside ragweed, but to the ragweed of city lots and fallow fields.

49 The booming sales of chemical crabgrass killers are another example of how readily unsound methods catch on. There is a cheaper and better way to remove crabgrass than to attempt year after year to kill it out with chemicals. This is to give it competition of a kind it cannot survive, the competition of other grass. Crabgrass exists only in an unhealthy lawn. It is a symptom, not a disease in itself. By providing a fertile soil and giving the desired grasses a good start, it is possible to create an environment in which crabgrass cannot grow, for it requires open space in which it can start from seed year after year.

50 Instead of treating the basic condition, suburbanites—advised by nurserymen who in turn have been advised by the chemical manufacturers—continue to apply truly astonishing amounts of crabgrass killers to their lawns each year. Marketed under trade names which give no hint of their nature, many of these preparations contain such poisons as mercury, arsenic, and chlordane. Application at the recommended rates leaves tremendous amounts of these chemicals on the lawn. Users of one product, for example, apply 60 pounds of technical chlordane to the acre if they follow directions. If they use another of the many available products, they are applying 175 pounds of metallic arsenic to the acre. The toll of dead birds, as we shall see in Chapter 8, is distressing. How lethal these lawns may be for human beings is unknown.

51 The success of selective spraying for roadside and right-of-way vegetation, where it has been practiced, offers hope that equally sound ecological methods may be developed for other vegetation programs for farms, forests, and ranges—methods aimed not at destroying a particular species but at managing vegetation as a living community.

52 Other solid achievements show what can be done. Biological control has achieved some of its most spectacular successes in the area of curbing unwanted vegetation. Nature herself has met many of the problems that now beset us, and she has usually solved them in her own successful way. Where man has been intelligent enough to observe and to emulate Nature he, too, is often rewarded with success.

53 An outstanding example in the field of controlling unwanted plants is the handling of the Klamath-weed problem in California. Although the Klamath weed, or goatweed, is a native of Europe (where it is called St. Johnswort), it accompanied man in his westward migrations, first appearing in the United States in 1793 near Lancaster, Pennsylvania. By 1900 it had reached California in the vicinity of the Klamath River, hence the name locally given to it. By 1929 it had occupied about 100,000 acres of rangeland, and by 1952 it had invaded some two and one half million acres.

54 Klamath weed, quite unlike such native plants as sagebrush, has no place in the ecology of the region, and no animals or other plants require its presence. On the contrary, wherever it appeared livestock became "scabby, sore-mouthed, and unthrifty" from feeding on this toxic plant. Land values declined accordingly, for the Klamath weed was considered to hold the first mortgage.

55 In Europe the Klamath weed, or St. Johnswort, has never become a problem because along with the plant there have developed various species of insects; these feed on it so extensively that its abundance is severely limited. In particular, two species of beetles in southern France, pea-sized and of metallic color, have their whole beings so adapted to the presence of the weed that they feed and reproduce only upon it.

56 It was an event of historic importance when the first shipments of these beetles were brought to the United States in 1944, for this was the first attempt in North America to control a plant with a plant-eating insect. By 1948 both species had become so well established that no further importations were needed. Their spread was accomplished by collecting beetles from the original colonies and redistributing

them at the rate of millions a year. Within small areas the beetles accomplish their own dispersion, moving on as soon as the Klamath weed dies out and locating new stands with great precision. And as the beetles thin out the weed, desirable range plants that have been crowded out are able to return.

57 A ten-year survey completed in 1959 showed that control of the Klamath weed had been "more effective than hoped for even by enthusiasts," with the weed reduced to a mere 1 per cent of its former abundance. This token infestation is harmless and is actually needed in order to maintain a population of beetles as protection against a future increase in the weed.

58 Another extraordinarily successful and economical example of weed control may be found in Australia. With the colonists' usual taste for carrying plants or animals into a new country, a Captain Arthur Phillip had brought various species of cactus into Australia about 1787, intending to use them in culturing cochineal insects for dye. Some of the cacti or prickly pears escaped from his gardens and by 1925 about 20 species could be found growing wild. Having no natural controls in this new territory, they spread prodigiously, eventually occupying about 60 million acres. At least half of this land was so densely covered as to be useless.

59 In 1920 Australian entomologists were sent to North and South America to study insect enemies of the prickly pears in their native habitat. After trials of several species, 3 billion eggs of an Argentine moth were released in Australia in 1930. Seven years later the last dense growth of the prickly pear had been destroyed and the once uninhabitable areas reopened to settlement and grazing. The whole operation had cost less than a penny per acre. In contrast, the unsatisfactory attempts at chemical control in earlier years had cost about £10 per acre.

60 Both of these examples suggest that extremely effective control of many kinds of unwanted vegetation might be achieved by paying more attention to the role of plant-eating insects. The science of range management has largely ignored this possibility, although these insects are perhaps the most selective of all grazers and their highly restricted diets could easily be turned to man's advantage.

Part III
Our Understanding of Human Understanding

Now we reflect on our own understanding. Scientists investigate the physical universe and the world of life with their minds, using the faculty of reasoning. But how does the human brain operate? What is the human mind? Do we have a soul? What limitations does reasoning have? Many scientists have tackled these problems. Two of them were Henri Poincaré and Eric Kandel.

In the excerpt (Text 7) from *Science and Method*, Poincaré shares his idea of beauty. He also looks into his experience of mathematical discovery and finds a possible mechanism for how mathematical ideas form. While Poincaré's study of the psychology of scientific discovery was an early attempt of scientists to understand the human mind, Kandel's *In Search of Memory* is a concise description of the current status of the subject. In the excerpt (Text 8) from this book, the Nobel Laureate in Physiology or Medicine raised the two central problems of the study of the human mind, one is the easy one, unity of the consciousness, and the other the hard one, the subjectivity of consciousness.

It is generally agreed that modern science emerged in Western Europe. However, there was a long period of time when the most advanced technology was found in China. Why, therefore, did modern science not emerge in China? Can Chinese culture make any contribution to modern science? These are very big questions that have attracted great minds. Joseph Needham was one of them. In an excerpt (Text 9) from *The Shorter Science and Civilisation in China*, Needham talks about how Chinese people understood the world with the five elements, *yin-yang* and associative thinking.

On the question why modern science emerged in the West instead of China, Needham took a socio-economical approach and suggested the rise of the bourgeoisie as the primary cause. Some people agree with him but some do not. In his often cited paper (Text 10a), Nathan Sivin points out that there were sciences in China. However, unity in Chinese science did not exist as evidenced by Shen Kua's system of thought. Sivin points out two loopholes in historical reasoning and suggests that there was indeed a scientific revolution in China. In the excerpt (Text 10b) chosen from Shen Kua's *Brush Talks from Dream Brook* (《夢溪筆談》), the reader can take a closer look into Chinese science.

The *Principia* (Text 3b) begins with definitions and axioms, from which Newton derives all theorems of motion. This style of writing originated from Euclid's *Elements*, which was written in c. 300 B.C. in ancient Greece. The great achievement of Euclid is that he has demonstrated how the whole body of geometry can be derived from several irreducible definitions, postulates and common notions. *Elements* is not an easy text. It is written in a precise and concise manner that many readers find it boring. In the excerpt (Text 11a) from William Dunham's *The Mathematical Universe*, Dunham gives an excellent introduction to the basic ideas of *Elements*. Text 11b is an excerpt from *Elements*. The reader will appreciate how Euclid systematically proved the well-known theorem that the sum of any two sides of a triangle is longer than the third.

Text 7

from

Science and Method

by Henri Poincaré

I. THE SELECTION OF FACTS

1 Tolstoi explains somewhere in his writings why, in his opinion, "Science for Science's sake" is an absurd conception. We cannot know all the facts, since they are practically infinite in number. We must make a selection; and that being so, can this selection be governed by the mere caprice of our curiosity? Is it not better to be guided by utility, by our practical, and more especially our moral, necessities? Have we not some better occupation than counting the number of lady-birds in existence on this planet?

2 It is clear that for him the word *utility* has not the meaning assigned to it by business men, and, after them, by the greater number of our contemporaries. He cares but little for the industrial applications of science, for the marvels of electricity or of automobilism, which he regards rather as hindrances to moral progress. For him the useful is exclusively what is capable of making men better.

3 It is hardly necessary for me to state that, for my part, I could not be satisfied with either of these ideals. I have no liking either for a greedy and narrow plutocracy, or for a virtuous unaspiring democracy, solely occupied in turning the other cheek, in which we should find good people devoid of curiosity, who, avoiding all excesses, would not die of any disease—save boredom. But it is all a matter of taste, and that is not the point I wish to discuss.

From Henri Poincaré, *The Value of Science: Essential Writings of Henri Poincaré* (2001), published by Modern Library. Reproduced from the public domain.

4 Nonetheless the question remains, and it claims our attention. If our selection is only determined by caprice or by immediate necessity, there can be no science for science's sake, and consequently no science. Is this true? There is no disputing the fact that a selection must be made: however great our activity, facts outstrip us, and we can never overtake them; while the scientist is discovering one fact, millions and millions are produced in every cubic inch of his body. Trying to make science contain nature is like trying to make the part contain the whole.

5 But scientists believe that there is a hierarchy of facts, and that a judicious selection can be made. They are right, for otherwise there would be no science, and science does exist. One has only to open one's eyes to see that the triumphs of industry, which have enriched so many practical men, would never have seen the light if only these practical men had existed, and if they had not been preceded by disinterested fools who died poor, who never thought of the useful, and yet had a guide that was not their own caprice.

6 What these fools did, as Mach has said, was to save their successors the trouble of thinking. If they had worked solely in view of an immediate application, they would have left nothing behind them, and in face of a new requirement, all would have had to be done again. Now the majority of men do not like thinking, and this is perhaps a good thing, since instinct guides them, and very often better than reason would guide a pure intelligence, at least whenever they are pursuing an end that is immediate and always the same. But instinct is routine, and if it were not fertilized by thought, it would advance no further with man than with the bee or the ant. It is necessary, therefore, to think for those who do not like thinking, and as they are many, each one of our thoughts must be useful in as many circumstances as possible. For this reason, the more general a law is, the greater is its value.

7 This shows us how our selection should be made. The most interesting facts are those which can be used several times, those which have a chance of recurring.We have been fortunate enough to be born in a world where there are such facts. Suppose that instead of eighty chemical elements we had eighty millions, and that they were not some common and others rare, but uniformly distributed. Then each time we picked up a new pebble there would be a strong probability that it was composed of some unknown substance. Nothing that we knew of other pebbles would tell us anything about it. Before each new object we should be like a new-born child; like him we could but obey our caprices or our necessities. In such a world there would

be no science, perhaps thought and even life would be impossible, since evolution could not have developed the instincts of self-preservation. Providentially it is not so; but this blessing, like all those to which we are accustomed, is not appreciated at its true value. The biologist would be equally embarrassed if there were only individuals and no species, and if heredity did not make children resemble their parents.

Which, then, are the facts that have a chance of recurring? In the first place, 8
simple facts. It is evident that in a complex fact many circumstances are united by chance, and that only a still more improbable chance could ever so unite them again. But are there such things as simple facts? and if there are, how are we to recognize them? Who can tell that what we believe to be simple does not conceal an alarming complexity? All that we can say is that we must prefer facts which appear simple, to those in which our rude vision detects dissimilar elements. Then only two alternatives are possible; either this simplicity is real, or else the elements are so intimately mingled that they do not admit of being distinguished. In the first case we have a chance of meeting the same simple fact again, either in all its purity, or itself entering as an element into some complex whole. In the second case the intimate mixture has similarly a greater chance of being reproduced than a heterogeneous assemblage. Chance can mingle, but it cannot unmingle, and a combination of various elements in a well-ordered edifice in which something can be distinguished, can only be made deliberately. There is, therefore, but little chance that an assemblage in which different things can be distinguished should ever be reproduced. On the other hand, there is great probability that a mixture which appears homogeneous at first sight will be reproduced several times. Accordingly facts which appear simple, even if they are not so in reality, will be more easily brought about again by chance.

It is this that justifies the method instinctively adopted by scientists, and what 9
perhaps justifies it still better is that facts which occur frequently appear to us simple just because we are accustomed to them.

But where is the simple fact? Scientists have tried to find it in the two 10
extremes, in the infinitely great and in the infinitely small. The astronomer has found it because the distances of the stars are immense, so great that each of them appears only as a point and qualitative differences disappear, and because a point is simpler than a body which has shape and qualities. The physicist, on the other hand, has sought the elementary phenomenon in an imaginary division of bodies into

infinitely small atoms, because the conditions of the problem, which undergo slow and continuous variations as we pass from one point of the body to another, may be regarded as constant within each of these little atoms. Similarly the biologist has been led instinctively to regard the cell as more interesting than the whole animal, and the event has proved him right, since cells belonging to the most diverse organisms have greater resemblances, for those who can recognize them, than the organisms themselves. The sociologist is in a more embarrassing position. The elements, which for him are men, are too dissimilar, too variable, too capricious, in a word, too complex themselves. Furthermore, history does not repeat itself; how, then, is he to select the interesting fact, the fact which is repeated? Method is precisely the selection of facts, and accordingly our first care must be to devise a method. Many have been devised because none holds the field undisputed. Nearly every sociological thesis proposes a new method, which, however, its author is very careful not to apply, so that sociology is the science with the greatest number of methods and the least results.

11 It is with regular facts, therefore, that we ought to begin; but as soon as the rule is well established, as soon as it is no longer in doubt, the facts which are in complete conformity with it lose their interest, since they can teach us nothing new. Then it is the exception which becomes important. We cease to look for resemblances, and apply ourselves before all else to differences, and of these differences we select first those that are most accentuated, not only because they are the most striking, but because they will be the most instructive. This will be best explained by a simple example. Suppose we are seeking to determine a curve by observing some of the points on it. The practical man who looked only to immediate utility would merely observe the points he required for some special object; these points would be badly distributed on the curve, they would be crowded together in certain parts and scarce in others, so that it would be impossible to connect them by a continuous line, and they would be useless for any other application. The scientist would proceed in a different manner. Since he wishes to study the curve for itself, he will distribute the points to be observed regularly, and as soon as he knows some of them, he will join them by a regular line, and he will then have the complete curve. But how is he to accomplish this? If he has determined one extreme point on the curve, he will not remain close to this extremity, but will move to the other end. After the two extremities, the central point is the most instructive, and so on.

Thus when a rule has been established, we have first to look for the cases 12
in which the rule stands the best chance of being found in fault. This is one of many reasons for the interest of astronomical facts and of geological ages. By making long excursions in space or in time, we may find our ordinary rules completely upset, and these great upsettings will give us a clearer view and better comprehension of such small changes as may occur nearer us, in the small corner of the world in which we are called to live and move. We shall know this corner better for the journey we have taken into distant lands where we had no concern.

But what we must aim at is not so much to ascertain resemblances and 13
differences, as to discover similarities hidden under apparent discrepancies. The individual rules appear at first discordant, but on looking closer we can generally detect a resemblance; though differing in matter, they approximate in form and in the order of their parts. When we examine them from this point of view, we shall see them widen and tend to embrace everything. This is what gives a value to certain facts that come to complete a whole, and show that it is the faithful image of other known wholes.

I cannot dwell further on this point, but these few words will suffice to show 14
that the scientist does not make a random selection of the facts to be observed. He does not count lady-birds, as Tolstoi says, because the number of these insects, interesting as they are, is subject to capricious variations. He tries to condense a great deal of experience and a great deal of thought into a small volume, and that is why a little book on physics contains so many past experiments, and a thousand times as many possible ones, whose results are known in advance.

But so far we have only considered one side of the question. The scientist 15
does not study nature because it is useful to do so. He studies it because he takes pleasure in it, and he takes pleasure in it because it is beautiful. If nature were not beautiful it would not be worth knowing, and life would not be worth living. I am not speaking, of course, of that beauty which strikes the senses, of the beauty of qualities and appearances. I am far from despising this, but it has nothing to do with science. What I mean is that more intimate beauty which comes from the harmonious order of its parts, and which a pure intelligence can grasp. It is this that gives a body a skeleton, so to speak, to the shimmering visions that flatter our senses, and without this support the beauty of these fleeting dreams would be

imperfect, because it would be indefinite and ever elusive. Intellectual beauty, on the contrary, is self-sufficing, and it is for it, more perhaps than for the future good of humanity, that the scientist condemns himself to long and painful labours.

16 It is, then, the search for this special beauty, the sense of the harmony of the world, that makes us select the facts best suited to contribute to this harmony; just as the artist selects those features of his sitter which complete the portrait and give it character and life. And there is no fear that this instinctive and unacknowledged preoccupation will divert the scientist from the search for truth. We may dream of a harmonious world, but how far it will fall short of the real world! The Greeks, the greatest artists that ever were, constructed a heaven for themselves; how poor a thing it is beside the heaven as we know it!

17 It is because simplicity and vastness are both beautiful that we seek by preference simple facts and vast facts; that we take delight, now in following the giant courses of the stars, now in scrutinizing with a microscope that prodigious smallness which is also a vastness, and now in seeking in geological ages the traces of a past that attracts us because of its remoteness.

18 Thus we see that care for the beautiful leads us to the same selection as care for the useful. Similarly economy of thought, that economy of effort which, according to Mach, is the constant tendency of science, is a source of beauty as well as a practical advantage. The buildings we admire are those in which the architect has succeeded in proportioning the means to the end, in which the columns seem to carry the burdens imposed on them lightly and without effort, like the graceful caryatids of the Erechtheum.

19 Whence comes this concordance? Is it merely that things which seem to us beautiful are those which are best adapted to our intelligence, and that consequently they are at the same time the tools that intelligence knows best how to handle? Or is it due rather to evolution and natural selection? Have the peoples whose ideal conformed best to their own interests, properly understood, exterminated the others and taken their place? One and all pursued their ideal without considering the consequences, but while this pursuit led some to their destruction, it gave empire to others. We are tempted to believe this, for if the Greeks triumphed over the barbarians, and if Europe, heir of the thought of the Greeks, dominates the world, it is due to the fact that the savages loved garish colours and the blatant noise of the drum, which appealed to their senses, while the Greeks loved the intellectual beauty hidden behind sensible beauty, and that it is this beauty which gives certainty and

strength to the intelligence.

No doubt Tolstoi would be horrified at such a triumph, and he would refuse 20
to admit that it could be truly useful. But this disinterested pursuit of truth for its own beauty is also wholesome, and can make men better. I know very well there are disappointments, that the thinker does not always find the serenity he should, and even that some scientists have thoroughly bad tempers.

Must we therefore say that science should be abandoned, and morality alone 21
be studied? Does anyone suppose that moralists themselves are entirely above reproach when they have come down from the pulpit?

[. . .]

III. MATHEMATICAL DISCOVERY

The genesis of mathematical discovery is a problem which must inspire the 1
psychologist with the keenest interest. For this is the process in which the human mind seems to borrow least from the exterior world, in which it acts, or appears to act, only by itself and on itself, so that by studying the process of geometric thought we may hope to arrive at what is most essential in the human mind.

This has long been understood, and a few months ago a review called 2
L'Enseignement mathématique, edited by MM. Laisant and Fehr, instituted an enquiry into the habits of mind and methods of work of different mathematicians. I had outlined the principal features of this article when the results of the enquiry were published, so that I have hardly been able to make any use of them, and I will content myself with saying that the majority of the evidence confirms my conclusions. I do not say there is unanimity, for on an appeal to universal suffrage we cannot hope to obtain unanimity.

One first fact must astonish us, or rather would astonish us if we were not 3
too much accustomed to it. How does it happen that there are people who do not understand mathematics? If the science invokes only the rules of logic, those accepted by all well-formed minds, if its evidence is founded on principles that are common to all men, and that none but a madman would attempt to deny, how does it happen that there are so many people who are entirely impervious to it?

4 There is nothing mysterious in the fact that everyone is not capable of discovery. That everyone should not be able to retain a demonstration he has once learnt is still comprehensible. But what does seem most surprising, when we consider it, is that anyone should be unable to understand a mathematical argument at the very moment it is stated to him. And yet those who can only follow the argument with difficulty are in a majority; this is incontestable, and the experience of teachers of secondary education will certainly not contradict me.

5 And still further, how is error possible in mathematics? A healthy intellect should not be guilty of any error in logic, and yet there are very keen minds which will not make a false step in a short argument such as those we have to make in the ordinary actions of life, which yet are incapable of following or repeating without error the demonstrations of mathematics which are longer, but which are, after all, only accumulations of short arguments exactly analogous to those they make so easily. Is it necessary to add that mathematicians themselves are not infallible?

6 The answer appears to me obvious. Imagine a long series of syllogisms in which the conclusions of those that precede form the premisses of those that follow. We shall be capable of grasping each of the syllogisms, and it is not in the passage from premisses to conclusion that we are in danger of going astray. But between the moment when we meet a proposition for the first time as the conclusion of one syllogism, and the moment when we find it once more as the premise of another syllogism, much time will sometimes have elapsed, and we shall have unfolded many links of the chain; accordingly it may well happen that we shall have forgotten it, or, what is more serious, forgotten its meaning. So we may chance to replace it by a somewhat different proposition, or to preserve the same statement but give it a slightly different meaning, and thus we are in danger of falling into error.

7 A mathematician must often use a rule, and, naturally, he begins by demonstrating the rule. At the moment the demonstration is quite fresh in his memory he understands perfectly its meaning and significance, and he is in no danger of changing it. But later on he commits it to memory, and only applies it in a mechanical way, and then, if his memory fails him, he may apply it wrongly. It is thus, to take a simple and almost vulgar example, that we sometimes make mistakes in calculation, because we have forgotten our multiplication table.

8 On this view special aptitude for mathematics would be due to nothing but a very certain memory or a tremendous power of attention. It would be a quality

analogous to that of the whist player who can remember the cards played, or, to rise a step higher, to that of the chess player who can picture a very great number of combinations and retain them in his memory. Every good mathematician should also be a good chess player and *vice versa*, and similarly he should be a good numerical calculator. Certainly this sometimes happens, and thus Gauss was at once a geometrician of genius and a very precocious and very certain calculator.

But there are exceptions, or rather I am wrong, for I cannot call them exceptions, 9
otherwise the exceptions would be more numerous than the cases of conformity with the rule. On the contrary, it was Gauss who was an exception. As for myself, I must confess I am absolutely incapable of doing an addition sum without a mistake. Similarly I should be a very bad chess player. I could easily calculate that by playing in a certain way I should be exposed to such and such a danger; I should then review many other moves, which I should reject for other reasons, and I should end by making the move I first examined, having forgotten in the interval the danger I had foreseen.

In a word, my memory is not bad, but it would be insufficient to make me 10
a good chess player. Why, then, does it not fail me in a difficult mathematical argument in which the majority of chess players would be lost? Clearly because it is guided by the general trend of the argument. A mathematical demonstration is not a simple juxtaposition of syllogisms; it consists of syllogisms *placed in a certain order*, and the order in which these elements are placed is much more important than the elements themselves. If I have the feeling, so to speak the intuition, of this order, so that I can perceive the whole of the argument at a glance, I need no longer be afraid of forgetting one of the elements; each of them will place itself naturally in the position prepared for it, without my having to make any effort of memory.

It seems to me, then, as I repeat an argument I have learnt, that I could have 11
discovered it. This is often only an illusion; but even then, even if I am not clever enough to create for myself, I rediscover it myself as I repeat it.

We can understand that this feeling, this intuition of mathematical order, 12
which enables us to guess hidden harmonies and relations, cannot belong to everyone. Some have neither this delicate feeling that is difficult to define, nor a power of memory and attention above the common, and so they are absolutely incapable of understanding even the first steps of higher mathematics. This applies to the majority of people. Others have the feeling only in a slight degree, but they

are gifted with an uncommon memory and a great capacity for attention. They learn the details one after the other by heart, they can understand mathematics and sometimes apply them, but they are not in a condition to create. Lastly, others possess the special intuition I have spoken of more or less highly developed, and they cannot only understand mathematics, even though their memory is in no way extraordinary, but they can become creators, and seek to make discovery with more or less chance of success, according as their intuition is more or less developed.

13 What, in fact, is mathematical discovery? It does not consist in making new combinations with mathematical entities that are already known. That can be done by anyone, and the combinations that could be so formed would be infinite in number, and the greater part of them would be absolutely devoid of interest. Discovery consists precisely in not constructing useless combinations, but in constructing those that are useful, which are an infinitely small minority. Discovery is discernment, selection.

14 How this selection is to be made I have explained above. Mathematical facts worthy of being studied are those which, by their analogy with other facts, are capable of conducting us to the knowledge of a mathematical law, in the same way that experimental facts conduct us to the knowledge of a physical law. They are those which reveal unsuspected relations between other facts, long since known, but wrongly believed to be unrelated to each other.

15 Among the combinations we choose, the most fruitful are often those which are formed of elements borrowed from widely separated domains. I do not mean to say that for discovery it is sufficient to bring together objects that are as incongruous as possible. The greater part of the combinations so formed would be entirely fruitless, but some among them, though very rare, are the most fruitful of all.

16 Discovery, as I have said, is selection. But this is perhaps not quite the right word. It suggests a purchaser who has been shown a large number of samples, and examines them one after the other in order to make his selection. In our case the samples would be so numerous that a whole life would not give sufficient time to examine them. Things do not happen in this way. Unfruitful combinations do not so much as present themselves to the mind of the discoverer. In the field of his consciousness there never appear any but really useful combinations, and some that he rejects, which, however, partake to some extent of the character of useful combinations. Everything happens as if the discoverer were a secondary

examiner who had only to interrogate candidates declared eligible after passing a preliminary test.

But what I have said up to now is only what can be observed or inferred by 17
reading the works of geometricians, provided they are read with some reflection.

It is time to penetrate further, and to see what happens in the very soul of the 18
mathematician. For this purpose I think I cannot do better than recount my personal recollections. Only I am going to confine myself to relating how I wrote my first treatise on Fuchsian functions. I must apologize, for I am going to introduce some technical expressions, but they need not alarm the reader, for he has no need to understand them. I shall say, for instance, that I found the demonstration of such and such a theorem under such and such circumstances; the theorem will have a barbarous name that many will not know, but that is of no importance. What is interesting for the psychologist is not the theorem but the circumstances.

For a fortnight I had been attempting to prove that there could not be any 19
function analogous to what I have since called Fuchsian functions. I was at that time very ignorant. Every day I sat down at my table and spent an hour or two trying a great number of combinations, and I arrived at no result. One night I took some black coffee, contrary to my custom, and was unable to sleep. A host of ideas kept surging in my head; I could almost feel them jostling one another, until two of them coalesced, so to speak, to form a stable combination. When morning came, I had established the existence of one class of Fuchsian functions, those that are derived from the hyper-geometric series. I had only to verify the results, which only took a few hours.

Then I wished to represent these functions by the quotient of two series. This 20
idea was perfectly conscious and deliberate; I was guided by the analogy with elliptical functions. I asked myself what must be the properties of these series, if they existed, and I succeeded without difficulty in forming the series that I have called Theta-Fuchsian.

At this moment I left Caen, where I was then living, to take part in a geological 21
conference arranged by the School of Mines. The incidents of the journey made me forget my mathematical work. When we arrived at Coutances, we got into a break to go for a drive, and, just as I put my foot on the step, the idea came to me, though nothing in my former thoughts seemed to have prepared me for it, that the transformations I had used to define Fuchsian functions were identical with

those of non-Euclidian geometry. I made no verification, and had no time to do so, since I took up the conversation again as soon as I had sat down in the break, but I felt absolute certainty at once. When I got back to Caen I verified the result at my leisure to satisfy my conscience.

22 I then began to study arithmetical questions without any great apparent result, and without suspecting that they could have the least connexion with my previous researches. Disgusted at my want of success, I went away to spend a few days at the seaside, and thought of entirely different things. One day, as I was walking on the cliff, the idea came to me, again with the same characteristics of conciseness, suddenness, and immediate certainty, that arithmetical transformations of indefinite ternary quadratic forms are identical with those of non-Euclidian geometry.

23 Returning to Caen, I reflected on this result and deduced its consequences. The example of quadratic forms showed me that there are Fuchsian groups other than those which correspond with the hyper-geometric series; I saw that I could apply to them the theory of the Theta-Fuchsian series, and that, consequently, there are Fuchsian functions other than those which are derived from the hyper-geometric series, the only ones I knew up to that time. Naturally, I proposed to form all these functions. I laid siege to them systematically and captured all the outworks one after the other. There was one, however, which still held out, whose fall would carry with it that of the central fortress. But all my efforts were of no avail at first, except to make me better understand the difficulty, which was already something. All this work was perfectly conscious.

24 Thereupon I left for Mont-Valérien, where I had to serve my time in the army, and so my mind was preoccupied with very different matters. One day, as I was crossing the street, the solution of the difficulty which had brought me to a standstill came to me all at once. I did not try to fathom it immediately, and it was only after my service was finished that I returned to the question. I had all the elements, and had only to assemble and arrange them. Accordingly I composed my definitive treatise at a sitting and without any difficulty.

25 It is useless to multiply examples, and I will content myself with this one alone. As regards my other researches, the accounts I should give would be exactly similar, and the observations related by other mathematicians in the enquiry of *L'Enseignement mathématique* would only confirm them.

One is at once struck by these appearances of sudden illumination, obvious indications of a long course of previous unconscious work. The part played by this unconscious work in mathematical discovery seems to me indisputable, and we shall find traces of it in other cases where it is less evident. Often when a man is working at a difficult question, he accomplishes nothing the first time he sets to work. Then he takes more or less of a rest, and sits down again at his table. During the first half-hour he still finds nothing, and then all at once the decisive idea presents itself to his mind. We might say that the conscious work proved more fruitful because it was interrupted and the rest restored force and freshness to the mind. But it is more probable that the rest was occupied with unconscious work, and that the result of this work was afterwards revealed to the geometrician exactly as in the cases I have quoted, except that the revelation, instead of coming to light during a walk or a journey, came during a period of conscious work, but independently of that work, which at most only performs the unlocking process, as if it were the spur that excited into conscious form the results already acquired during the rest, which till then remained unconscious. 26

There is another remark to be made regarding the conditions of this unconscious work, which is, that it is not possible, or in any case not fruitful, unless it is first preceded and then followed by a period of conscious work. These sudden inspirations are never produced (and this is sufficiently proved already by the examples I have quoted) except after some days of voluntary efforts which appeared absolutely fruitless, in which one thought one had accomplished nothing, and seemed to be on a totally wrong track. These efforts, however, were not as barren as one thought; they set the unconscious machine in motion, and without them it would not have worked at all, and would not have produced anything. 27

The necessity for the second period of conscious work can be even more readily understood. It is necessary to work out the results of the inspiration, to deduce the immediate consequences and put them in order and to set out the demonstrations; but, above all, it is necessary to verify them. I have spoken of the feeling of absolute certainty which accompanies the inspiration; in the cases quoted this feeling was not deceptive, and more often than not this will be the case. But we must beware of thinking that this is a rule without exceptions. Often the feeling deceives us without being any less distinct on that account, and we only detect it 28 when we attempt to establish the demonstration. I have observed this fact most notably with regard to ideas that have come to me in the morning or at night when I have been in bed in a semi-somnolent condition.

29 Such are the facts of the case, and they suggest the following reflections. The result of all that precedes is to show that the unconscious ego, or, as it is called, the subliminal ego, plays a most important part in mathematical discovery. But the subliminal ego is generally thought of as purely automatic. Now we have seen that mathematical work is not a simple mechanical work, and that it could not be entrusted to any machine, whatever the degree of perfection we suppose it to have been brought to. It is not merely a question of applying certain rules, of manufacturing as many combinations as possible according to certain fixed laws. The combinations so obtained would be extremely numerous, useless, and encumbering. The real work of the discoverer consists in choosing between these combinations with a view to eliminating those that are useless, or rather not giving himself the trouble of making them at all. The rules which must guide this choice are extremely subtle and delicate, and it is practically impossible to state them in precise language; they must be felt rather than formulated. Under these conditions, how can we imagine a sieve capable of applying them mechanically?

30 The following, then, presents itself as a first hypothesis. The subliminal ego is in no way inferior to the conscious ego; it is not purely automatic; it is capable of discernment; it has tact and lightness of touch; it can select, and it can divine. More than that, it can divine better than the conscious ego, since it succeeds where the latter fails. In a word, is not the subliminal ego superior to the conscious ego? The importance of this question will be readily understood. In a recent lecture, M. Boutroux showed how it had arisen on entirely different occasions, and what consequences would be involved by an answer in the affirmative. (See also the same author's *Science et religion*, pp. 313 *et seq.*)

31 Are we forced to give this affirmative answer by the facts I have just stated? I confess that, for my part, I should be loth to accept it. Let us, then, return to the facts, and see if they do not admit of some other explanation.

32 It is certain that the combinations which present themselves to the mind in a kind of sudden illumination after a somewhat prolonged period of unconscious work are generally useful and fruitful combinations, which appear to be the result of a preliminary sifting. Does it follow from this that the subliminal ego, having

divined by a delicate intuition that these combinations could be useful, has formed none but these, or has it formed a great many others which were devoid of interest, and remained unconscious?

Under this second aspect, all the combinations are formed as a result of the automatic action of the subliminal ego, but those only which are interesting find their way into the field of consciousness. This, too, is most mysterious. How can we explain the fact that, of the thousand products of our unconscious activity, some are invited to cross the threshold, while others remain outside? Is it mere chance that gives them this privilege? Evidently not. For instance, of all the excitements of our senses, it is only the most intense that retain our attention, unless it has been directed upon them by other causes. More commonly the privileged unconscious phenomena, those that are capable of becoming conscious, are those which, directly or indirectly, most deeply affect our sensibility. 33

It may appear surprising that sensibility should be introduced in connexion with mathematical demonstrations, which, it would seem, can only interest the intellect. But not if we bear in mind the feeling of mathematical beauty, of the harmony of numbers and forms and of geometric elegance. It is a real aesthetic feeling that all true mathematicians recognize, and this is truly sensibility. 34

Now, what are the mathematical entities to which we attribute this character of beauty and elegance, which are capable of developing in us a kind of aesthetic emotion? Those whose elements are harmoniously arranged so that the mind can, without effort, take in the whole without neglecting the details. This harmony is at once a satisfaction to our aesthetic requirements, and an assistance to the mind which it supports and guides. At the same time, by setting before our eyes a well-ordered whole, it gives us a presentiment of a mathematical law. Now, as I have said above, the only mathematical facts worthy of retaining our attention and capable of being useful are those which can make us acquainted with a mathematical law. Accordingly we arrive at the following conclusion. The useful combinations are precisely the most beautiful, I mean those that can most charm that special sensibility that all mathematicians know, but of which laymen are so ignorant that they are often tempted to smile at it. 35

What follows, then? Of the very large number of combinations which the subliminal ego blindly forms, almost all are without interest and without utility. But, for that very reason, they are without action on the aesthetic sensibility; 36

the consciousness will never know them. A few only are harmonious, and consequently at once useful and beautiful, and they will be capable of affecting the geometrician's special sensibility I have been speaking of; which, once aroused, will direct our attention upon them, and will thus give them the opportunity of becoming conscious.

37 This is only a hypothesis, and yet there is an observation which tends to confirm it. When a sudden illumination invades the mathematician's mind, it most frequently happens that it does not mislead him. But it also happens sometimes, as I have said, that it will not stand the test of verification. Well, it is to be observed almost always that this false idea, if it had been correct, would have flattered our natural instinct for mathematical elegance.

38 Thus it is this special aesthetic sensibility that plays the part of the delicate sieve of which I spoke above, and this makes it sufficiently clear why the man who has it not will never be a real discoverer.

39 All the difficulties, however, have not disappeared. The conscious ego is strictly limited, but as regards the subliminal ego, we do not know its limitations, and that is why we are not too loth to suppose that in a brief space of time it can form more different combinations than could be comprised in the whole life of a conscient being. These limitations do exist, however. Is it conceivable that it can form all the possible combinations, whose number staggers the imagination? Nevertheless this would seem to be necessary, for if it produces only a small portion of the combinations, and that by chance, there will be very small likelihood of the *right* one, the one that must be selected, being found among them.

40 Perhaps we must look for the explanation in that period of preliminary conscious work which always precedes all fruitful unconscious work. If I may be permitted a crude comparison, let us represent the future elements of our combinations as something resembling Epicurus's hooked atoms. When the mind is in complete repose these atoms are immovable; they are, so to speak, attached to the wall. This complete repose may continue indefinitely without the atoms meeting, and, consequently, without the possibility of the formation of any combination.

41 On the other hand, during a period of apparent repose, but of unconscious work, some of them are detached from the wall and set in motion. They plough through space in all directions, like a swarm of gnats, for instance, or, if we prefer

a more learned comparison, like the gaseous molecules in the kinetic theory of gases. Their mutual collisions may then produce new combinations.

What is the part to be played by the preliminary conscious work? Clearly it is to liberate some of these atoms, to detach them from the wall and set them in motion. We think we have accomplished nothing, when we have stirred up the elements in a thousand different ways to try to arrange them, and have not succeeded in finding a satisfactory arrangement. But after this agitation imparted to them by our will, they do not return to their original repose, but continue to circulate freely. 42

Now our will did not select them at random, but in pursuit of a perfectly definite aim. Those it has liberated are not, therefore, chance atoms; they are those from which we may reasonably expect the desired solution. The liberated atoms will then experience collisions, either with each other, or with the atoms that have remained stationary, which they will run against in their course. I apologize once more. My comparison is very crude, but I cannot well see how I could explain my thought in any other way. 43

However it be, the only combinations that have any chance of being formed are those in which one at least of the elements is one of the atoms deliberately selected by our will. Now it is evidently among these that what I called just now the *right* combination is to be found. Perhaps there is here a means of modifying what was paradoxical in the original hypothesis. 44

Yet another observation. It never happens that unconscious work supplies *ready-made* the result of a lengthy calculation in which we have only to apply fixed rules. It might be supposed that the subliminal ego, purely automatic as it is, was peculiarly fitted for this kind of work, which is, in a sense, exclusively mechanical. It would seem that, by thinking overnight of the factors of a multiplication sum, we might hope to find the product ready-made for us on waking; or, again, that an algebraic calculation, for instance, or a verification could be made unconsciously. Observation proves that such is by no means the case. All that we can hope from these inspirations, which are the fruits of unconscious work, is to obtain points of departure for such calculations. As for the calculations themselves, they must be made in the second period of conscious work which follows the inspiration, and in which the results of the inspiration are verified and the consequences deduced. The rules of these calculations are strict and complicated; they demand discipline, 45 attention, will, and consequently consciousness. In the subliminal ego, on the contrary, there reigns what I would call liberty, if one could give this name to the mere absence of discipline and to disorder born of chance. Only, this very disorder permits of unexpected couplings.

46 I will make one last remark. When I related above some personal observations, I spoke of a night of excitement, on which I worked as though in spite of myself. The cases of this are frequent, and it is not necessary that the abnormal cerebral activity should be caused by a physical stimulant, as in the case quoted. Well, it appears that, in these cases, we are ourselves assisting at our own unconscious work, which becomes partly perceptible to the overexcited consciousness, but does not on that account change its nature. We then become vaguely aware of what distinguishes the two mechanisms, or, if you will, of the methods of working of the two egos. The psychological observations I have thus succeeded in making appear to me, in their general characteristics, to confirm the views I have been enunciating.

47 Truly there is great need of this, for in spite of everything they are and remain largely hypothetical. The interest of the question is so great that I do not regret having submitted them to the reader.

from

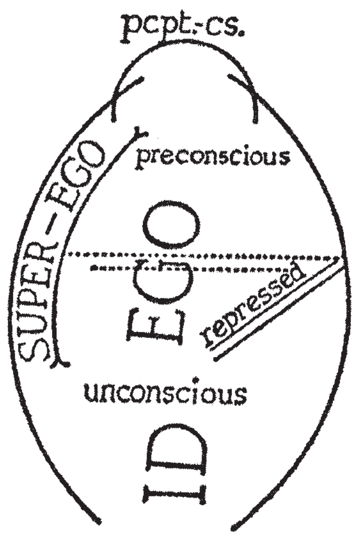
In Search of Memory: The Emergence of a New Science of Mind

by Eric R. Kandel

CHAPTER 4
ONE CELL AT A TIME

I entered Harry Grundfest's laboratory at Columbia University for a six-month elective period in the fall of 1955, hoping to learn something about higher brain functions. I did not anticipate embarking on a new career, a new way of life. But my very first conversation with Grundfest gave me reason to reflect. In that conversation I described my interest in psychoanalysis and my hope of learning something about where in the brain the ego, the id, and the superego might be located. 1

My desire to find these three psychic agencies had been sparked by a diagram Freud published in the course of summarizing his new structural theory of mind, which he developed in the decade 1923 to 1933 (figure 4-1). That new theory maintained his earlier distinction between conscious and unconscious mental functions, but it added three interacting psychic agencies: the ego, the id, and the superego. Freud saw consciousness as the *surface* of the mental apparatus. Much 2 of our mental function is submerged below that surface, Freud argued, just as the bulk of an iceberg is submerged below the surface of the ocean. The deeper a mental function lies below the surface, the less accessible it is to consciousness. Psychoanalysis provided a way of digging down to the buried mental strata, the preconscious and the unconscious components of the personality.



4-1 Freud's structural theory. *Freud conceived of three main psychic structures—the ego, the id, and the superego. The ego has a conscious component (perceptual consciousness, or* pcpt.-cs.*) that receives sensory information and is in direct contact with the outside world, as well as a preconscious component, an aspect of unconscious processing that has ready access to consciousness. The ego's unconscious components act through repression and other defenses to inhibit the instinctual urges of the id, the generator of sexual and aggressive instincts. The ego also responds to the pressures of the superego, the largely unconscious carrier of moral values. The dotted lines indicate the divisions between those processes that are accessible to consciousness and those that are completely unconscious. (From* New Introductory Lectures on Psychoanalysis *[1933]).*

3 What gave Freud's new model a dramatic turn was the three interacting psychic agencies. Freud did not define the ego, the id, and the superego as either conscious or unconscious, but as differing in cognitive style, goal, and function.

According to Freud's structural theory, the ego (the "I," or autobiographical 4
self) is the executive agency, and it has both a conscious and an unconscious component. The conscious component is in direct contact with the external world through the sensory apparatus for sight, sound, and touch; it is concerned with perception, reasoning, the planning of action, and the experiencing of pleasure and pain. In their work, Hartmann, Kris, and Lowenstein emphasized that this conflict-free component of the ego operates logically and is guided in its actions by the reality principle. The unconscious component of the ego is concerned with psychological defenses (repression, denial, sublimation), the mechanisms whereby the ego inhibits, channels, and redirects both the sexual and the aggressive instinctual drives of the id, the second psychic agency.

The id (the "it"), a term that Freud borrowed from Friedrich Nietszche, is 5
totally unconscious. It is not governed by logic or by reality but by the hedonistic principle of seeking pleasure and avoiding pain. The id, according to Freud, represents the primitive mind of the infant and is the only mental structure present at birth. The superego, the third governor, is the unconscious moral agency, the embodiment of our aspirations.

Although Freud did not intend his diagram to be a neuroanatomical map 6
of mind, it stimulated me to wonder where in the elaborate folds of the human brain these psychic agencies might live, as it had earlier stimulated the curiosity of Kubie and Ostow. As I mentioned, these two psychoanalysts with a keen interest in biology had encouraged me to study with Grundfest.

Grundfest listened patiently as I told him of my rather grandiose ideas. 7
Another biologist might well have dismissed me, wondering what to do with this naïve and misguided medical student. But not Grundfest. He explained that my hope of understanding the biological basis of Freud's structural theory of mind was far beyond the grasp of contemporary brain science. Rather, he told me, to understand mind we needed to look at the brain one cell at a time.

One cell at a time! I initially found those words demoralizing. How could one 8
address psychoanalytic questions about the unconscious motivation of behavior, or the action of our conscious life, by studying the brain on the level of single nerve cells? But as we talked I suddenly remembered that in 1887, when Freud began his own career, he had sought to solve the hidden riddles of mental life by studying the brain one nerve cell at a time. Freud started out as an anatomist, studying single

nerve cells, and had anticipated a key point of what later came to be called the neuron doctrine, the view that nerve cells are the building blocks of the brain. It was only later, after he began treating mentally ill patients in Vienna, that Freud made his monumental discoveries about unconscious mental processes.

9 I found it ironic and remarkable that I was now being encouraged to take that journey in reverse, to move from an interest in the top-down structural theory of mind to the bottom-up study of the signaling elements of the nervous system, the intricate inner worlds of nerve cells. Harry Grundfest offered to guide me into this new world.

[. . .]

CHAPTER 28
CONSCIOUSNESS

1 Psychoanalysis introduced us to the unconscious in its several forms. Like many scientists now working on the brain, I have long been intrigued by the biggest question about the brain: the nature of consciousness and how various unconscious psychological processes relate to conscious thought. When I first talked with Harry Grundfest about Freud's structural theory of mind—the ego, the id, and the superego—the central focus of my thinking was: How do conscious and unconscious processes differ in their representation in the brain? But only recently has the new science of mind developed the tools for exploring this question experimentally.

2 To develop productive insights into consciousness, the new science of mind first had to settle on a working definition of consciousness as a state of perceptual awareness, or selective attention writ large. At its core, consciousness in people, is an awareness of self, an awareness of being aware. Consciousness thus refers to our ability not simply to experience pleasure and pain but to attend to and reflect upon those experiences, and to do so in the context of our immediate lives and our life history. Conscious attention allows us to shut out extraneous experiences and focus on the critical event that confronts us, be it pleasure or pain, the blue of the sky, the cool northern light of a Vermeer painting, or the beauty and calm we experience at the seashore.

Understanding consciousness is by far the most challenging task confronting 3
science. The truth of this assertion can best be seen in the career of Francis Crick, perhaps the most creative and influential biologist of the second half of the twentieth century. When Crick first entered biology, after World War II, two great questions were thought to be beyond the capacities of science to answer: What distinguishes the living from the nonliving world? And what is the biological nature of consciousness? Crick turned first to the easier problem, distinguishing animate from inanimate matter, and explored the nature of the gene. By 1953, after just two years of collaboration, he and Jim Watson had helped solve that mystery. As Watson later described in *The Double Helix*, "at lunch Francis winged into the Eagle [Pub] to tell everyone within hearing distance that we had found the secret of life." In the next two decades, Crick helped crack the genetic code: how DNA makes RNA and RNA makes protein.

In 1976, at age sixty, Crick turned to the remaining scientific mystery: 4
the biological nature of consciousness. This he studied for the rest of his life in partnership with Christof Koch, a young computational neuroscientist. Crick brought his characteristic intelligence and optimism to bear on the question; moreover, he made consciousness a focus of the scientific community, which had previously ignored it. But, despite almost thirty years of continuous effort, Crick was able to budge the problem only a modest distance. Indeed, some scientists and philosophers of mind continue to find consciousness so inscrutable that they fear it can never be explained in physical terms. How can a biological system, a biological machine, they ask, feel anything? Even more doubtful, how can it think about itself?

These questions are not new. They were first posed in Western thought during 5
the fifth century B.C. by Hippocrates and by the philosopher Plato, the founder of the Academy in Athens. Hippocrates was the first physician to cast superstition aside, basing his thinking on clinical observations and arguing that all mental processes emanate from the brain. Plato, who rejected observations and experiments, believed that the only reason we can think about ourselves and our mortal body is that we have a soul that is immaterial and immortal. The idea of an immortal soul was subsequently incorporated into Christian thought and elaborated upon by St. Thomas Aquinas in the thirteenth century. Aquinas and later religious thinkers

held that the soul—the generator of consciousness—is not only distinct from the body, it is also of divine origin.

6 In the seventeenth century, René Descartes developed the idea that human beings have a dual nature: they have a body, which is made up of material substance, and a mind, which derives from the spiritual nature of the soul. The soul receives signals from the body and can influence its actions but is itself made up of an immaterial substance that is unique to human beings. Descartes' thinking gave rise to the view that actions like eating and walking, as well as sensory perception, appetites, passions, and even simple forms of learning, are all mediated by the brain and can be studied scientifically. Mind, however, is sacred and as such is not a proper subject of science.

7 It is remarkable to reflect that these seventeenth-century ideas were still current in the 1980s. Karl Popper, the Vienna-born philosopher of science, and John Eccles, the Nobel laureate neurobiologist, espoused dualism all of their lives. They agreed with Aquinas that the soul is immortal and independent of the brain. Gilbert Ryle, the British philosopher of science, referred to the notion of the soul as "the ghost in the machine."

8 Today, most philosophers of mind agree that what we call consciousness derives from the physical brain, but some disagree with Crick as to whether it can ever be approached scientifically. A few, such as Colin McGinn, believe that consciousness simply cannot be studied, because the architecture of the brain poses limitations on human cognitive capacities. In McGinn's view, the human mind may simply be incapable of solving certain problems. At the other extreme, philosophers such as Daniel Dennett deny that there is any problem at all. Dennett argues, much as neurologist John Hughlings Jackson did a century earlier, that consciousness is not a distinct operation of the brain; rather, it is the combined result of the computational workings of higher-order areas of the brain concerned with later stages of information processing.

9 Finally, philosophers such as John Searle and Thomas Nagel take a middle position, holding that consciousness is a discrete set of biological processes. The processes are accessible to analysis, but we have made little headway in understanding them because they are very complex and represent more than the sum of their parts. Consciousness is therefore much more complicated than any property of the brain that we understand.

Searle and Nagel ascribe two characteristics to the conscious state: unity 10
and subjectivity. The unitary nature of consciousness refers to the fact that our experiences come to us as a unified whole. All of the various sensory modalities are melded into a single, coherent, conscious experience. Thus when I approach a rosebush in the botanical garden at Wave Hill near my house in Riverdale, I sniff the exquisite fragrance of the blossoms at the same time that I see their beautiful red color—and I perceive this rosebush against the background of the Hudson River and the cliffs of the Palisade mountain ridge behind it. My perception is not only whole during the moment I experience it, it is also whole two weeks later, when I engage in mental time travel to recapture the moment. Despite the fact that there are different organs for smell and vision, and that each uses its own individual pathways, they converge in the brain in such a way that my perceptions are unified.

The unitary nature of consciousness poses a difficult problem, but perhaps not 11
an insurmountable one. This unitary nature can break down. In a surgical patient whose brain is severed between the two hemispheres, there are two conscious minds, each with its own unified percept.

Subjectivity, the second characteristic of conscious awareness, poses the 12
more formidable scientific challenge. Each of us experiences a world of private and unique sensations that is much more real to us than the experiences of others. We experience our own ideas, moods, and sensations directly, whereas we can only appreciate another person's experience indirectly, by observing or hearing about it. We therefore can ask, Is your response to the blue you see and the jasmine you smell—the meaning it has for you—identical to my response to the blue I see and the jasmine I smell and the meaning these have for me?

The issue here is not one of perception per se. It is not whether we each see 13
a very similar shade of the same blue. That is relatively easy to establish by recording from single nerve cells in the visual system of different individuals. The brain does reconstruct our perception of an object, but the object perceived—the color blue or middle C on the piano—appears to correspond to the physical properties of the wavelength of the reflected light or the frequency of the emitted sound. Instead, the issue is the significance of that blue and that note for each of us. What we do not understand is how electrical activity in neurons gives rise to the meaning we ascribe to that color or that wavelength of sound. The fact that conscious experience is unique to each person raises the question of whether it is possible

to determine objectively any characteristics of consciousness that are common to everyone. If the senses ultimately produce experiences that are completely and personally subjective, we cannot, the argument goes, arrive at a general definition of consciousness based on personal experience.

14 Nagel and Searle illustrate the difficulty of explaining the subjective nature of consciousness in physical terms as follows: Assume we succeed in recording the electrical activity of neurons in a region known to be important for consciousness while the person being studied carries out some task that requires conscious attention. For example, suppose we identified the cells that fire when I look at and become aware of a red image of the blossoms on a rosebush at Wave Hill. We have now taken a first step in studying consciousness—namely, we have found what Crick and Koch have called the neural correlate of consciousness for this one percept. For most of us, this would be a great advance because it pinpoints a material concomitant of conscious perception. From there we could go on to carry out experiments to determine whether these correlates also meld into a coherent whole, that is, the background of the Hudson River and the Palisades. But for Nagel and Searle, this is the easy problem of consciousness. The hard problem of consciousness is the second mystery, that of subjective experience.

15 How is it that I respond to the red image of a rose with a feeling that is distinctive to me? To use another example, what grounds do we have for believing that when a mother looks at her child, the firing of cells in the region of the cortex concerned with face recognition accounts for the emotions she feels and for her ability to summon the memory of those emotions and that image of her child?

16 As yet, we do not know how the firing of specific neurons leads to the subjective component of conscious perception, even in the simplest case. In fact, according to Searle and Nagel, we lack an adequate theory of how an objective phenomenon, such as electrical signals in the brain, can cause a subjective experience, such as pain. And because science as we currently practice it is a reductionist, analytical view of complicated events, while consciousness is irreducibly subjective, such a theory lies beyond our reach for now.

17 According to Nagel, science cannot take on consciousness without a significant change in methodology, a change that would enable scientists to identify and analyze the elements of subjective experience. Those elements are likely to be basic components of brain function, much as atoms and molecules

are basic components of matter, but to exist in a form we cannot yet imagine. The reductions performed routinely in science are not problematic, Nagel holds. Biological science can readily explain how the properties of a particular type of matter arise from the objective properties of the molecules of which it is made. What science lacks are rules for explaining how subjective properties (consciousness) arise from the properties of objects (interconnected nerve cells).

Nagel argues that our complete lack of insight into the elements of subjective 18
experience should not prevent us from discovering the neural correlates of consciousness and the rules that relate conscious phenomena to cellular processes in the brain. In fact, it is only by accumulating such information that we will be in a position to think about the reduction of something subjective to something physical and objective. But to arrive at a theory that supports this reduction, we will first have to discover the elements of subjective consciousness. This discovery, says Nagel, will be enormous in its magnitude and its implications, requiring a revolution in biology and most likely a complete transformation of scientific thought.

The aim of most neural scientists working on consciousness is much more 19
modest than this grand perspective would imply. They are not deliberately working toward or anticipating a revolution in scientific thought. Although they must struggle with the difficulties of defining conscious phenomena experimentally, they do not see those difficulties as precluding all experimental study under existing paradigms. Neural scientists believe, and Searle for one agrees with them, that they have been able to make considerable progress in understanding the neurobiology of perception and memory without having to account for individual experience. For example, cognitive neural scientists have made advances in understanding the neural basis of the perception of the color blue without addressing the question of how each of us responds to the same blue.

What we do not understand is the hard problem of consciousness—the 20
mystery of how neural activity gives rise to subjective experience. Crick and Koch have argued that once we solve the easy problem of consciousness, the unity of consciousness, we will be able to manipulate those neural systems experimentally to solve the hard problem.

The unity of consciousness is a variant of the binding problem first identified 21
in the study of visual perception. An intimate part of my experiencing the subjective pleasure of the moment at Wave Hill is how the look and the smell of roses in the

gardens are bound together and unified with my view of the Hudson, the Palisades, and all the other component images of my perception. Each of these components of my subjective experience is mediated by different brain regions within my visual and olfactory and emotional systems. The unity of my conscious experience implies that the binding process must somehow connect and integrate all of these separate areas in the brain.

22 As a first step toward solving the easy problem of consciousness, we need to ask whether the unity of consciousness—a unity thought to be achieved by neural systems that mediate selective attention—is localized in one or just a few sites, which would enable us to manipulate them biologically. The answer to this question is by no means clear. Gerald Edelman, a leading theoretician on the brain and consciousness, has argued effectively that the neural machinery for the unity of consciousness is likely to be widely distributed throughout the cortex and thalamus. As a result, Edelman asserts, it is unlikely that we will be able to find consciousness through a simple set of neural correlates. Crick and Koch, on the other hand, believe that the unity of consciousness will have direct neural correlates because they most likely involve a specific set of neurons with specific molecular or neuroanatomical signatures. The neural correlates, they argue, probably require only a small set of neurons acting as a searchlight: the spotlight of attention. The initial task, they argue, is to locate within the brain that small set of neurons whose activity correlates best with the unity of conscious experience and then to determine the neural circuits to which they belong.

23 How are we to find this small population of nerve cells that could mediate the unity of consciousness? What criteria must they meet? In Crick and Koch's last paper (which Crick was still correcting on his way to the hospital a few hours before he died, on July 28, 2004), they focused on the claustrum, a sheet of brain tissue that is located below the cerebral cortex, as the site that mediates unity of experience. Little is known about the claustrum except that it connects to and exchanges information with almost all of the sensory and motor regions of the cortex as well as the amygdala, which plays an important role in emotion. Crick and Koch compare the claustrum to the conductor of an orchestra. Indeed, the neuroanatomical connections of the claustrum meet the requirements of a conductor; it can bind together and coordinate the various brain regions necessary for the unity of conscious awareness.

The idea that obsessed Crick at the end of his life—that the claustrum is the spotlight of attention, the site that binds the various components of any percept together—is the last in a series of important ideas he advanced. Crick's enormous contributions to biology (the double helical structure of DNA, the nature of the genetic code, the discovery of messenger RNA, the mechanisms of translating messenger RNA into the amino acid sequence of a protein, and the legitimizing of the biology of consciousness) put him in a class with Copernicus, Newton, Darwin, and Einstein. Yet his intense, lifelong focus on science, on the life of mind, is something he shares with many in the scientific community, and that obsession is symbolic of science at its best. The cognitive psychologist Vilayanur Ramachandran, a friend and colleague of Crick's, described Crick's focus on the claustrum during his last weeks: 24

> Three weeks prior to his death I visited him in his home in La Jolla. He was eighty-eight, had terminal cancer, was in pain, and was on chemotherapy; yet he had obviously been working away nonstop on his latest project. His very large desk—occupying half the room—was covered by articles, correspondence, envelopes, recent issues of *Nature*, a laptop (despite his dislike of computers), and recent books on neuroanatomy. During the whole two hours that I was there, there was no mention of his illness—only a flight of ideas on the neural basis of consciousness. He was especially interested in a tiny structure called the claustrum which, he felt, had been largely ignored by mainstream pundits. As I was leaving he said: "Rama, I think the secret of consciousness lies in the claustrum—don't you? Why else would this tiny structure be connected to so many areas in the brain?"—And gave me a sly, conspiratorial wink. It was the last time I saw him.

Since so little is known about the claustrum, Crick continued, he wanted to start an institute to focus on its function. In particular, he wanted to determine whether the claustrum is switched on when unconscious, subliminal perception of a given stimulus by a person's sensory organs turns into a conscious percept. 25

One example of such switching that intrigued crick and Koch is binocular rivalry. Here, two different images—say, vertical stripes and horizontal stripes— 26 are presented to a person simultaneously in such a way that each eye sees only one set of stripes. The person may combine the two images and report seeing a plaid, but more commonly the person will see first one image, then the next, with horizontal and vertical stripes alternating back and forth spontaneously.

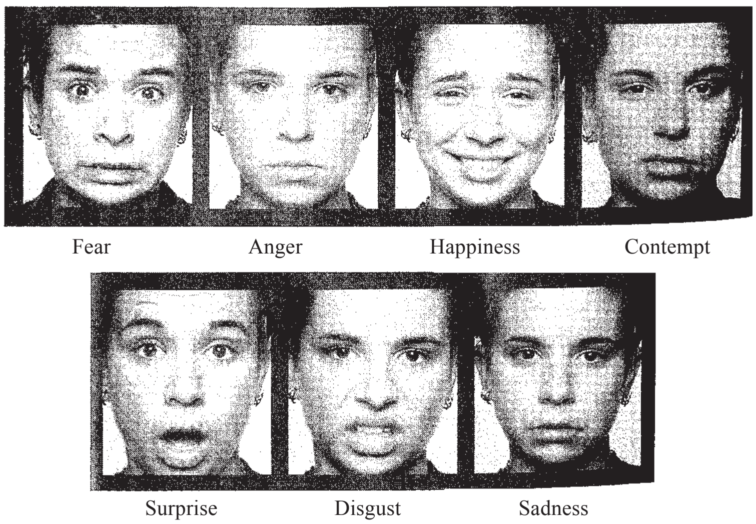
27 Using MRI, Eric Lumer and his colleagues at University College, London have identified the frontal and parietal areas of the cortex as the regions of the brain that become active when a person's conscious attention switches from one image to another. These two regions have a special role in focusing conscious attention on objects in space. In turn, the prefrontal and posterior parietal regions of the cortex seem to relay the decision regarding which image is to be enhanced to the visual system, which then brings the image into consciousness. Indeed, people with damage to the prefrontal cortex have difficulty switching from one image to the other in situations of binocular rivalry. Crick and Koch might argue that the frontal and parietal areas of the cortex are recruited by the claustrum, which switches attention from one eye to the other and unifies the image presented to conscious awareness by each eye.

28 As these arguments make clear, consciousness remains an enormous problem. But through the efforts of Edelman on the one hand, and Crick and Koch on the other, we now have two specific and testable theories worthy of exploration.

29 As someone interested in psychoanalysis, I wanted to take the Crick-Koch paradigm of comparing unconscious and conscious perception of the same stimulus to the next step: determining how visual perception becomes endowed with emotion. Unlike simple visual perception, emotionally charged visual perception is likely to differ between individuals. Therefore, a further question is, How and where are unconscious emotional perceptions processed?

30 Amit Etkin, a bold and creative M.D.-Ph.D. student, and I undertook a study in collaboration with Joy Hirsch, a brain imager at Columbia, in which we induced conscious and unconscious perceptions of emotional stimuli. Our approach paralleled in the emotional sphere that of Crick and Koch in the cognitive sphere. We explored how normal people respond consciously and unconsciously to pictures of people with a clearly neutral expression or an expression of fear on their faces. The pictures were provided by Peter Ekman at the University of California, San Francisco.

Ekman, who has cataloged more than 100,000 human expressions, was able to show, as did Charles Darwin before him, that irrespective of sex or culture, conscious perceptions of seven facial expressions—happiness, fear, disgust, contempt, anger, surprise, and sadness—have virtually the same meaning to everyone (figure 28-1). We therefore argued that fearful faces should elicit a similar response from the healthy young medical and graduate student volunteers in our study, regardless of whether they perceived the stimulus consciously or unconsciously. We produced a conscious perception of fear by presenting the fearful faces for a long period, so people had time to reflect on them. We produced unconscious perception of fear by presenting the same faces so rapidly that the volunteers were unable to report which type of expression they had seen. Indeed, they were not even sure they had seen a face! 31



28-1 Ekman's seven universal facial expressions. *(Courtesy of Paul Ekman.)*

Since even normal people differ in their sensitivity to a threat, we gave all of the volunteers a questionnaire designed to measure background anxiety. In contrast to the momentary anxiety most people feel in a new situation, background anxiety reflects an enduring baseline trait. 32

Not surprisingly, when we showed the volunteers pictures of faces with fearful expressions, we found prominent activity in the amygdala, the structure 33 deep in the brain that mediates fear. What was surprising was that conscious and unconscious stimuli affected different regions of the amygdala, and they did so to differing degrees in different people, depending on their baseline anxiety.

34 Unconscious perception of fearful faces activated the basolateral nucleus. In people, as in mice, this area of the amygdala receives most of the incoming sensory information and is the primary means by which the amygdala communicates with the cortex. Activation of the basolateral nucleus by unconscious perception of fearful faces occurred in direct proportion to a person's background anxiety: the higher the measure of background anxiety, the greater the person's response. People with low background anxiety had no response at all. Conscious perception of fearful faces, in contrast, activated the dorsal region of the amygdala, which contains the central nucleus, and it did so regardless of a person's background anxiety. The central nucleus of the amygdala sends information to regions of the brain that are part of the autonomic nervous system—concerned with arousal and defensive responses. In sum, unconsciously perceived threats disproportionately affect people with high background anxiety, whereas consciously perceived threats activate the fight-or-flight response in all volunteers.

35 We also found that unconscious and conscious perception of fearful faces activates different neural networks outside the amygdala. Here again, the networks activated by unconsciously perceived threats were recruited only by the anxious volunteers. Surprisingly, even unconscious perception recruits participation of regions within the cerebral cortex.

36 Thus viewing frightening stimuli activates two different brain systems, one that involves conscious, presumably top-down attention and one that involves unconscious, bottom-up attention, or vigilance, much as a signal of salience does in explicit and implicit memory in *Aplysia* and in the mouse.

37 These are fascinating results. First, they show that in the realm of emotion, as in the realm of perception, a stimulus can be perceived both unconsciously and consciously. They also support Crick and Koch's idea that in perception, distinct areas of the brain are correlated with conscious and unconscious awareness of a stimulus. Second, these studies confirm biologically the importance of the psychoanalytic idea of unconscious emotion. They suggest that the effects of anxiety are exerted most dramatically in the brain when the stimulus is left to the imagination rather than when it is perceived consciously. Once the image of

a frightened face is confronted consciously, even anxious people can accurately appraise whether it truly poses a threat.

A century after Freud suggested that psychopathology arises from conflict occurring on an unconscious level and that it can be regulated if the source of the conflict is confronted consciously, our imaging studies suggest ways in which such conflicting processes may be mediated in the brain. Moreover, the discovery of a correlation between volunteers' background anxiety and their unconscious neural processes validates biologically the Freudian idea that unconscious mental processes are part of the brain's system of information processing. While Freud's ideas have existed for more than one hundred years, no previous brain-imaging study had tried to account for how differences in people's behavior and interpretations of the world arise from differences in how they unconsciously process emotion. The finding that unconscious perception of fear lights up the basolateral nucleus of the amygdala in direct proportion to a person's baseline anxiety provides a biological marker for diagnosing an anxiety state and for evaluating the efficacy of various drugs and forms of psychotherapy. 38

In discerning a correlation between the activity of a neural circuit and the unconscious and conscious perception of a threat, we are beginning to delineate the neural correlate of an emotion—fear. That description might well lead us to a scientific explanation of consciously perceived fear. It might give us an approximation of how neural events give rise to a mental event that enters our awareness. Thus, a half century after I left psychoanalysis for the biology of mind, the new biology of mind is getting ready to tackle some of the issues central to psychoanalysis and consciousness. 39

One such issue is the nature of free will. Given Freud's discovery of psychic determinism—the fact that much of our cognitive and affective life is unconscious—what is left for personal choice, for freedom of action? 40

A critical set of experiments on this question was carried out in 1983 by Benjamin Libet at the University of California, San Francisco. Libet used as his starting point a discovery made by the German neuroscientist Hans Kornhuber. In his study, Kornhuber asked volunteers to move their right index finger. He then measured this voluntary movement with a strain gauge while at the same time recording the electrical activity of the brain by means of an electrode on the skull. After hundreds of trials, Kornhuber found that, invariably, each movement was 41 preceded by a little blip in the electrical record from the brain, a spark of free will! He called this potential in the brain the "readiness potential" and found that it occurred 1 second before the voluntary movement.

42 Libet followed up on Kornhuber's finding with an experiment in which he asked volunteers to lift a finger whenever they felt the urge to do so. He placed an electrode on a volunteer's skull and confirmed a readiness potential about 1 second before the person lifted his or her finger. He then compared the time it took for the person to will the movement with the time of the readiness potential. Amazingly, Libet found that the readiness potential appeared not after, but 200 milliseconds before a person felt the urge to move his or her finger! Thus by merely observing the electrical activity of the brain, Libet could predict what a person would do before the person was actually aware of having decided to do it.

43 This finding has caused philosophers of mind to ask: If the choice is determined in the brain before we decide to act, where is free will? Is our sense of willing our movements only an illusion, a rationalization after the fact for what has happened? Or is the choice made freely, but not consciously? If so, choice in action, as in perception, may reflect the importance of unconscious inference. Libet proposes that the process of initiating a voluntary action occurs in an unconscious part of the brain, but that just before the action is initiated, consciousness is recruited to approve or veto the action. In the 200 milliseconds before a finger is lifted, consciousness determines whether it moves or not.

44 Whatever the reasons for the delay between decision and awareness, Libet's findings also raise the moral question: How can one be held responsible for decisions that are made without conscious awareness? The psychologists Richard Gregory and Vilayanur Ramachandran have drawn strict limits on that argument. They point out that "our conscious mind may not have free will, but it does have free won't." Michael Gazzaniga, one of the pioneers in the development of cognitive neuroscience and a member of the American Council of Bioethics, has added, "Brains are automatic, but people are free." One cannot infer the sum total of neural activity simply by looking at a few neural circuits in the brain.

Text 9

from

The Shorter Science and Civilisation in China

by Joseph Needham

10 THE FUNDAMENTAL IDEAS OF CHINESE SCIENCE

We now approach a field of vital importance for the history of scientific thought in China: the fundamental ideas and theories that were worked out from the earliest times by the Chinese naturalists. The subject conveniently divides into three sections: first, the theory of the Five Elements (*wu hsing*); secondly, that of the Two Fundamental Forces (*Yin* and *Yang*); thirdly, the scientific, or to be more precise, the proto-scientific use of that elaborate symbolic structure, the Book of Changes (*I Ching*). In the light of modern research, our discussion will differ considerably from the Chinese traditions taken over rather uncritically by the early Western sinologists, and it will be as well if we preface it by a glance at the origin and development of some of the Chinese words most important for scientific thought. 1

Origins of Some of the Most Important Chinese Scientific Words

Before any science can develop there must be a suitable stock of words, and now, due to the discovery of the Anyang oracle-bones already described (Chapter 4), 2

Table 8. ***Ideographic etymologies of some of the words important in scientific thinking***

No.	Word	Modern Chinese romani-sation	Modern Chinese character	Ancient oracle-bone, bronze, or seal, form	K no. (see notes, p. 199)	Remarks
1a	affirmatory final particle (equivalent to 'x *is* y')	*yeh*	也		4	(a) Drawing of a cobra-like serpent. The semantic link, if any, would have been 'affirmation of danger'. The word for serpent, *shê* 蛇, is certainly related to it. But this explanation is less widely accepted than the following.
					4	(b) Drawing of the female external genitalia, the vulva. The semantic link, if any, would have been 'gate of being', hence 'affirmations of Being', and all lesser affirmations of qualities and attributes contained therein. This explanation was never challenged throughout Chinese history in spite of centuries of Confucian prudery.
1b	affirmatory verb-noun, to be, is, existence	*shih*	是		866c	Drawing of the sun with a foot and other strokes below it, probably composing the word *chêng* 正, 'correct, straight, fair and square'; not illusory. Thus 'that which exists under the sun'. Vision was here taken as representative of all the other means by which we collect sense-data.
2a	negative particle, 'not'	*pu*	不		999	Drawing of a flower-head on a stalk with two drooping leaves; thus, as regards sense, a borrowed homophone. The traditional explanation (Hsü Shen) was that it was an abstract concept symbol, i.e. a bird soaring aloft and *not* allowing itself to be caught.
2b	negative verb-noun, not to be, is not, non- existence	*fei*	非		579	Traditionally explained as the lower part of the word *fei* 飛, to fly, itself an old drawing of a bird; therefore two wings (or perhaps birds) back to back, i.e. *not* facing each other. Therefore (if Hsü was right) an abstract concept symbol.
		(*fei*, to fly)				There are, of course, a number of other words signifying the affirmative and the negative with various nuances.

3	different	*i*	異		954	A frontal and linear representation of a man with arms raised protecting his head or making a gesture of respect. The latter meaning is found in bronze inscriptions; perhaps the gesture had reference to the assumed effulgence of a noble interlocutor. If the character is not purely a borrowed homophone, social difference between lords and people may thus have led to the idea of 'otherness, strangeness, difference' in general. The head itself is drawn in an unusual exaggerated way, possibly to represent a mask.
4	like, similar to	*ju*	如		94g	The 'woman' and 'month' radicals combined. A very early borrowed homophone or phonetic loan-word. No archaic significance.
5	if	*jo*	若		777	A person kneeling, perhaps gathering plants. Some have thought that submissiveness is implied, hence, 'to be harmonious, to concur, complaisant', hence, by further extension, 'granted that, if. . .'. But it is more likely to be purely a borrowed homophone.
6	change, permutation	*i*	易		850	Drawing of a lizard, the meaning being derived either from colour-changes (cf. the chameleon), or rapid shifts of position.
7	change, especially gradual change, and change of form	*pien*	變		1780	Apparently not found in bone or bronze inscriptions, therefore of relatively late invention. The meaning of the drawing is uncertain, but it contains two hanks of silk and Hsü Shen said that it meant 'to bring into order', as in spinning or reeling. The radical, placed below, shows a hand holding a stick, signifying 'movement, action'. If the character is not purely a phonetic loan-word, it may have implied change from disorder to order.

Table 8 ***(cont.)***

No.	Word	Modern Chinese: romanisation	Modern Chinese: character	Ancient oracle-bone, bronze, or seal, form	K no. (see notes, p. 199)	Remarks
8	change, especially sudden change, and change of substance	*hua*	化		19	Drawing of two knives, i.e. coins of knife-money. Currency exchange would thus have given rise to one expression of the idea of change in general. Cf. no. 27.
9	origin, first	*yuan*	元		257	Drawing of a figure of a man in profile, with emphasis on the head, therefore 'first, beginning', the head being the most important part of the body. Since the ancient people doubtless knew that the head grows faster than other parts of the body during the embryonic life of vertebrates, and is relatively larger then than later, there may be an echo of primitive biological knowledge here.
10	cause, to rely on, following	*yin*	因		370	Drawing of a mat with woven pattern. Hence a basis, 'something to be relied on'; the meaning being extended from the static to the temporal. It may be noted that the same drawing occurs again in *hsiu* 宿, the resting-place for the night, a term of importance in astronomy.
11	cause, reason, fact	*ku*	故		49*i*	The left-hand side of the ancient bronze graph is the radical meaning 'old, ancient'; its significance is not exactly known, but it originates from a drawing of a shield stored in an open rack. The right-hand side shows the hand holding the stick, symbolising action. The general meaning is clearly 'precedent' or 'prior action'.
12	make, do, act	*wei*	為		27	Drawing of an elephant, with a man's hand on its trunk, symbolising prehensility and dexterousness.

13	begin	*shih*	始		976*p*, *e*′, *g*′, *h*′	Drawing of a foetus (upside down) and a woman. Closely related to *thai* 胎, womb, and embryo. Hence here it probably signified a female embryo, a 'beginning of beginnings'.
14	go, move	*hsing*	行		748	Diagram of a crossroads.
15	go away, deprive, send away	*chhü*	去		642	Drawing of a rice-basket covered with a lid. A homophone borrowed for the present meaning.
16	come to, reach, attain	*chih*	至		413	An arrow hitting its target, or the ground.
17	stop	*chih*	止		961	Drawing of a human foot.
18	end, finished, exhausted	*chin*	盡		381	Drawing of a hand cleaning out a vessel with a brush.
19	true, the truth, real	*chen*	真		375	Seal form of uncertain representational significance, but (as we know from all the related derivative words) almost surely implying 'full, filled up, solid'. Hence the derived meaning of truth as opposed to 'empty, unreal'. The drawing of a full sack standing on a stool, if that is what it is, would thus be an abstract concept symbol.
20	above, to ascend, to hand up	*shang*	上		726	Geometrical pictograph.
21	below, to descend, to hand down	*hsia*	下		35	Geometrical pictograph.

K no. from B. Kalgren's etymological dictionary *Grammatica Serica* in *Bulletin of the Museum of Far Eastern Antiquities*, Stockholm, 1940, vol. 12, p. 1 (Photographically reproduced as a separate volume, Peking, 1941).

as well as the characters inscribed on Shang and Chou bronze vessels, there is a copious graphic vocabulary for study. Not all the characters have been identified, but even so there are enough for a selection to be made of those ideographs that throw light on the origins of Chinese scientific terms. It is true that the ancient words probably had little influence on the thinking of the exponents of the proto-sciences in Chhin and Han times, but for us these early ideographs are of interest in themselves because they help us to understand the ancient Chinese approach to science. A selection of some of the important words is to be found in the accompanying table (Table 8), where the English word is followed by the modern pronunciation of its Chinese equivalent, then the Chinese character, next its ancient form, and finally a brief explanation of its archaic significance.

3 A study of the table shows that the fundamental terms necessary for the beginnings of science were formed just as one might expect, given the principle of the ideograph. Only two (numbers 20 and 21) can be regarded as purely geometrical symbols; the remaining seventy-eight are drawings of one kind or another. Of these one still defies analysis, at least eight are borrowed like-sounding words (homophones), and three or four concern abstract ideas, but otherwise the characters depict natural objects, the human body and its parts, and human activities. This might be expected, but what is remarkable about them is that characters concerned with technology and communication are the most numerous of all. Such a bias might change if we analysed a larger sample, but even so we can see how from everyday life ideographs were developed which could later acquire quite abstract meanings. They demonstrate, too, how technical terminology for everyday thinking and experimentation actually arose.

[. . .]

The Stabilised Five-Element Theory

13 We are now in a position to consider the Five-Element theory as it was finalised in Han times, and so handed down to all later ages. Two aspects merit special attention: the Enumeration Orders and the Symbolic Correlations.

14 The *Enumeration Orders* are orders in which the five elements were named in various ancient and mediaeval presentations of the subject. They were far from always being the same, but the four most important were as follows:

i	The Cosmogonic Order	*w F W M E*
ii	The Mutual Production Order	*W F E M w*
iii	The Mutual Conquest Order	*W M F w E*
iv	The ‘Modern’ Order	*M W w F E*

where *w* is Water, *W* Wood, *F* Fire, *M* Metal and *E* Earth.

The Cosmogonic Order was the order in which the elements were supposed to have come into being. It begins with Water, thus echoing the recurrent emphasis in Chinese writings on Water as the primeval element. The Mutual Production Order was the order in which the elements were supposed to give rise to one another; it gave the seasons in correct order, beginning with Wood for the Spring and Water for Winter (Earth corresponding to a month situated in between summer and autumn). 15

The Mutual Conquest Order described the series in which each element was supposed to conquer its predecessor. It was, in a sense, the most venerable order since it was the one associated with the teaching of Tsou Yen himself. It was based on a logical sequence of ideas that had their basis in everyday scientific facts: for instance that Wood conquers Earth because, presumably, when in the form of a spade, it can dig up earth. Again, Metal conquers Wood since it can cut and carve it; Fire conquers Metal for it can melt or even vaporise it; Water conquers Fire because it can extinguish it; and, finally, Earth conquers Water because it can dam it and contain it —a very natural metaphor for people to whom irrigation and hydraulic engineering were so important. This order was also considered significant from the political point of view; it was put forward as an explanation of the course of history, with the implication that it would continue to apply in the future and was, therefore, useful for prediction. Lastly, there was the ‘Modern’ Order, of obscure significance. However, it is this order that has come down to us in Chinese colloquial speech, where everyone still learns ‘Metal, Wood, Water, Fire, Earth’, even in nursery rhymes. 16

There are two interesting secondary principles involved in the Enumeration Order, principles that concern Rate of Change; these are the Principles of Control and Masking. The Principle of Control was derived solely from the Mutual Conquest Order, and according to it a given process of conquest is said to be controlled by the element that conquers the conqueror. For example, Metal conquers Wood but Fire controls the process; Fire conquers Metal but Water controls the process, and 17 so on. The idea was used in fate-calculations, but nevertheless the Chinese were following in it perfectly logical paths of thought which, in our own time, have been found applicable in numerous fields of experimental science. For instance, in the ecological balance of animal species, the abundance of various forms that prey on one another in a sequence depends on their sizes and habits; increases in the numbers of a particular bird, for example, will indirectly benefit the population of greenfly because of the thinning effect on ('conquest of') ladybirds, which eat the greenfly but are themselves eaten ('controlled') by the birds.

18 The Chinese principle of control could of course be criticised on the grounds that as a cyclic process it meant that nothing could ever happen because every element would always inhibit another. But the Chinese never supposed that all elements were effectively present everywhere at the same time, so the criticism is purely formal. And when related to the Mutual Production Order (*W, F, E, M, w*) as well as the Mutual Conquest Order (*W, M, F, w, E*), it followed that the controlling element is always the one produced by the conquered element, and the system could carry on. For example, Wood conquers Earth in a process controlled by Metal, but Metal is a product of Earth, so there is bound to be a feedback. This idea could have social consequences as when the Confucians adopted it to prove that a son had the right to take revenge on the enemy of his father. But its primary connotations were scientific. The idea that something acting on something else destroys it, and yet, in so doing, is itself affected so as to bring about its own change or destruction was known in ancient alchemy. It is also familiar to the chemist of today, where the chemical reaction which comes to a stop because of an accumulation of reaction products is one example. Biochemistry has even more telling ones, e.g. one of the contractile proteins of our muscles, myosin, is itself the enzyme adenosine-triphosphatase that breaks down the substance bringing energy to it.

19 The second principle, the Principle of Masking, depends on both the Mutual Production and the Mutual Conquest Orders. It refers to the masking of a process of change by some other process which creates more of the material than is being destroyed, or makes it faster. Thus, Wood destroys (conquers) Earth, but Fire masks the process, since Fire will destroy Wood and make Earth (ash) at a greater rate than Wood can destroy Earth. Again, as with the Principle of Control, there are modern biological and ecological examples of this, as when the action of large carnivores in devouring lemmings in Norway is masked by other factors that enormously increase the lemming population.

It is important to note that in both these principles there lurks a strong 20
quantitative element: the conclusions depend on quantities, speeds and rates. This arose, perhaps, from questions raised simply by the Enumeration Orders themselves, for we should realise that the early Chinese thinkers were not satisfied with them. This can be seen, for instance, in the *Mo Ching* where a fragment of the criticisms of the Naturalists by the late Mohists is preserved:

> 'C The Five Elements do *not* perpetually overcome one another.
> CS The five are Metal, Water, Earth, Wood, and Fire. Quite apart [from any cycle] Fire naturally melts Metal, if there is enough Fire. Or Metal may pulverise a burning Fire to cinders, if there is enough Metal. Metal will store Water [but does not produce it]. Fire attaches itself to Wood [but is not produced from it].'*

This attack on Tsou Yen's Mutual Conquest theory may be a retort to his supercilious attitude to the logical studies of the Mo-Ming schools, but is interesting all the same as a demonstration of the quantitative approach in Mohist scientific thinking.

The Symbolic Correlations

The Five Elements gradually came to be associated with every conceivable category 21
of things in the universe that it was possible to classify in fives. Table 9 sets forth some of these, but it must be realised that this is only a selection.

If one divides the correlations into groups, as has been done in drawing up the 22
table, a careful analysis makes it seem likely that the different groups were compiled by different groups of scholars. The astronomical group, for instance, may go back as early as the ninth century B.C., though it seems to show definite evidence of the hand of the great astronomer Kan Tê of the fourth century B.C., and probably the astrologer Kan Chung-Kho as well, who was of the same family but three centuries later. Then there were the Naturalist groups associated with Tsou Yen—of which the Yin-Yang group and the group of human psycho-physical functions are given here—and, finally, two groups of particular scientific interest, one primarily agricultural and one mainly medical. A striking point about the agricultural groups is that they contain no reference

* [All brackets in the quotations in this text are Needham's.] e.d.

to rice in their lists of grains, although it does appear in a medical group (not given here); presumably then the agricultural correlations arose in northern China, or at an earlier date. Actually the medical philosophers had from very early times a sixfold series complementing or replacing the fivefold one of the scientific thinkers. There were thus six, not five, internal Yang organs of the body and six Yin ones, and many other examples of this classification can be found in medicine. Perhaps it originated from Babylonia where counting in sixties, not tens, was dominant; and perhaps the fivefold order was more indigenously Chinese.

23 As we might imagine, these correlations met with criticism, sometimes severe, because they led to many absurdities, as pointed out in the first century A.D. by Wang Chhung:

> 'The [cyclical] sign *yin* corresponds to Wood, and its proper animal is the tiger. *Hsü* corresponds to Earth, and its animal is the dog. *Chhou* and *wei* likewise correspond to Earth, *chhou* having as animal the ox, and *wei* having the sheep. Now Wood conquers Earth, therefore the tiger overcomes the dog, ox, and sheep. Again, *hai* goes with Water, its animal being the boar. *Ssu* goes with fire, having the serpent as its animal. *Tzu* also signifies Water, its animal being the rat. *Wu*, conversely, goes with Fire, and its animal manifestation is the horse. Now Water conquers Fire, therefore the boar devours the serpent, and horses, if they eat rats [are injured by] a swelling of their bellies. [So run the usual arguments.]
>
> 'However when we go into the matter more thoroughly, we find that in fact it very often happens that animals do not overpower one another as they ought to do on these theories. The horse is connected with *wu* (Fire), the rat with *tzu* (Water). If Water really conquers Fire, [it would be much more convincing if] rats normally attacked horses and drove them away. Then the cock is connected with *yu* (Metal) and the hare with *mao* (Wood). If metal really conquers wood, why do cocks not devour hares?'

This attack by Wang Chhung was part of the Chinese sceptical tradition, which will be discussed in the next chapter, but it has been quoted here because it is not untypical of the kind of strictures that appeared on the correlations as well as on the whole Five-Element theory. Yet in spite of such criticisms, it seems that in the beginning these correlations were helpful to scientific thought in China. They were certainly no worse than the Greek theory of the elements that dominated European mediaeval thinking, and it was only when they became over-elaborate and fanciful, too far removed from the observation of Nature, that they were positively harmful. As an example of their positive use, it is worth quoting from the *Mêng Chhi Pi Than* (Dream Pool Essays) of Shen Kua. Shen Kua, whose book appeared in A.D. 1086, was one of the most wide-ranging thinkers that China produced in any age, and his use of the theory is therefore of considerable interest:

> 'In the Chhien Shan district of Hsinchow there is a bitter spring which forms a rivulet at the bottom of a gorge. When its water is heated it becomes *tan fan* [bitter alum, literally "gall-alum"—probably impure copper sulphate]. When this is heated it gives copper. If this "alum" is heated for a long time in an iron pan, the pan is changed to copper. Thus Water can be transformed into Metal—an extraordinary change of substance.
>
> 'According to the (*Huang Ti Nei Ching*) *Su Wên* [the medical classic] there are five elements in the sky, and five elements on the Earth. The *chhi* of Earth, when in the sky, is moisture. Earth [we know] produces metal and stone [as ores in the mountains], and here we see that Water can also produce metal and stone. These instances are therefore proofs that the principles of the *Su Wên* are right.'

Of course Shen Kua had no clear understanding of chemical change since like everyone else at the time he was dominated by the Five-Element theory. In Europe at the same time and much later too, down to the seventeenth century, the chemical reaction that Shen Kua describes was taken as evidence of transmutation. In fact his description shows that he was a fine observer, and his account is probably the first record in any language of the precipitation of metallic copper by iron, with the consequent formation of iron sulphate.

Numerology and Scientific Thinking

Before we can leave the Five-Element system, two important examples of Chinese 24
naturalistic thinking must be mentioned. Both are contained in the *Ta Tai Li Chi* (Record

of Rites of the Elder Tai), a compilation made between A.D. 85 and 105, although the quotations given probably date from the second century B.C. The first runs:

> 'Sanchü Li asked Tsêng Tzu saying, "It is said Heaven is round and Earth square, is that really so?". . .
>
> 'Tsêng Tzu said, "That to which Heaven gives birth has its head on the upper side; that to which Earth gives birth has its head on the under side. The former is called round, the latter is called square. If Heaven were really round and the Earth really square the four corners of the Earth would not be properly covered. Come nearer and I will tell you what I learnt from the Master [Confucius]. He said that the Tao of Heaven was round and that of the Earth square. The square is dark and the round bright. The bright radiates *chhi*, therefore there is light outside it. The dark imbibes *chhi*, therefore there is light within it. Thus it is that Fire and the Sun have an external brightness, while Metal and Water have an internal brightness. That which irradiates is active, that which imbibes radiation is reactive. Thus the Yang is active and the Yin reactive.
>
> '"The seminal essence [*ching*] of the Yang is called *shen*. The germinal essence of the Yin is called *ling*. The *shen* and *ling* (vital forces) are the root of all living creatures; and the ancestors of [such high developments as] rites and music, human-heartedness and righteousness; and the makers of good and evil, as well as of social order and disorder.
>
> '"When the Yin and Yang keep precisely to their proper positions, then there is quiet and peace . . .
>
> '"Hairy animals acquire their coats before coming into the world, feathered ones similarly first acquire their feathers. Both are born of the power of Yang. Animals with carapaces and scales on their bodies likewise come into the world with them; they are born by the power of Yin. Man alone comes naked into the world; [this is because] he has the [balanced] essences of both Yang and Yin.
>
> '"The essence [or most representative example] of hairy animals is the unicorn, that of feathered ones is the phoenix [or pheasant]; that of the carapace-animals is the tortoise, and that of the scaly ones is the dragon. That of the naked ones is the Sage."'

This passage, which acts as an introduction to the next, is significant especially for the remark that the sage is the chief representative of the naked animals; this really is a supreme example of the fact that Chinese thought refused to separate man from Nature, or individual man from social man. But is also contains the view, which could not be bettered even by modern evolutionists, that the basic forces seen at work in the lowest creatures are the same as those which at higher levels will develop the highest manifestations of human social and ethical life. The second passage is almost entirely biological:

> 'The Master said, "[The Principle of] Change has brought into existence men, birds, animals and all the varieties of creeping things, some living solitary, some in pairs, some flying and some running on the ground. And no one knows how things seem to each of them. And he alone who profoundly scrutinises the virtue of the Tao can grasp their basis and their origin.
>
> '"Heaven is 1, Earth is 2, Man is 3. 3 × 3 makes 9. 9 × 9 makes 81. 1 governs the Sun. The Sun's number is 10. Therefore Man is born in the tenth month of development.
>
> '"8 × 9 makes 72. Here an even number follows after an odd one. Odd numbers govern time. Time governs the Moon. The Moon governs the horse. Therefore the horse has a gestation period of 11 months.
>
> '"7 × 9 makes 63. 3 governs the Great Bear [the Plough or Northern Dipper]. This constellation governs the dog. Therefore the dog is born after only 3 months.". . .
>
> 'Now birds and fishes are born under the sign of the Yin, but they belong to the Yang. This is why birds and fishes both lay eggs. Fishes swim in the waters, birds fly among the clouds. But in winter, the swallows and starlings go down into the sea and change into mussels.
>
> 'The habits of the various classes of animals are very different. Thus silkworms eat but do not drink, while cicadas drink but do not eat, and ephemeral gnats and flies do neither. Animals with scales and carapaces eat during the summer, and in winter hibernate. Animals with beaks [birds] have 8 openings of the body and lay eggs. Animals which masticate [mammals] have 9 openings of the body and nourish their young in wombs.'

Here we see that the Naturalists, or whoever it was who wrote it, were not only close observers of Nature, but also that their observations were fitted into a framework of number-mysticism. Traces of this mysticism were already evident in the table of symbolic connections (Table 9, pp. 154–5), and it appears to have exerted a fascination for a very long time, for although it began in the third century B.C., or even a little before, it was still active as late as the twelfth century A.D. This we shall have to take into account in our assessment of Chinese scientific ideas.

The Theory of the Two Fundamental Forces

25 So far, more has been said about the Five Elements and their symbolic correlations than about the two fundamental forces of the Yin and Yang. This is merely because we know more about the Five Elements than the Two Forces, which do not appear in the surviving fragments of Tsou Yen though his school was called the Yin–Yang Chia, and in later books discussion of them was generally credited to him. Nevertheless, there can be little doubt that the philosophical use of the terms began at the start of the fourth century B.C., and that passages in older texts which use them are later interpolations.

26 Some facts about Yin and Yang are, however, clear. We know, for instance, that the Chinese characters for Yin and Yang are connected with darkness and light. Yin involves graphs for hill (-shadows) and clouds, and the Yang character has slanting sun-rays or a flag fluttering in the sunshine (although the latter may represent someone holding a perforated jade disc, which was the symbol of Heaven, and probably the oldest of all astronomical instruments). These correspond well with the way in which the terms are used, for example, in the *Shih Ching* (Book of Odes), a collection of ancient folksongs. Here Yin evokes the idea of cold and cloud, of rain, of femaleness, of that which is inside, dark like the underground chambers where ice is kept for summer use. Yang, on the other hand, evokes the idea of sunshine and warmth, of spring and summer months, of maleness and brightness. Yin and Yang also had more factual meanings: Yin the shady side of a mountain or valley, Yang the sunny side.

27 As philosophical terms their explicit use appears in the appendix of the third-century B.C. classic, the *I Ching* (Book of Changes), which will be discussed in more detail later. Here there are sayings, the sense of which is that there are only

two fundamental forces or operations in the universe, now one dominating, now the other, in a wave-like succession. And there are other early mentions. The fourth-century B.C. *Mo Tzu* book, for instance, has two references, one where it is said that every living creature partakes of the nature of Heaven and Earth and the harmony of the Yin and the Yang, the other where the sage-kings are said to have brought the Yin and the Yang, the rain and the dew, in timely season. Then in that other fourth-century work, the *Tao Tê Ching* (Canon of Virtue of the Tao), it is said that living creatures are surrounded by Yin and envelop Yang, and that the harmony of their life processes depends on the harmony of the two forces.

When we move forward to the second century B.C., the time of the Han 28
Confucians, we come across some more specific remarks. In the words of Tung Chung-Shu:

> 'Heaven has Yin and Yang, so has man. When the Yin *chhi* of Heaven and Earth begins [to dominate], the Yin *chhi* of man responds by taking the lead also. Or if the Yin *chhi* of man begins to advance, the Yin *chhi* of Heaven and Earth must by rights respond to it by rising also. Their Tao is one. Those who are clear about this [know that] if the rain is to come, then the Yin must be activated and its influence set to work. If the rain is to stop, then the Yang must be activated and its influence set to work.'

We shall shortly see that the *I Ching* contains some vitally important material 29
giving us a deeper insight into the ideas of Chinese scientific thought, but since the Yin and the Yang are so basic a part of the text and its numerous appendices, it will be best to consider something of this aspect here. The *I Ching* contains a series of 64 symbolic hexagrams, each of which is composed of six lines, whole or broken, corresponding to the Yang and the Yin. Each hexagram is primarily Yin or primarily Yang, and by a judicious arrangement it was found possible to derive all the 64 in such a way as to produce alternating Yin and Yang. Figure 23 gives an example of this, showing how the Yang splits into two, then four, then eight, and so on until 64 alternations exist. And the process need not, of course, stop there; the 64 could be extended to 128, and so *ad infinitum*. The Yin and Yang components never become completely separated, but at each stage, in any given fragment, only one is manifested. Now this has considerable scientific interest because this splitting and

Table 9. ***The symbolic correlations***

Elements *hsing* 行	Seasons *shih* 時	Cardinal points *fang* 方	Tastes *wei* 味	Smells *chhou* 臭	Stems (denary cyclical signs) *kan* 干	Branches (duodenary cyclical signs) and the animals pertaining to them *chih* 支	Numbers *shu* 數
WOOD	spring	east	sour	goatish	*chia i* 甲 乙	*yin* 寅 (tiger) and *mao* 卯 (hare)	8
FIRE	summer	south	bitter	burning	*ping ting* 丙 丁	*wu* 午 (horse) and *ssu* 巳 (serpent)	7
EARTH	—[a]	centre	sweet	fragrant	*wu chi* 戊 己	*hsü* 戌 (dog), *chhou* 丑 (ox), *wei* 未 (sheep) and *chhen* 辰 (dragon)	5
METAL	autumn	west	acrid	rank	*kêng hsin* 庚 辛	*yu* 酉 (cock) and *shen* 申 (monkey)	9
WATER	winter	north	salt	rotten	*jen kuei* 壬 癸	*hai* 亥 (boar) and *tzu* 子 (rat)	6

Elements *hsing* 行	Musical notes *yin* 音	Hsiu *hsiu* 宿 (mansions)	Star-palaces *kung* 宮	Heavenly bodies *chhen* 辰	Planets *hsing* 星	Weather *chhi* 氣	States *kuo* 國
WOOD	*chio* 角	1–7	Azure Dragon	stars	Jupiter	wind	Chhi
FIRE	*chih* 徵	22–28	Vermilion Bird	sun	Mars	heat	Chhu
EARTH	*kung* 宮	—	Yellow Dragon	earth	Saturn	thunder	Chou
METAL	*shang* 商	15–21	White Tiger	*hsiu* constellations	Venus	cold	Chhin
WATER	*yü* 羽	8–14	Sombre Warrior	moon	Mercury	rain	Yen

Elements hsing 行	Rulers[b] ti 帝	Yin-Yang 陰 陽	Human psychophysical functions shih 事	Styles of government *chêng* 政	Ministries *pu* 部	Colours *ssu* 色	Instruments *chhi* 器
WOOD	Yü the Great [Hsia]	Yin in Yang or lesser Yang	deameanour	relaxed	Agriculture	green	compasses
FIRE	Wên Wang [Chou]	Yang or greater Yang	vision	enlightened	War	red	weights and measures
EARTH	Huang Ti [pre-dyn.]	Equal balance	thought	careful	the Capital	yellow	plumblines
METAL	Thang the Victorious [Shang]	Yang in Yin or lesser Yin	speech	energetic	Justice	white	T-squares
WATER	Chhin Shih Huang Ti [Chhin]	Yin or greater Yin	hearing	quiet	Works	black	balances

Elements *hsing* 行	Classes of living animals *chhung* 蟲	Domestic animals *shêng* 牲	'Grains' *ku* 穀	Sacrifices *ssu* 祀	Viscera *tsang* 臟	Parts of the body *thi* 體	Sense-organs *kuan* 官	Affective states *chih* 志
WOOD	scaly (fishes)	sheep	wheat	inner door	spleen	muscles	eye	anger
FIRE	feathered (birds)	fowl	beans	hearth	lungs	pulse (blood)	tongue	joy
EARTH	naked (man)	ox	panicled millet	inner court	heart	flesh	mouth	desire
METAL	hairy (mammals)	dog	hemp	outer door	kidney	skin and hair	nose	sorrow
WATER	shell-covered (invertebrates)	pig	millet	well	liver	bones (marrow)	ear	fear

[a] The sixth month was sometimes supposed to be under the sign of Earth.

[b] There are many variants of this list; the names given above are those which appear in the fragment from Tsou himself, see Needham, vol. 2, p.238, adding that of the First (Chhin) Emperor who believed his sway to be under the sign of water.

re-splitting of two factors, with one dominant and one recessive, has parallels in modern scientific thinking, e.g. in genetics. So here once again we have a parallel to what was said about the supposed interactions of the Five Elements—they lead to paths of thought which only in our own time have been seen to have a valid application to Nature. In brief, some elements of the structure of the world as modern science sees it are foreshadowed in the speculation of the early Chinese philosophers.

‘Associative’ Thinking and Its Significance

30 We have now reached a stage when we can see that the scientific ideas of the Chinese involved two fundamental principles, the Two Forces and the Five Elements. Basically the Two Forces, the Yin and Yang, were derived from the negative and positive projections of Man’s own sexual experience, while the Five Elements were believed to lie behind every substance and every process. With these Five Elements were associated or correlated everything in the universe susceptible of a fivefold arrangement, although, since not everything could be grouped in this way, there was a larger region comprising everything else that was classifiable but which would only go into some other order (fours, nines, twenty-eights, etc.). This wider approach gave rise to another aspect of Chinese thought—number-mysticism, of which the main purpose was to relate the various numerical categories with one another.

31 Most European observers have written off this number-mysticism as pure superstition, and have claimed that it prevented the rise of true scientific thinking in China. Some modern Chinese scientists have also been inclined to take the same view, but at least they had the excuse that they had to deal with many thousands of traditional Chinese scholars who, unschooled in any modern scientific view, still imagined that the ancient thought-system of China was a viable alternative. But we are not concerned with this last problem, with the modernisation of Chinese society, which is quite capable of modernising itself; we have to discover whether the ancient and traditional thought-system was merely superstition or simply a variety of primitive thought; whether, perhaps, it contained something characteristic of the civilisation which produced it, and if it did, whether it contributed something positive to other civilisations.

A study of primitive magic shows that this seems to have operated on the basis of two 'laws', the 'law of similarity' according to which like produces like, and the 'law of contagion' whereby things that have once been in contact, but are not so any longer, still continue to act upon one another. Yet these are just the kind of 'laws' that lie behind the Chinese correlations or associations, which fit in completely with the ideas of similarity and contagion. Immediately, then, we have a clue to the motive lying behind the compilation of their immense correlative lists, it was magical; a fact that need not disturb us since we have already noted that in early times it was magic that nourished science and that probably the earliest scientists were magicians. After all, though magic arises in a wide variety of ways from the mystical life from which it draws strength, it then mingles with the life of the ordinary man to serve him. It tends to the concrete, the factual; it deals with reality. Indeed, magic works in the same sense as techniques work, as operations in chemistry, or in industry; it is an art 32

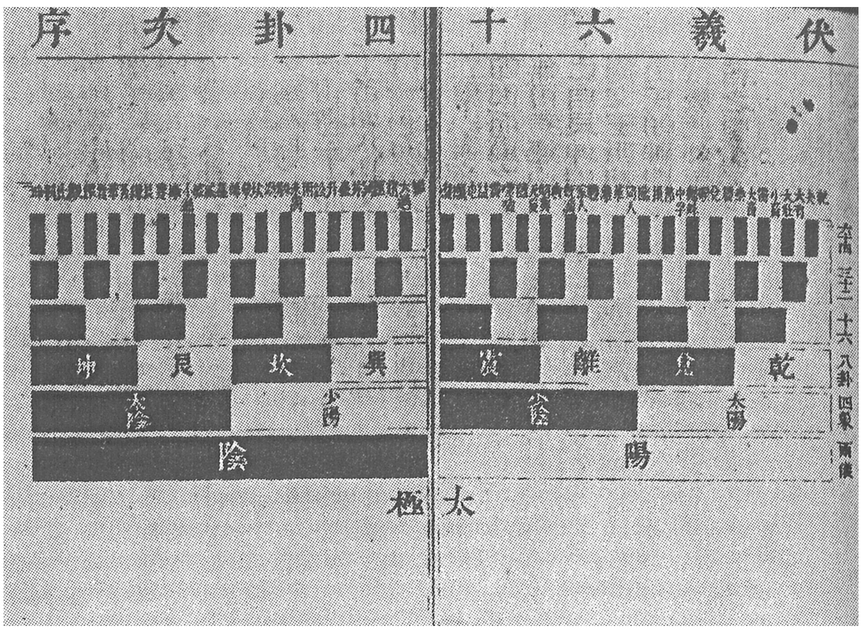


*Fig. 23. Segregation Table of the symbols of the Book of Changes (*Fu-Hsi Liu-shih-ssu Kua Tzhu Hsü*), from Chu Hsi's* Chou I Pên I Thu Shuo *(twelfth century A.D.). Yin and Yang separate, but each contains half of its opposite in a 'recessive' state, as is seen when the second division occurs. There is no logical end to the process but here it is not followed beyond the stage of the 64 hexagrams.*

of doing things. But there is little need to labour the point further, for it has already been stressed in the chapter on Taoism; what we see now, though, is that, because of this practical emphasis of magic, the symbolic correlations of the Chinese were just what a magician would need to practise his art: they brought order into things and did this at a time when no one could know what would bring success and what would not in any magical or experimental procedure. There had to be some guide to choosing the right conditions, and the correlations provided it. If one were experimenting or doing magic with water then it seemed logical, say, not to wear red, the colour of fire. Such associations might be more intuitive than anything else, but at the time what else could they be?

33 Modern scholars have called this kind of mental approach 'associative thinking' or 'co-ordinative thinking'. It is a system that works by association and intuition, and it has its own logic and its own laws of cause and effect. It is not just superstition, but a thought-form perfectly reasonable by its own standards, though, of course, it differs from the type of thinking characteristic of modern science, where the emphasis is on external causes. It does not classify its ideas in a series of ranks but side by side in a pattern. Things influence one another not by mechanical causes but by a kind of induction effect.

34 In the Chinese thought with which we are dealing, the key words are *Order* and *Pattern*; or one might almost say there is only one key-word, *Organism*, for certainly the symbolic correlations, the correspondences, the hexagrams of the *I Ching*, all formed part of one gigantic whole. Things behaved in particular ways not necessarily because of the prior actions of other things, but primarily because their position in the ever-changing cyclical universe was such that they were endowed with intrinsic natures which made such behaviour natural for them. If they did not behave in those particular ways they would lose their positions, and their relations to other things (which made them what they were) would alter and would turn them into something other than themselves. Their existence depended on the whole world-organism, and they reacted on one another by a kind of mysterious resonance.

35 Nowhere is this better expressed than by Tung Chung-Shu in the *Chhun Chhiu Fan Lu* (String of Pearls on the Spring and Autumn Annals) of the second century B.C. In a chapter entitled 'Things of the Same Genus Energise Each Other' we read:

> 'If water is poured on level ground it will avoid the parts which are dry and move towards those that are wet. If [two] identical pieces of firewood are exposed to the fire, the latter will avoid the damp and ignite the dry one. All things reject what is different [to themselves] and follow what is akin. Thus it is that if [two] *chhi* are similar, they will coalesce; if notes correspond they resonate. [The experimental proof of this is extraordinarily clear. Try tuning musical instruments.] The *kung* note or the *shang* note struck upon one lute will be answered by the *kung* or the *shang* notes from other stringed instruments. They sound by themselves. There is nothing miraculous, but the Five Notes being in relation; they are what they are according to the Numbers [whereby the world is constructed].
>
> '[Similarly] lovely things summon others among the class of lovely things; repulsive things summon others among the class of repulsive things. This arises from the complementary way in which a thing of the same class responds—as for instance if a horse whinnies another horse whinnies in answer, and if a cow lows, another cow lows in response.
>
> 'When a great ruler is about to arise auspicious omens first appear; when a ruler is about to be destroyed, there are baleful ones beforehand. Things indeed summon each other, like to like, a dragon bringing rain, a fan driving away heat. . .
>
> 'It is not only the two *chhi* of the Yin and Yang which advance and retreat, according to their categories. Even the origins of the varied fortunes, good and bad, of men, behave in the same way. There is no happening that does not depend for its beginning upon something prior, to which it responds because [it belongs to the same] category, [*lei*].'

The classification which Tung Chung-Shu uses is the capacity of various things in the universe to fit into a fivefold, or some other, numerical grouping. And it is interesting that he takes the acoustic resonance of stringed instruments as an example of this, for to those who knew nothing of sound waves it must have seemed very convincing, proving his point that things in the cosmos that belonged to the same class resonated with, or energised, one another. He did not, of course, take the very primitive view that anything could affect anything else: his relationships were part of a closely knit universe with selective effects. Indeed, to Tung Chung-Shu, and to his successors, causation was something very special, since it acted in a sort of stratified pattern, not at random. Nothing was uncaused, but nothing was caused mechanically. The organism of the universe was such that everything fitted into its place and acted according to an eternal dramatic cycle; if anything missed its cue it would cease to exist. But nothing ever did fail in this way. And here, as elsewhere in Chinese thought, the text shows that the regularity of natural processes is conceived of, not as a government of law, but of mutual adaptations to community life. Not only in human relationships but throughout the world of Nature, there was give and take, a kind of mutual courtesy rather than strife among inanimate powers and processes. Solutions were found by compromise.

36 If all this expresses something deeply true about the Chinese world-picture—and there is much to make us believe that it does—then the fivefold correlations are an abstract chart of the whole thought-system, and the scholars of Han and later times were far from being stuck in the mud of 'primitive thought'. In true primitive thought anything can be the cause of anything else: everything is credible; nothing is impossible or absurd. If a steamship with one funnel more than usual calls at a small seaport and an epidemic follows, the appearance of the steamship is just as likely as anything else to be regarded as the cause. But once things are categorised, as they were in the fivefold system, then anything can in no way be the cause of anything else.

37 With these thoughts in mind, we are driven to the conclusion that there are two ways of advancing from primitive truth. One was the way taken by some of the Greeks: to refine the ideas of causation in such a way that one ended up with a mechanical explanation of the universe, just as Democritus did with his atoms. The other way is to systematise the universe of things and events into a structural pattern which conditioned all the mutual influences of its different parts. On the Greek world-view, if a particle of matter occupied a particular place at a particular time, it was because another particle had pushed it there. On the other view, the particle's behaviour was governed by the fact that it was taking its place in a 'field of force' alongside other particles that are similarly responsive: causation here is not 'responsive' but 'environmental'.

38 The Greek Democritean approach may have been a necessary prelude to modern science, but that does not mean that the criticism of the Chinese view as mere superstition is correct; far from it. The idea that things belonging to the

same class resonated with, or energised, each other was echoed also in Greece. Aristotle, for instance, claimed that there were three kinds of 'motion': movement in space was explained by asserting that like attracts like; growth by asserting that like nourishes like; change of quality by saying that like affects like. This view, and others akin to it, echo the earliest Greek philosophers who spoke of 'love' and 'hatred' in natural phenomena. But the point to be emphasised here is that while Greek thought as a whole moved away from these views towards concepts of mechanical cause and effect, Chinese thought developed the organic concept. It is a mistake, and a serious one, to think of this Chinese outlook on Nature as essentially primitive. It was a precisely ordered universe, not governed either by the fiat of a supreme creator-lawgiver nor by the inexorable clashes of atoms, but by a harmony of wills, spontaneous but ordered in patterns, rather like the dancers in a country dance, none of whom are bound by law or pushed by the others, but who co-operate voluntarily. If the Moon stood in a certain constellation of stars at a certain time, it did so not because anyone ordered it to do so, nor yet because it was obeying some regularity, some isolatable cause, which could be expressed mathematically. It did so because it was the nature of the pattern of the universal organism that it should do so, and for no other reason. Looking back down the long avenues of time, we see the universe of Newton at the end of the Democritean view, but we do not find an emptiness at the other; instead there is the modern 'philosophy of organism' that stems from the twentieth-century mathematician and philosopher A. N. Whitehead, and of which we shall have more to say shortly.

The contrast between the two views of the universe, the traditional Chinese 39
and that generally accepted by modern science, comes out very clearly in their use of numbers. A large amount of creditable mathamatics was, of course, done in China, as we shall see, but the point at issue is the Chinese use of number-mysticism or numerology in connection with their associative thinking. Numerology—the kind of thing epitomised in the nineteenth-century fancies associated with the Great Pyramid, where the lengths and intersections of passages are taken as providing the dates of future events—is utterly distasteful to the modern scientific mind. In China too it seems to have contributed little of scientific value, but equally important, it does not appear to have had any really bad effect either; indeed, it can be claimed that even the most exaggerated 'numerical' correlations of the Five Elements were valid in their way. Certainly they played their part in the development of Chinese

scientific thinking, just as in Europe extravagances like the legal trials of animals for misdemeanours foreshadowed the conception of Laws of Nature.

40 Time and space were also looked on differently in the East and the West. For the ancient Chinese time was not a purely abstract quantity, but divided into separate seasons, each with their own subdivisions. Nevertheless there was continuity too, because time flowed in one direction only, and in China there was never any disposition to adopt the cyclical recurrent time of Indian philosophers, even though their beliefs were known. Space was not something abstractedly uniform, extending in all directions, but was divided into separate regions—south, north, east, west and centre—each connected with time and with the Five Elements into 'correspondences'. The east was indissolubly connected with the spring and with Wood, the south with summer and Fire, and so on. This compartmented world was quite similar to that of mediaeval Europe before Galileo and Newton had extended geometrical space and universal gravitation to the entire cosmos.

41 For the ancient Chinese, things were connected rather than caused, as Tung Chung-Shu put it in the second century B.C.:

> 'The constant course of Nature is that things in opposition to each other cannot both arise simultaneously. The Yin and Yang [for example] move parallel to each other, but not along the same road; they meet one another, and each in turn operates as controller. Such is their pattern.'

The universe is a vast organism, with now one component, now another, taking the lead at any one time, with all the parts co-operating in a mutual service which is perfect freedom.

42 In such a system as this, causality is not like a chain of events, but rather like what the modern biologist calls the 'endocrine orchestra' of mammals where, though all the endocrine glands work, it is not easy to find which element is taking the lead at any one time. And we should be clear about it; modern science needs concepts like this when considering questions like the higher nervous centres of mammals and even Man himself. But leaving modern science aside, it is clear that the concept of causality where the idea of succession was subordinated to that of interdependence dominated Chinese thinking.

[. . .]

Text 10a

Why the Scientific Revolution Did Not Take Place in China —or Didn't It?[*]

by Nathan Sivin

This essay makes the case for two conclusions. First, why the scientific revolution 1
did not take place in China is not a question that historical research can answer. It becomes a useful question primarily when one locates the fallacies that lead people to ask it. Second, a scientific revolution, by the criteria that historians of science use, did take place in China in the eighteenth century. It did not, however, have the social consequences that we assume a scientific revolution will have. The most obvious conclusion is that those assumptions are mistaken.[1]

* This is a revised version (revised on 24 August, 2005) of an essay first published in *Chinese Science*, 1982, 5: 45–66, and often anthologized.

1 This essay incorporates my current views on a historical issue to which I have returned regularly for some years. No doubt my views on this topic will be different in another decade; all I mean to accomplish with these ephemeral reflections is to transmit the idea that the issue is worth thinking about, to suggest how one might think about it, and to point out that certain ways of thinking about it are so burdened by suspect assumptions that they do not encourage clear explanation. I have addressed one aspect or another in previous writings, to which the reader is referred for documentation: "Copernicus in China," *Studia Copernicana*, 1973,

Originally published in *Chinese Science*, 1982, 5: 45–66, revised 2005.8.24, used by permission of the author.

2 Anyone who has looked into the history of science, technology, and medicine in the last generation or so has been aware that all the great civilizations of the ancient world had their own sophisticated traditions. The Chinese traditions, because they are recorded so fully, and because they were more independent of European influence than the Islamic and Indian ones, are particularly fascinating if we want to compare how understanding of Nature varies in different cultural circumstances. Beginning in the 1920's, Chinese and Japanese historians explained what the Chinese knew and did. My English colleague Joseph Needham, in the 1950's, began calling their work to the attention of educated people in the West, and encouraging them to add to it. By now the study of China is one of the most flourishing fields in the history of science, with perhaps a thousand specialists in China, Japan, Europe, the United States, and elsewhere.[2]

3 When people become aware of what we have turned up, they usually begin wondering why the transition to modern science first happened where it did. In 1969 Joseph Needham gave the "Scientific Revolution problem" its classic formulation: Why did modern science, the mathematization of hypotheses about Nature, with

---

6: 63–122; "Shen Kua" and "Wang Hsi-shan," *Dictionary of Scientific Biography, s.v.*; "Next Steps in Learning about Science from the Chinese Experience," *Proceedings*, XIVth International Congress of the History of Science (Tokyo and Kyoto, 19–27 August, 1974), I, 10–18; N. Sivin (Ed.), *Science and Technology in East Asia* (New York: Science History Publications, 1977), pp. xi–xxi; and "Chinesische Naturwissenschaft: Weber und Needham," in W. Schluchter (Ed.), *Max Webers Studie fiber Konfuzianismus und Taoismus. Interpretation und Kritik* (München, 1982). Joseph Needham has also provided a summary of our conversations and correspondence on the "Scientific Revolution Problem," an interesting attempt to specify differences and similarities in our views, in *Science and Civilisation in China*, Vol. 5, Part 2, "Spagyrical Discovery and Invention: Magisteries of Gold and Immortality" (Cambridge: Cambridge University Press, 1974), pp. xxii–xxvii. Here I set out my own somewhat different view of the divergences that accompany our very broad areas of agreement. In several points regarding the Scientific Revolution problem I have been anticipated by W. T. Chan, "Neo-Confucianism and Chinese Scientific Thought," *Philosophy East and West*, 1957, 6 (4): 309–332.

I use "Scientific Revolution" to refer primarily to the transition in the exact sciences between Galileo and Laplace and its wider repercussions by 1800. This is one of several definitions in current use. I adopt it for the purpose of this essay not because it is the best possible definition, but because it is the one most commonly presupposed by Sinologists and laymen who set out to compare developments in China and the West. Needham's usage of the term "Scientific Revolution" is often, but not consistently, broader. No definition is better than a historiographic expedient. Lack of a consensus about the significance of the term has led some historians of science to reject its use altogether.

2 [That was the case in 1982. By 2002, the number has probably doubled.]

all its implications for advanced technology, take its meteoric rise only in the West at the time of Galileo?" "Why modern science had not developed in Chinese civilization . . . ?" He adds a second question that makes the larger problem more interesting: "why, between the first century B.C. and the fifteenth century A.D., Chinese civilization was much more efficient than occidental in applying human natural knowledge to practical human needs."[3]

4 In that millennium and a half, European civilization was first experiencing a slow general collapse and then even more slowly recovering from it. It is obvious that we ought to be looking at the Western end of Eurasia, not the Far Eastern end, to account for European inferiority in technology over a span of 1400 years. But there are still other doubts to be expressed in connection with this second question, with its claim of Chinese superiority over many centuries. The natural knowledge that was being applied to human needs was not what we usually call Chinese science.

5 Early technology did not succeed or fail according to how well it applied the insights of early science. On the whole, it was members of the small educated class in China who did science, and passed down their understanding in books. Technology was a matter of craft and manufacturing skills that artisans privately transmitted to their children and apprentices. Most such artisans could not read the scientists' books. They had to depend on their own practical and esthetic knowledge. What that knowledge was like we can only reconstruct from the artifacts they left and from the scattered written testimony of literate people. Literacy spread considerably outside the elite over the last several centuries, but this did not lead to the substantial use of books to teach craft skills.

6 Comparing all of the scientific and engineering activity of one civilization with all that of another in a single generalization conceals more than it reveals, since it is only in modern times that these various kinds of work became closely connected. It is true that between the end of the Roman period and 1400 or so, a Chinese visiting Europe would have found it in many respects technologically backward. On the other hand, there was probably not a great deal to choose between Chinese and European medical practice before about 1850 (knowledge of anatomy

3 J. Needham, *The Grand Titration: Science and Society in East and West* (Toronto: University of Toronto Press, 1969), pp. 16 and 190. He first posed these questions in 1964.

and physiology had hardly any therapeutic applications earlier). Mathematical astronomy in China by its last high point about 1300 never quite reached the general level of predictive accuracy that Ptolemy had mastered eleven hundred years earlier.

7 I don't need to dwell on comparisons of this kind. They tell us nothing at all about what we can expect to learn from one culture or the other. After all, no one is going to propose that we stop studying the European tradition of alchemy just because the Chinese alchemical literature is richer in chemical knowledge.[4] What matters is that we are now able to begin comparing several strong traditions of science and technology based on the ideas and social arrangements of different civilizations. All of them have to be studied if we want to understand the general relations through history and across the globe between science and culture, science and society, science and politics, science and individual consciousness. Without that comparative understanding we will remain trapped in our own parochial viewpoints.[5] Historians have more urgent work to do than trying to prove that every other culture was inferior to the one they specialize in.

8 As an example of how studying the Chinese experience can suggest clues about the character of early science in general, let me dwell briefly on the case of Shen Kua (1031–1095), one of the most versatile figures in the history of Chinese science and engineering. Just to give a few examples, he is famous for the first discussion of magnetic declination and of printing with movable type, the only application of permutations in traditional Chinese mathematics, a proposal for daily records of the lunar and planetary positions, the first suggestion in East Asia of a purely solar calendar, an explanation of the process of land formation by both deposition of silt and erosion, and an important book on the theory and practice of medicine. In addition to his technical activities, his writing has to be consulted by every student of early Chinese archeology, music, art and literary criticism, economic theory, and diplomacy. He made his early reputation as a land reclamation expert. He was deeply involved as a high official in the 1060's in the most important political reform movement for some centuries.

---

4 See in particular J. Needham, *Science and Civilisation in China*, Vol. 5, Parts 2–4, (Cambridge: Cambridge University Press, 1974–1980).

5 For a sample of such work, see G. E. R. Lloyd and N. Sivin, *The Way and the Word. Science and Medicine in Early China and Greece* (New Haven: Yale University Press, 2004).

Shen's combination of unlimited curiosity and involvement in the affairs of his time had a special interest for my own education. For some time, through a series of studies roaming through different historic periods and technical disciplines, I have been trying to piece together bits of answers to a large question that I find boundlessly interesting. How did Chinese scientists in traditional times explain to themselves what they were doing? In other words, what was their understanding of nature and of their relation to it as conscious individuals living in a society? How did the insights of the various sciences hang together to form this understanding? I had gradually formed a general idea of the sciences as defined in early China, but I couldn't see how their insights were combined to form that general understanding. It occurred to me that I might do well to study how the sciences fit together in the mind of a person who was involved in all of them. The obvious person to study was Shen Kua. 9

The pattern that emerged wasn't unexpected, but I had to take stock of it for the first time. One aspect was that there does not seem to have been a systematic connection between all the sciences in the minds of the people who did them. They were not integrated under the dominion of philosophy, as schools and universities integrated them in Europe and Islam. They had sciences but no science, no single conception or word for the overarching sum of all of them. 10

In Shen Kua's memoirs there is a classification called "regularities underlying the phenomena." Under this heading he like many others grouped together physical and numerological aspects of astronomy, astrology, cosmology, and divination, which refract the pattern of physical reality in their various ways. A section called "technical skills" puts together medicine, engineering, and mathematics (including astronomical mathematics), because they share purely instrumental value. There they fit alongside architecture and games. His chapter on "strange occurrences" sets out his thoughts on the origin of plant fossils, the first recorded description of a tornado in East Asia, an account of his experiment on the formation of rainbows, and similar gems, all sitting alongside unlikely hearsay and ephemeral curiosities. 11

You can see that what makes us think of Shen Kua as a scientist was widely scattered through his own scheme of human knowledge. That scheme cohered not on the level of science, but on a much more general level. In his writing, there are no clear boundaries between material that fits the modern conception of science and material that doesn't. That modern conception does not help us to understand what Shen Kua was getting at. 12

13 Shen Kua, in the second half of the eleventh century, made his turn on the stage of history at a time when a great upsurge in social mobility was broadening the group that ruled China. Many of these new men were interested in all sorts of practical affairs that well-born people in earlier times would have considered beneath them. Civil servants were expected to be competent and versatile, and might work their way to the highest posts of the empire as financial specialists. At that time the government was basing merit ratings of officials on the bottom line—quantitative measures of efficiency in collecting taxes, reclaiming land, and so on, instead of entirely on virtue, breeding, and orthodoxy, as had been the case earlier. At leisure too, this large group that Shen Kua belonged to was free to indulge curiosity—in an amateur way, of course—about anything in the universe, including technical matters that earlier were fit only for clerks or artisans.

14 Only after Shen's lifetime did this evolving amateur ideal settle on philosophy, the arts, and literature as the appropriate realms to be universal within, once again leaving the study of the earth and sky largely to the mere technicians. In the eleventh century Shen was only one of a number of polymaths whose scientific and technological interests, however amateur, all emerged in connection with their varied official responsibilities. The intellectual consistency of Shen's style in scientific thought seems to reflect only the consistency of his public career, in which that style was formed. What connected his research interests, in other words, were the remarkably diverse responsibilities and commitments of his civil service appointments.[6]

15 The astronomer in the court computing calendars to be issued in the emperor's name, the doctor curing sick people in whatever part of society he was born into, the alchemist pursuing archaic secrets in mountain haunts of legendary teachers, had no reason to relate their arts to each other. Philosophers were in no position to define a common discipline for all of them, as Aristotle and his successors had done

6 On other scientists pertinent to this point, in addition to frequent references in *Science and Civilisation in China,* see S. Miyasita, "Su Sung," Vol. 3, pp. 969–970 in H. Franke (Ed.), *Sung Biographies* (3 Vols., Wiesbaden: Steiner, 1976); K. M. Teng 鄧廣銘 and C.T. Wang 王振鐸, "Su Sung〈蘇頌〉," pp. 123–134 in Institute for the History of Science, Chinese Academy of Sciences (Ed.), *Chung-kuo ku-tai k'o-hsueh-chia*《中國古代科學家》(Ancient Chinese Scientists, Beijing: Science Press, 1959); and Wang Chin-kuang 王錦光, "*Sung-tai k'ohsueh-chia Yen Su*〈宋代科學家燕肅〉(The Sung Scientist Yen Su)," *Hang-chou ta-hsueh hsueh-pao*《杭州大學學報》, 1979, 3: 34–38.

in Europe, and so philosophers had practically no influence on the development of these special pursuits.

If anyone was going to seek out the common ground of the sciences in China, 16
it was people like Shen Kua, who were mastering them all. But Shen put his own understanding together in ways that did not directly link the fields of Chinese science, and in ways that intimately associate what today would be considered scientific with what would be called grossly superstitious. That distinction simply gets in the way of understanding the way Shen Kua's thought was connected. Surely it is necessary to understand thought before one begins to label it.

I would have to say that I failed to find the internal unity of Chinese science 17
that I was looking for in the mind of Shen Kua. By way of compensation, I did learn the importance of an issue that I hadn't paid enough attention to before, that is, the relations of the sciences to other kinds of knowledge.

* * * * * * * * * *

Now back to the Scientific Revolution problem. It is striking that this question— 18
Why didn't the Chinese beat Europeans to the Scientific Revolution?—happens to be one of the few questions that people often ask publicly about why something didn't happen in history. It is analogous to the question of why your name did not appear on page 3 of today's newspaper. It belongs to an infinite set of questions that historians don't organize research programs around because they have no direct answers. They translate into questions about the rest of the world. The one that concerns us, for instance, translates into "in what circumstances did the Scientific Revolution take place in the seventeenth and eighteenth centuries in Western Europe?"

Why do people keep asking why the Scientific Revolution did not take place 19
in China when they know enough not to waste time explaining why their names did not appear on page 3 of today's newspaper? Because the question encourages exploration of a fascinating topic and provides some order for thinking about it. It is, in other words, heuristic. Heuristic questions are useful at the beginning of an inquiry. As we comprehend enough to deal with complicated patterns, heuristic questions tend to grow murky, and finally to lose their interest compared with the emerging clarity of what did happen.

20 So much for heuristic questions in general. Why do we tend to take this one more seriously than the general run? Somehow the Scientific Revolution problem holds a special urgency.

21 That urgency is there, I suggest, because this problem relies on certain Western assumptions, shaky assumptions that we do not feel comfortable about questioning. Above all we usually assume that the Scientific Revolution is what everybody ought to have had. But it is not at all clear that that is what everybody wanted before it became, in recent times, an urgent matter of survival amidst violent change. This change resulted from, among other things, the Scientific Revolution that did take place. In fact we have made very little progress so far in understanding how Europeans originally came to want that revolution in one country after another, since the attention of historians has been concentrated on how it took place.

22 There is usually the equally sentimental assumption that civilizations which had the potential for a scientific revolution ought to have had the kind that took place in the West, that led to the sorts of institutional and social changes that appeared in the West.

23 These assumptions are usually linked to a faith that European civilization—above all in its current American form—was somehow in touch with reality in a way no other civilization could be, and that its great share of the world's wealth and power comes from some intrinsic fitness to inherit the earth that was there all along. Historical study does not suggest that Europe by 1600 had a concentration of intelligence, imagination, talent, or virtue that no other civilization could match. It does suggest that the privileged position of the West comes instead from a head start in the technological exploitation of nature and the political exploitation of societies not technologically equipped to defend themselves.

24 Finally there is the conviction among scientists that, since science has so quickly and thoroughly become international, it transcends European historical and philosophic biases, and is as universal, objective, and value-free as the Nature that it seeks to understand and manipulate.

25 What seems to be common sense in that last assumption (or in the self-conception that all the articles of faith I have mentioned are part of) does not stand up to thoughtful examination. Modern science is still too marked by the special circumstances of its development in Europe to be considered universal.

26 Chinese science got along without dichotomies between mind and body, objective and subjective, even wave and particle. In the West the first two were

entrenched in scientific thought by the time of Plato. Galileo, Descartes, and others carried them into modern times to mark off the realm of physical science from the province of the soul, which was decidedly off limits to secular innovators like themselves. These distinctions let early modern scientists claim authority over the physical world on the ground that purely natural knowledge could not conflict with and therefore could not threaten the authority of established religion.

Science and religion have long since learned to coexist, but we are still 27
living with these sharp distinctions between mind and body and so on. If they are European peculiarities, and perpetual sources of trouble at that, why hasn't modern science managed to rid itself of them? It is evidently not a simple matter to root them out. Until we do, there is something to be said for frankly admitting a certain parochialism in the foundations of science. The mathematical equations may be universal, but the allocation of human effort among the possibilities of natural knowledge is not.

Science and technology have spread throughout the world, but that has not 28
made them universal, in the sense of transcending European patterns of thought. In one society after another the encounter between old and new ideas has been abortive, resolved by social change and political legislation. Traditional ideas are simply excluded (on the grounds that they are backward, superstitious, regressive, fit only for the lower classes, etc.) from the educational systems created to teach a new technical and managerial elite the values of technology alongside its theory and practice.

Modern technology is clearly more powerful than that of traditional societies; 29
but to a larger extent than we generally realize, its strength emerges in application to needs and expectations that do not exist until it generates them.[7] True universality would require modern technology to coexist with and serve cultural diversity rather than standardizing it out of existence.

I am arguing that the notion of a universal and value-free modern science, 30
which has somehow become independent of its social and historical origins, is wishful thinking. It is easy even for an intelligent reader to be led astray on this point. The narrow limits of the certainty from which this notion arises are never defined carefully by those who set out to explain science to non-scientists.

7 This point has been most persuasively argued in L. Winner, *Autonomous Technology: Technics-out-of-control as a Theme in Political Thought* (Cambridge: The MIT Press., 1977).

31 It would be foolish to deny that modern science has attained a verifiability, an internal consistency, a taxonomic grasp, a precision in accounting for physical phenomena, and an accuracy in prediction that no other kind of activity shares, and that lay far outside the grasp of early sciences. The rigor that makes these remarkable characteristics possible quickly disappears, however, once the formulation of a law or theory in mathematical equations, matrices of categories, or exactly defined technical concepts and models has been translated into the ordinary language and general discourse of a given culture. That translation into analogies and metaphors steeped in values must precede all public discussion of science, and almost all philosophic discussion. It even precedes most reflection by scientists on fields outside their own disciplines.

32 Beyond the narrow, abstract realm in which exactitude is possible, values and subjective judgments come to bear on every activity situated within a society. There are, for instance, profound differences between the character of modern scientific activity in the contemporary People's Republic of China and United States. They reflect different predominant convictions about the relations between basic and applied science, the relation of both to general culture, the roles of scientists in defining research programs, procedures for planning and supporting individuals' research projects, expectations about the social aims to which scientific work will contribute, the organization and status of professional scientists, the connections of political ideas and scientific knowledge, and the division of national resources between science and other priorities, and between various scientific activities. That certain equations and models are invariant between the two societies is a factor in all these consensuses, but then so is the ubiquity of opposable thumbs. Despite the invariance, a given constellation of values will determine that certain laws and hypotheses can be developed further, and that others will be abandoned unless they are among the very few that individuals can explore at their private discretion and their own expense. The great disparity in Chinese and American definitions of psychology is only one particularly obvious example that affects the life and death of particular theories in one society or the other.

33 So long as there is variation of such magnitude in the balance between the cognitive, practical, normative, and social dimensions of science, such words as "international" and "universal" are out of place. When applied to the narrow, rigorous technical realm of scientific cognition alone, they constitute a modest claim indeed.

Nor can one accept uncritically the idea that modern science is in every essential respect European in its social and historical origins. To those familiar with the science of other cultures, any account of the early history of science is lopsided, and misleading on the most fundamental issues, if it restricts itself substantially to discoveries made and understandings worked out at the Western end of Eurasia; if it loses sight of the constant movement of ideas back and forth between civilizations from the New Stone Age to the present; if it does not adequately consider what Europeans had learned by 1600 about Islamic, Indian, and Chinese science; or if it ignores the impact of exotic technologies and materials on the experiences of Europeans. 34

Fallacies of Historical Reasoning

Growing awareness of the high level of science and technology in ancient China has led to cascades and avalanches of hypotheses from one scholar or another about factors that inhibited the evolution of modern science in China, or characteristics unique to the West that made possible or furthered a major scientific revolution.[8] These often incorporate elementary fallacies of historical reasoning that deserve notice. 35

8 The most significant early contributions to this literature, in order of their appearance, are Jen Hungchün 任鴻雋, "Shuo Chung-kuo wu k'o-hsueh chih yuan-yin〈說中國無科學之原因〉" (The Reason for China's Lack of Science), *K'o-hsueh*《科學》, 1915, 1 (1): 8–13; Y. L. Fung, "Why China Has No Science—An Interpretation of the History and Consequences of Chinese Philosophy," *The International Journal of Ethics*, 1922, 32 (3): 237–263; H. H. Dubs, "The Failure of the Chinese to Produce Philosophical Systems," *T'oung Pao*《通報》, 1929, 26 (1–2): 96–109; D. Bodde, "The Attitude Toward Science and Scientific Method in Ancient China," *T'ien Hsia Monthly*, 1936, 2: 139–160; and R. Murphey, "The Nondevelopment of Science in Traditional China," *Papers on China*, 1947, 1: 1–30 (for others see the bibliographies of *Science and Civilisation in China*, esp. Vol. 2). Jen claims that science failed to develop in China after the Han period because of inattention to "the inductive method." Fung claims "it is because of the fact that the Chinese ideal prefers enjoyment to power that China has no need of science . . ." (p. 261). Dubs refutes the silly prejudice that the character of the Chinese language made systematic thought impossible, but argues that "the result of the absence of mathematical systems was that the Chinese philosophers attacked the world piecemeal . . . by empirical rather than by rational methods" (p. 108). He has nothing whatever to say about Chinese scientists. Bodde considers attitudes toward science, and is aware of a few isolated scientific accomplishments despite his disregard for the technical literature, but suggests that the most important "retarding effect upon scientific innovation . . . has been

36 For roughly two-thirds of a century, historians have argued that although Ch'ing dynasty thinkers took the world as observable, nominalistic fact, just as Francis Bacon (1561–1626) did, unlike him they did not develop a scientific methodology. Despite the positivist bias of such arguments, they did not even consider whether Bacon's scientific method has survived in the practice of contemporary science.[9] It was, in fact, largely Scholastic in its origins, concerned with taxonomies rather than theories of natural phenomena, and resolutely

---

the ideographic nature of the Chinese written language" (p. 158). Murphey, dependent upon Western-language sources and influenced by the stereotypes of F. S. C. Northrup, concludes "a naturalistic philosophy which might be called a reliance on the aesthetic continuum . . . clearly had no place for the inductive hypotheses necessary for science" (p. 15). Writings of this sort are full of acute observations, particularly about philosophic attitudes expressed in the early classics, but it is obvious that their authors failed to examine the literature of the Chinese scientific traditions. They may make a case that Lao-tzu or Hsun-tzu would have been mediocre biologists or mathematicians, but they do not help us account for the theoretical analyses, mathematical proofs, and programs of empirical discovery so profusely documented in the writings of those actually engaged in studies of nature.

Because of his knowledge of the Chinese sciences and the breadth of his hypotheses, Needham's is the earliest discussion of the Scientific Revolution problem that still commands attention, and is still the best. The most useful critiques of Needham's writings on this subject are, from Sinologists, D. Bodde, "Evidence for 'Laws of Nature' in Chinese Thought," *Harvard Journal of Asiatic Studies*, 1957 (publ. 1959), 20 (3–4): 709–727, and "Chinese 'Laws of Nature': A Reconsideration," ibid., 1979, 39 (1): 139–155, A. C. Graham, "China, Europe, and the Origins of Modern Science: Needham's The Grand Titration," pp. 45–69 in S. Nakayama and N. Sivin (Ed.), *Chinese Science. Explorations of an Ancient Tradition* (Cambridge: The MIT Press, 1973); from a historian of science, S. Nakayama, "Joseph Needham, Organic Philosopher," ibid., pp. 23–43; from a philosopher, R. S. Cohen, "The Problem of 19 (k)," *Journal of Chinese Philosophy*, 1973, 1: 103–117; and from sociologists, B. Nelson, "Sciences and Civilizations, 'East' and 'West.' Joseph Needham and Max Weber," *Boston Studies in the Philosophy of Science*, 1974, 11: 445–493, and S. Restivo, "Joseph Needham and the Comparative Sociology of Chinese and Modern Science," *Research in Sociology of Knowledge, Sciences and Art*, 1979, 2: 25–51. Kenneth G. Robinson sets to rest once and for all claims that literary Chinese is an inferior vehicle of science in "Literary Chinese as a Language for Science," in *Science and Civilisation in China*, J. Needham (Ed.), et al. (Cambridge: Cambridge University Press, 2004), Vol. 7, Part 2, pp. 95–198.

9 Jen, loc. cit.: J. R. Levenson, *Confucian China and Its Modern Fate. The Problem of Intellectual Continuity* (London: Routledge and Paul, 1958), pp. 3–14; and D. E. Mungello, "On the Significance of the Question 'Did China Have Science?'" *Philosophy East and West*, 1972, 22 (4): 467–478, and my comments on this article in the same journal, 1973, 23: 413–416. Fung also refers to Bacon in connection with the Scientific Revolution problem, but there the issue is not method but, more pertinently, the relation of science and power (Fung does not specify what kind of power).

unconcerned with mathematical measurement. Of the major early modern attempts to define how physical science might fruitfully proceed it was probably the most sterile, in contrast to Bacon's very influential convictions about the organization and ideology of scientific activity.

This pattern of thought, then, faults the Chinese for not developing a scientific 37
method that later proved abortive in the West. The same habit shows up in many other forms. A well-known sociological study of astronomy in the last two centuries B.C. explains the failure to develop a "unified scientific system." One reason is that Chinese astronomers "were not interested in applied technical sciences, e.g., in developing theoretical tools which could be used to control the flight of a cannon shell or to direct ships safely across the sea."[10] So much for the first civilization to note the declination of the compass needle. So much for the astronomy of an era more than a millennium before the invention of the cannon. The same lack of interest is prominent in the impetus theoreticians from John Philoponus (fl. ca. 530) to Jean Buridan (ca. 1295–ca. 1358) and others of the School of Paris whose investigations furnished much of the basis for Galilean mechanics. How then did what is presented as a disastrous shortcoming in China fail to prevent in Italy—in fact, according to the conventional wisdom, help directly to bring about—the mathematical study of bodies in motion?

Considered generally, this fallacy amounts to claiming that if an important 38
aspect of the European Scientific Revolution cannot be found in another civilization, the whole ensemble of fundamental changes could not have happened there. The flaw of reasoning that underlies it is the arbitrary assumption, never explicit, never discussed, that a given circumstance amounts to a necessary condition. It is almost invariably arbitrary because if we trace the prehistory of the actual Scientific Revolution backward far enough, in most cases we can find a point when the circumstance is absent in Europe. In that case, on what grounds can it be considered a necessary condition? In most cases one need not go back very far. That is why, despite their currency among Sinologists, in the past generation necessary conditions have practically disappeared from the armamentarium of discriminating historians of science.

10 W. Eberhard, "The Political Function of Astronomy and Astronomers in Han China," pp. 33–70, 345–352 in J. K. Fairbank (Ed.), *Chinese Thought and Institutions* (Chicago: The University of Chicago Press, 1957), p. 66.

39 The mirror image of this fallacy may be seen in an influential estimate of the *Chou i* 《周易》, the Book of Changes, as a deterrent to science. Here is the way Joseph Needham put it in 1956:

> while the five-element (*wu-hxing* 五行) and two-force (*yin-yang* 陰陽) theories were favourable rather than inimical to the development of scientific thought in China, the elaborated symbolic system of the Book of Changes was almost from the start a mischievous handicap. It tempted those who were interested in Nature to rest in explanations that were no explanations at all. The Book of Changes was a system for *pigeon-holing* novelty and then doing nothing more about it.

40 Nearly two decades later Ho Peng Yoke assured us that if Chinese scientists "were fully satisfied with an explanation they could find from the system of the *Book of Changes* they would go no further to look for mathematical formulations and experimental verifications of their scientific studies. Looking at the system of the *Book of Changes* in this light, one may regard it as one of the inhibiting factors in the development of scientific ideas in China."[11]

41 In these instances one is tempted to counter the arguments with matters of fact. Although Needham's extended discussion treats the Book of Changes predominantly as a static classificatory system of concepts, we find that natural philosophers most often used it to construct dynamic explanations of change. One also looks in vain for a habit among early Chinese scientists of constructing purely mathematical formulations and experimental verifications. If one cannot prove that this tendency was evolving steadily to a certain point, if there is no tangible evidence that without the Book of Changes they would have "gone further," there is no warrant for introducing from modern biology the metaphor of inhibition.[12]

---

11 J. Needham, *Science and Civilisation in China*, Vol. 2, pp. 336 and 340; P. Y. Ho, "The System of the *Book of Changes* and Chinese Science," *Japanese Studies in the History of Science*, 1972, 11: 23–39. Attempts to explain scientific revolutions by lists of positive and negative factors abstracted from context have been criticized by R. K. Merton in *Science, Technology and Society in Seventeenth-Century England* (New York: Howard Fertig, 1970), p. x.

12 Although Needham has given considerable weight to the notion of inhibition, as one would expect of a first-rate biologist he is cautious about using it in relation to processes that he

Exactly what does "inhibiting factor" mean in such contexts? Consider one 42
of these often used to explain why China failed to beat Europe to the Scientific Revolution despite a putative early head start, namely the predominance of a scholar-bureaucrat class immersed in books, faced toward the past, and oriented toward human institutions rather than toward Nature as the matrix of the well-lived life. But in Europe at the onset of the Scientific Revolution we are faced with the predominance in the universities of Schoolmen and dons, immersed in books, faced toward the past, and oriented toward human institutions rather than toward Nature. They did not prevent the great changes that swept over Europe. It would take a more imaginative historian than myself to swear that those changes would have taken place sooner had Scholasticism never existed.

The confusion about "inhibiting factors" is no less a confusion when it has 43
to do with ideas or techniques. One might just as well call Euclidean geometry an inhibiting factor for the development of non-Euclidean geometry, since so long as people were satisfied with it they didn't move on to a new step. But can one argue that non-Euclidean geometry would have developed sooner without it? It is unfortunate to see the remarkably interesting technical language of the Book of Changes, so powerful in systematically relating broader ranges of human experience than modern science attempts to encompass, written off as an obstacle before anyone has taken the trouble to comprehend all of its dimensions.

The first fallacy, then, confuses for a cause or necessary condition what 44
merely describes an earlier state of a culture, or a culture's way of doing something. In its complement, as can be seen by the examples just given, the absence of the

---

cannot prove were under way. Writers who draw on his work are not so discriminating. This point is easily demonstrated by examining the perceptive list of 29 "factors inhibiting the emergence of modern science in China and Western Europe" compiled from Needham's writing in Restivo (see note 8), pp. 46–47. In only four of these 29 does Needham actually invoke the concept of inhibition, and all are tautologous or too vague to challenge (e.g., "it is a matter for reflection how far Chinese algebra was inhibited from developments of post-Renaissance type by its failure to produce a sign which would permit the setting up of equations in modern form," *Science and Civilisation in China*, Vol. 3, p. 115). In a half-dozen other places Needham uses wording which suggests inhibition, generally in a similarly vague way (e.g., Confucian rationalism and humanism as "fundamental tendencies which paradoxically helped the germs of science on the one hand and injured them on the other," Vol. 2, p. 12). In the remaining score of instances Restivo has read inhibition into statements about failure, lack, and inadequacy.

subsequent state is confused with an inhibitor. One who commits this second fallacy is stopping growth that never took place. Both of the confusions I have described—blaming the earlier state for delaying the later state, and using the early absence of something modern to prove that modernity was unattainable later—confound continuity with stasis. They are bad history because they are bad reasoning.

45 I recur to the assumptions about ourselves that I have discussed earlier, for they are at the root of both these fallacies. They turn the history of world science into a saga of Europe's success and everyone else's failure, or at best inherently flawed and transitory success, until the advent of redemption through modernization.

46 Joint use of the pair of fallacies makes it easy to prove that the European breakthrough is not simply a fact of history, but was inevitable since history began. Was the horse and buggy a necessary preliminary to the invention of the automobile, or did it delay that invention? Would the automobile have emerged sooner if the buggy had never been invented, so that people would have been dissatisfied with less adequate vehicles? If we find the horse and buggy in Europe, by fallacy 1 its absence in China made the invention of some analogue of the automobile impossible. If we find some analogue of the horse and buggy in China, we apply fallacy 2 and make it an inhibiting factor. Thus medieval European impetus theory, abstract and unconcerned with application, was a stage in the evolution of inertial guidance; if Chinese who thought about physical questions were equally uninterested in their application, inertial guidance could never have originated in East Asia.

47 This is an infallible formula for reading the strength and power of modern science into the historic past—but only the past of Europe. For the past of other civilizations the test is always anticipation of or approximation to some aspect of early European science, or modern science. Why does the science of early Europe not need to be tested? Because of the *assumption* that it and only it gave birth to the Scientific Revolution. Other civilizations shine only as they reflect the light of the European tradition. Or so the prophets of modernization suppose.

48 I claim, therefore, that the fallacies that so often accompany discussions of the Scientific Revolution problem reflect a set of disastrous assumptions that lie beneath the obvious "heuristic" interest and charm of such discourse. They are disastrous because they assure us there is no point in comprehending on their

own terms the technical inquiries of non-Western cultures.[13] We now find these assumptions accepted not only in Europe but to some extent in every country in which the history of China is studied.

Why should intellectuals in a non-European country, which owes little of its 49
culture before modern times to Western influence, accept this bias? That is perhaps inevitable, considering that modern education establishes itself (as it did originally in Europe) by teaching the rejection of the traditional past, or its demotion to a cultural exhibit that may be of use for nurturing nationalism (and, in the era of mass-market foreign vacations, for enticing tourists). Since Japan has had over a century's experience with a modern educational system and the self-consciousness it produces, Nishijima Sadao's acute analysis does not come as a surprise:

> The 'static character' hypothesis holds that Chinese society lacked the capacity progressively to form a new era through its own efforts. This hypothesis was afforded particular emphasis by the viewpoint that the modernization of Chinese society was retarded. . . . Originally the 'static character' hypothesis, in company with that of 'Oriental despotism,' was advocated, in contrasts with Western European society, as a notion in polar opposition, for the sake of validating the selfconsciousness that came into being with the formation of modern Western European society. That is, it was a postulate to serve as an element in the recognition of the value of modern Western European society. . . . In our country, when we deal with Chinese society from the point of view that makes the formation of the modern ego identical with the equal valuation of individuals in Western European civilization, we are led uncritically to use the 'static character'

13 Although Needham consciously assumes "that there is only one unitary science of Nature, approached more or less closely, built up more or less successfully and continuously, by various groups of mankind from time to time," he sees this as a reason to study, rather than to ignore, non-European traditions. See his discussion cited in note 1 above.

The sorts of scholars who affirm, without troubling themselves to peruse the Chinese scientific literature, that it could not possibly have any value (see notes 7 and 8) have recently provoked a reaction, equally uninformed, that claims European science was markedly inferior to that of China as recently as three hundred years ago. See J. Gribbin, "Did Chinese Cosmology Anticipate Relativity?" *Nature*, 1975, 256 (5519) : 619–620, and for a critical discussion, N. Sivin, "Chinese Cosmology," ibid., 1976, 259 (5540): 249.

hypothesis. This has brought about our tendency to be controlled by the inverted logic that makes the goal of understanding Chinese society equivalent to grasping the origins or even the mechanism responsible for the persistence of its 'static character.'[14]

50 In a few words, anyone who begins by assuming that the paramount issue in the study of China is accounting for the inevitability of its backwardness is unlikely to question whether backwardness was inevitable, to ask whether there were not in her history prominent patterns of success from which we might learn, or to reexamine the assumptions about the modernized West that organize European history as a crescendo of success (with setbacks, to be sure, adding to the complexity and thus the charm of the crescendo), and that of other civilizations as a static picture of failure. Thus Nishijima states the convictions that justified and supported not only Japan's esteem for European civilization but also its political aspirations in East Asia before and during the Sino-Japanese War.

51 One more fallacy often appears in connection with the Scientific Revolution problem, when historians select the aspects of the European experience that are appropriate for comparison with other civilizations. I mean the fallacious assumption that one can make sense of the evolution of science by looking at intellectual factors alone, or socioeconomic factors alone, according to preference. Some people think of science predominantly as an intellectual quest after truths hidden in nature. They tend to think of China's failure to beat England to modern science as an failure of thought. Other people,who think of science as primarily a social or economic phenomenon, tend to see the defeat as a matter of Chinese social or economic backwardness. But neither of these exclusive approaches to explanation is adequate. The distinction between intellectual and social factors or between internal and external factors is not out there in the events we study, but in the mental habits and professional associations, in the division of labor, of historians.[15]

---

14 Nishijima Sadao 西島定生, *Chūgoku keizaishi kenkyū*《中国経済史研究》(Studies in Chinese Economic History; Tokyo: Tokyo University Press, 1966), pp. 3–4. Nishijima's remarks are part of an effort to explain the slow development of studies in Chinese agriculture.

15 G. E. R. Lloyd and N. Sivin, 2002 (see note 5), pp. 3, 234–238, uses this methodology. See also N. Sivin, "A Multidimensional Approach to Research on Ancient Science," forthcoming in *East Asian Science, Technology, and Medicine*, 2005, 23.

Dimensions of the Scientific Revolution

The Scientific Revolution and its consequences cut across the boundaries of historical 52
specializations. Let me make this clear by defining its important dimensions.

To begin at the intellectual end, the Scientific Revolution was a transformation 53
of our knowledge of the external world. It changed the questions we asked, the means we used to explore them, and the character of the answers. It established for the first time the dominion of number and measure over every physical phenomenon.

Ernst Gellner has pointed out a particular way in which the European Scientific 54
Revolution is more than a leap to a new form of knowing. It is natural to assume that in science the crucial test has always been "is it true?" But earlier that was only one of several equally important questions: Is it beautiful? Is it conventional? Is it morally improving? Does it lead to perception of the Good? Does it conform to certain esthetic patterns that all truth must, as astronomers up to Kepler believed that celestial orbits must be compounded of perfect circular motions? In science the test of truth has displaced most of these and redefined the others. This demand for truth above all was an appeal to fact—fact that was in principle public, verifiable, morally neutral, that did not change with the social circumstances of the observer, that was immune from interference by magician, or god, or human need. But the new science did more than appeal to facts. It created facts of that kind for the first time, knowledge that had no value except truth value. That is an awesomely original creation. It took place in Europe between the time of Copernicus and Laplace and has spread across the world since.[16]

The same leap was not taken in seventeenth-century China. People there 55
considered the idea of objective knowledge without wisdom, without moral or esthetic significance, grotesque.

The Scientific Revolution in Europe also meant redefining the connections of 56
natural philosophy (i.e., scientific thought) to other kinds of knowledge. It meant redefining man's orientation toward the past and the future. It meant redefining what authority should determine what uses may be made of knowledge. It meant redefining what knowledge of nature is socially desirable, and what socially

16 N. Gellner, *Legitimation of Belief* (Cambridge: Cambridge University Press, 1974), *passim*. The physicist-philosopher Robert Cohen makes the same point incidentally but a little more broadly when he speaks of "the Galilean turn" as a "rush toward dynamic, functionalized mathematics and abstract quality-stripped epistemology" (see note 7), p. 114.

undesirable. It meant redefining how knowledge ought to be related to human individuality and to the active relations of man and nature.

57 Galileo and his friends and successors could not have got round the authority of the Church on the strength of ideas alone. That message was conveyed to Galileo by the Congregation of the Index in 1616, and then with drastic finality when he was condemned in 1633. But he and his fellow spirits had begun constructing a new intellectual community outside the old establishment. A hundred years earlier there had been no organized alternative to the Church and its scholastic educational system; then even Galileo himself might have died an archbishop.[17] But in the Counter-Reformation the Church, threatened by Protestantism, became defensive and obsessed with thought control. It naturally became less attractive to the most talented and ambitious (and of course there was less room for those who were attracted). A variety of new careers was emerging. Among them the profession of scientist was being invented. This profession could not provide structures that paid salaries for specialist careers until about 1800.[18] Nevertheless, from the start it assumed for its amateurs, devotees, and enthusiasts, independent authority to formulate the laws of nature. Scientists took that authority away, in fact, from the Scholastics, for whom science could never be more than a collaborator of faith. Secular learning remade the universities and displaced other ancient institutions while over several centuries of evolution and revolution it formed a technical establishment.

58 This outline of the Scientific Revolution's many dimensions is meant to suggest how much we are likely to miss if we care only about social factors, or only about intellectual factors, as we survey the situation in China. Until recently, for instance, people concerned with that topic, including myself, have overlooked a significant piece of the Chinese picture, which I will now consider.

Scientific Revolution in China

59 By conventional intellectual criteria, China had its own scientific revolution in the

17 On the earlier position of the Church as a locus of careers open to talent, see A. Murray, *Reason and Society in the Middle Ages* (New York: Oxford University Press, 1978), pp. 282–314.

18 See, for instance, A. Thackray, "Natural Knowledge in Cultural Context: The Manchester Model," *The American Historical Review*, 1974, 79 (3): 672–709, esp. p. 692.

seventeenth century. This is a point of no small interest if we are meditating about why China couldn't have had one.

Western mathematics and mathematical astronomy were introduced to China 60
beginning a little after 1600, in a form that before long would be obsolete in those parts of Europe where readers were permitted access to current knowledge. Several Chinese scholars quickly responded and began reshaping the way astronomy was done in China. They radically and permanently reoriented the sense of how one goes about comprehending the celestial motions. They changed the sense of which concepts, tools, and methods are centrally important, so that geometry and trigonometry largely replaced traditional numerical or algebraic procedures. Such issues as the absolute sense of rotation of a planet and its relative distance from the earth became important for the first time. Chinese astronomers came to believe for the first time that mathematical models can explain the phenomena as well as predict them. These changes amount to a conceptual revolution in astronomy.

That revolution did not generate the same pitch of tension as the one going on 61
in Europe at the same time. It did not burst forth in as fundamental a reorientation of thought about Nature. It did not cast doubt on *all* the traditional ideas of what constitutes an astronomical problem. It did not narrow people's views of what meaning astronomical prediction can have for the ultimate understanding of Nature and of man's relation to it.

The most striking long-range outcome of the encounter with European science, 62
in fact, was a revival of traditional Chinese astronomy, a rediscovery of forgotten methods, that were studied once again in combination with the new ideas and that supported what might be called a new classicism. Rather than replacing traditional values, the new values implicit in the foreign astronomical writings were used to renovate traditional values.[19]

Why didn't this conceptual revolution have the social consequences that 63
historians of Western science have encouraged us to expect? The old and new astronomy were not in antagonistic competition, once the Chinese acknowledged that the European techniques yielded much more reliable predictions. By the mid seventeenth century European civilization had had no appreciable political or social impact, and astronomy was making its way on its own merits.

---

19 B. Elman, *From Philosophy to Philology: Intellecutal and Social Aspects of Change in Late Imperial China* (Cambridge: Council on East Asian Studies, Harvard University, 1984).

64 One is tempted to see the later process by which Western astronomy became rooted in China as the last major face-to-face encounter of non-Western and European science in world history. By the eighteenth century modern science was crossing national boundaries on the coattails of Empire, and competition between sciences, literatures, religions, etc., on the basis of their particular merits had become a thing of the past. Even in the mid seventeenth century, despite the high drama of eclipse prediction contests in the late Ming court, the fact remains that the triumph of European computational techniques came about not through a consensus of great minds but by an imperial decision to hand over operational control of the Astronomical Bureau to Jesuit missionaries.

65 The foreign techniques, powerful though they were, offered Chinese students no alternative route to security and fame, and the civil service examination system left no room for one. The only astronomers who could respond to the Jesuits' writings were members of the old intellectual elite. They were bound to evaluate innovations in the light of established ideals that they felt an individual responsibility to strengthen and pass on to the next generation.

66 Revolutions in science as well as in politics take place at the margins of society, but the people who made the one in seventeenth-century China were firmly attached to the dominant values of their culture.[20] At the time there were no students of astronomy motivated to cast off traditional values. There were no groups of intellectuals alienated enough to follow ideas where they led even if the society around them fell apart.

67 The most influential first-generation champions of Western astronomy were men of the lower Yangtze region who lived through the Manchu invasion of the 1640's. They adopted the traditional role of the loyalist who would not serve a new dynasty, particularly what they saw as a non-Chinese dynasty. Having refused to strive for conventional careers in a society that in their view *had* fallen apart, they were motivated to spend their lives studying and teaching the new mathematics and astronomy while they used them to master the neglected techniques of their

---

20 See "Wang Hsi-shan," pp. 159 and 164, and for further details, N. Sivin, "Wang Hsi-shan," *Dictionary of Ming Biography, s.v.*

own tradition.[21] They rejected the Ch'ing present not for a modernist future but to keep alive the lost cause of the Ming for one more generation. Wang Hsi-shan even avoided using the Ch'ing dating system. Despite his superb critical acumen he was the opposite of Descartes, for whom every ancient institution had to justify itself by the new criterion of clear and distinct ideas or be considered a dead relic.

If then we seek in China those for whom science was not a means to conservative ends, those for whom a proven fact outweighed values that had evolved for thousands of years, we do not find them until the late nineteenth century. Then it was people with little or no stake in the old society who became the first modern scientists. By that time foreigners exempt from Chinese law and backed by gunboats could do what they wanted in China. They constructed new institutions and new career lines that let them attract and educate talented young people who had no other prospects. We can no longer talk about the encounter of the old and new astronomy. Social and political change had left nothing for the old to do. It became rare as time passed for modern scientists to be aware that their country had had its own scientific tradition. Only in the last couple of generations has that awareness became general. 68

Conclusion

My frustrations in trying to make sense of science in China arise partly because of the many levels of human activity that have to be encompassed over such a great sweep of time and human experience. They arise partly because the European Scientific Revolution seems to call for an understanding in greater breadth and depth than its historians have insisted upon. Once we keep in mind the many dimensions of scientific change and their complex relations, it becomes less surprising that the Scientific Revolution took place only when and where it did. The process increasingly comes to resemble any historic evolution, which is always the sum of human decisions and acts, some arbitrary, many wrongheaded—in other 69

21 B. Elman, 1984 (see note 19). Another signal contribution on the influence of astronomy is J. B. Henderson, *The Development and Decline of Chinese Cosmology* (New York: Columbia University Press, 1984).

words, muddling through. We do not need to appeal to fate, determinism, teleology, cultural superiority, an inexorably unfolding inner logic, or the hidden operation of some World Spirit.

70 Looking at these two scientific revolutions—the one we think we know so well in Europe, and the one that wasn't what we expect it to be in seventeenth-century China—suggests that we have a great deal to learn about the specific circumstances of each, seen in all its dimensions, before we are ready to tell the world why it couldn't have happened in other times or places.

71 I believe that the breakthroughs coming up in the study of Chinese science will be of another kind altogether. They will have to do with understanding in depth and in an integral way the circumstances of people who did science and technology: how their technical ideas related to the rest of their thought; what the scientific communities were—that is, who formed a consensus that certain phenomena were problematic, and that certain kinds of answers were legitimate; how those communities were related to the rest of society; how they were supported; how the responsibility of men of knowledge to their colleagues in science was reconciled with their responsibility to society; and what larger ends the sciences served, that kept their laws conformable to the laws of Chinese painting and to the basic principles of moral conduct.[22]

72 These are issues about which we understand very little with respect to China or to Europe. It will take much further study and reflection on both sides before the comparative history of science is ready to take off. My prognostication is that by that time we will no longer be asking why the transition to modern science did not first take place in China.

22 The language in which I pose these questions is more or less that of the sociology of knowledge. It is interesting that A. C. Crombie has phrased a very similar set of topics, also intended to provide an integrated view of the Scientific Revolution problem in terms familiar to intellectual historians and thus to the majority of historians of science: "conceptions of nature and of science, of scientific inquiry and scientific explanation, of the identity of natural science within an intellectual culture, and the intellectual commitments and expectations that affect attitudes to innovation and change" ("Science and the Arts in the Renaissance: The Search for Truth and Certainty, Old and New" *History of Science*, 1980, 18: 233–246, especially pp. 234–235).

APPENDIX

CHAPTER HEADINGS IN SHEN KUA, BRUSH TALKS FROM DREAM BROOK

No.	Heading	Item
1	Ancient Usages	1
3	Philological Criticism	42
5	Music and Mathematical Harmonics	82
7	Regularities Underlying the Phenomena	116
9	Human Affairs	151
11	Civil Service	189
13	Wisdom in Emergencies	224
14	Literature	245
17	Calligraphy and Painting	277
18	Technical Skills	298
20	The Supernormal	338
21	Strange Occurrences	357
22	Errors	388
23	Wit and Satire	401
24	Miscellaneous	420
26	Materia Medica	480

There are a total of 507 jottings in *Brush Talks from Dream Brook*, or 609 including its two sequels. For a classification of the book's contents according to field of knowledge treated, see J. Needham, *Science and Civilisation in China*, Vol. 1, 136.

Text 10b

《夢溪筆談》

節錄

沈括 著

卷十八 技藝

304. 棋局[1]都數[2] *

小說，唐僧一行[3]曾算棋局都數，凡若干局盡之。予嘗思之，此固易耳。但數多非世間名數[4]可能言之。今略舉大數，凡方[5]二路[6]，用四子，可變八千

1　棋局，這裏指在圍棋棋盤上布子的形勢。
2　都數，即總數。都，可指總滙、總括或總計，在唐代，張九齡〈敕安西節度王斛斯書〉之三已云：「卿狀但言都數，其中不列姓名，已令勘責，可速以實報。」
*　標題的數字乃現代學者加上。沈括所注的《夢溪筆談》、《補筆談》和《續筆談》經學者校正後，共得六百零九條，每條註一號碼，方便引用及尋檢。
3　一行法師，本姓張，名遂。生於唐高宗弘道元年（公元683年），幼年即以聰慧及博通經史聞名，後為避武三思的招攬，於嵩山嵩陽寺剃度出家，改名敬賢，法號一行。一行大師精天文曆法，創製大衍曆，為中國著名的科學家。
4　名數，帶計量單位，如十、百、千、萬等的數字。世間名數，指常用的帶計量單位數字，如一萬五千等。
5　方，此處指縱向和橫向地計算，《論語．先進》云：「方六七十，如五六十」，即「縱橫各六七十里，或縱橫各五六十里的地方」。
6　路，指下棋的位置。二路是兩個下棋的位置，但因為是「方二路」，也即縱向和橫向都有兩個位置，所以構成四個可以放棋子的位置。

文言文節錄自沈括，《新校正夢溪筆談》（1975），由中華書局（香港）有限公司授權翻印。語譯及注釋由香港中文大學巢立仁撰寫。

十一局[7]；方三路，用九子，可變一萬九千六百八十三局[8]；方四路，用十六子，可變四千三百四萬六千七百二十一局[9]；方五路，用二十五子，可變八千四百七十二億八千八百六十萬九千四百四十三局[10]；古法：十萬為億，十億為兆，萬兆為秭。算家以萬萬為億，萬萬億為兆，萬萬兆為垓。今但以算家數計之。方六路，用三十六子，可變十五兆九十四萬六千三百五十二億八千二百三萬一千九百二十六局[11]；方七路以上，數多無名可記。盡三百六十一路，大約連書萬字四十三即是局之大數[12]。萬字四十三，最下萬字即萬局，第二是萬萬局，第三是萬億局，第四是一兆局，第五是萬兆局，第六是萬萬兆，謂之一垓，第七是垓局[13]，第八是萬萬垓，第九是萬億萬萬垓[14]。此外無名可紀，但四十三次萬倍乘之，即是都大數，零中數不與。[. . .]

【譯文】

304. 棋局總數

（某）筆記小說（記載）：唐代僧人一行曾計算（圍棋）棋局的總數，（發現）大概某一定數（就能）算盡圍棋的棋局。我曾思考這個數目，這個算法其實（是）容易的。然而數目大（得）並非世間的帶單位數字能夠表述它。

現在簡略說明幾個大數，（在棋盤上的）縱橫各二路，使用四枚棋子，可以變出八十一種棋局；縱橫各三路，使用九枚棋子，可以變出一萬九千六百八十三種棋局；縱橫各四路，使用十六枚棋子，可以變出四千三百四萬六千七百二十一種棋局；縱橫各五路，使用二十五枚棋子，可以變出八千四百七十二億八千八百六十萬九千四百四十三種棋局；**古代的命數法：十萬為億，十億為兆，萬兆為秭。算學家（則）以萬萬為億，萬萬億為兆，萬萬兆**

7 局指棋局，但二路四個位置，沒可能變化出逾八千種棋局，千字誤，此處應為八十一局，即八十一種棋局。因每一個位置都有三種可能，即有一枚黑子，一枚白子或留空，所以四個位置總合的布子可能，就變成$3×3×3×3$，等於八十一種棋局。

8 即$3^9 = 19683$。

9 即$3^{16} = 43046721$。

10 即$3^{25} = 847288609443$。

11 棋局的正確數字應為$3^{36} = 150094635296999121$，沈括在千萬位開始算錯。

12 棋局總數是$3^{361} \approx 1.74×10^{172} = 1.74×10^{4×43} = 1.74 × 10000^{43}$，因此如沈括所說要連寫萬字四十三次。

13 第七是垓局，應為第七是萬垓局。審核文句（前句方見總結之前「謂之一垓」，承接句不應即指更後的「第七」又「是垓局」），萬字或因作者採用古文隔句省略寫法而省法，或因文獻本身訛脫而消失。

14 此處應為萬萬萬垓局，或萬億垓局，原文應誤，不知乃沈括之誤，抑文獻傳刻之誤。

為垓。現在只用算學家的命數法來計算棋局數字。縱橫各六路，使用三十六枚棋子，可以變出十五兆九十四萬六千三百五十二億八千二百三萬一千九百二十六種棋局；縱橫各七路以上，（棋局的）數目太大沒有單位可以表述，（如果）盡（用）三百六十一個位置，大約連寫萬字四十三次就是棋局的大概總數。**四十三個萬字，最後的萬字代表萬局，第二（個萬字）代表萬萬局，第三（個萬字）代表萬億局，第四（個萬字）代表一兆局，第五（個萬字）代表萬兆局，第六（個萬字）代表萬萬兆（局），稱為一垓，第七（個萬字）代表（萬）垓局，第八（個萬字）代表萬萬垓（局），第九（個萬字）代表萬億萬萬垓（局）。這個數目以外，就沒有單位可以紀錄，只是將萬字乘四十三次，（所得的積）就是（全部棋局）總合的大概數，零頭數不包括。**[. . .]

307. 活板印刷

板印書籍，唐人尚未盛為之[1]。自馮瀛王[2]始印五經[3]，已後典籍，皆為板本。慶曆[4]中，有布衣畢昇，又為活板。其法用膠泥刻字，薄如錢脣[5]，每字為一印，火燒令堅。先設一鐵板，其上以松脂臘[6]和紙灰之類冒[7]之，欲印則以一鐵範[8]置鐵板上，乃密布字印，滿鐵範為一板，持就火煬[9]之，藥稍鎔，則以一平板按其面，則字平如砥[10]。若止印三、二本，未為簡易；若印數十百千本，則極為神速。常作二鐵板，一板印刷，一板已自布字，此印者纔畢，

1　錢存訓《李約瑟中國科學技術史・第五卷化學及相關技術・第一分冊紙和印刷術》：「對中國雕版印刷的問世，歷來有各種不同的說法，其年代從六世紀中葉到九世紀末不等。現存文物中，沒有早於八世紀的。但是有文獻紀錄說明，印刷的開始年代，可能早於八世紀。」（頁131。上海：科學出版社/上海古籍出版社，1990年一版。）宋代的筆記有唐人作雕板印刷的記載，例如葉夢得在《石林燕語・卷二》提到唐人「雕版印紙」，而朱翌在《猗覺寮雜記・卷六》則提到「唐末益州始有墨版」。

2　馮瀛王，即馮道。馮道，字可道，自號長樂老，死後，謚號文懿，追封瀛王，瀛州景城（今河北泊頭市交河鎮）人。於五代時，歷仕後唐、後晉、後漢、後周等四朝十一君，任宰相、太師及中書令等要位。後唐時，與李愚同奏請由國子監刻印《九經》，那是五代有大規模雕板印書的明證，同時也是官府刻書的最早確記。結果，其事歷經二十餘年，至後周時，方見峻工。

3　五經，即《易經》、《尚書》、《詩經》、《禮記》及《春秋》。

4　慶曆，宋仁宗趙楨的年號，公元1041年11月至公元1048年。

5　錢脣，即銅錢的邊緣。

6　臘，通蠟。

7　冒，即覆蓋，蓋滿。

8　鐵範，即鐵製的框架、模子。

9　煬，即烤。

10　砥，即磨刀石。

則第二板已具。更互用之，瞬息可就。每一字皆有數印，如之、也等字，每字有二十餘印，以備一板內有重複者。不用則以紙貼之，每韻[11]為一貼，木格貯之。有奇字素無備者，旋刻之，以草火燒，瞬息可成。不以木為之者，木理有疏密，沾水則高下不平，兼與藥相粘，不可取，不若燔土，用訖再火令藥鎔，以手拂之，其印自落，殊不沾污[12]。昇死，其印為予羣從[13]所得，至今寶藏。

【譯文】

307. 活板印刷

用雕板印刷書籍，唐代人還沒有廣泛地實行這種方法。自從馮道印刷五經開始，以後的經書要籍，（才）全部都是以雕板印刷的本子。慶曆年間，有平民（名）畢昇（者），又創造了活字板（印刷術）。他的方法（是）使用膠泥刻製字形，（字形的突出部分）薄得像銅錢的邊緣，每字都做成一個方印，用火燒（印）使（印）堅硬。先設置一塊鐵板，在板上面以松脂蠟摻和紙灰的藥料塗滿它，要印刷就將一個鐵框放置在鐵板上，（在其中）緊密地鋪排字形方印，（排）滿了鐵框即為一板，拿着它到火上烤；（覆蓋板上的）藥料稍為鎔化（時），則以（另）一塊平板按壓字印的上面，那麼（整版）字（就會）平滑如磨刀石了。如果只是印刷三、兩本，（這種方法）不算簡單方便；（但）如果印刷數十、（數）百（或數）千本，（那）就極為神速（了）。通常準備兩塊鐵板，一板（在）印刷（時），（另）一板已經在排字，這一板字方才（印）好，而第二板已經備好。更替輪互地運用兩板，（工作）很快就可以完成。每一個字形都有數個印，好像之、也等字，每個字（更）有二十多個印，以備一板之內（的文字）有重複的。不使用（時）就以紙條標簽這些字印，每一個韻部是一個標籤，（並以）木格子貯存它們。遇有素來沒有準備的冷僻字，立刻（就）製作字印，以草（生）火燒焙，轉眼可以製成。（之所以）不用木材製作字印，（是因為）木材的紋理有疏（有）密，如果沾了水（字板）就會高低不平，而且還（會）與藥料黏在一起，不可以拿下來。不像燒泥（做）的字印，用完再用火烤（它）使藥料鎔化，（然後）用手一拂，板上的字印就會自行掉落，完全不會沾上污垢。畢昇死（後），他的活字印被我的堂兄弟 / 子侄得到，到現在還珍重地收藏（着）。

11 韻，指韻書中的韻部，如《廣韻》的東、同、中、蟲等。

12 沾污，指沾上屬於藥料的殘餘物。

13 群從，可指堂兄弟或子侄。

卷二十一 異事異疾附

357. 虹

世傳虹能入溪澗飲水，信然。

熙寧[1]中，予使契丹[2]，至其極北黑水境永安山[3]下卓帳[4]。是時新雨霽[5]，見虹下帳前澗中。予與同職扣[6]澗觀之，虹兩頭皆垂澗中。使人過澗，隔虹對立，相去數丈，中間如隔綃縠[7]。自西望東則見；蓋夕虹也。立澗之東西望，則為日所鑠[8]，都無所覩。久之稍稍正東，踰山而去。次日行一程，又復見之。孫彥先[9]云：「虹乃雨中日影也，日照雨則有之。」

【譯文】

357. 虹

世人傳說彩虹能夠到溪澗中飲水，這事（是）真確的。

熙寧年間，我出使契丹，到了它最北邊的黑水境內永安山下支起了帳篷。當時剛好雨後初晴，看見彩虹下探到帳篷前的溪澗裏。我與同事（一起）貼近溪澗觀察它，（發現）彩虹的兩首都垂入溪澗之中。着人渡過溪澗，（大家）隔着彩虹相對而立，相距數丈，中間好像隔了（一層）輕紗。（這道彩虹）從西向東觀望就（可以）看到。**因為（是）傍晚的彩虹啊**。站在溪澗的東面向西觀望，就會被太陽晃眼，（甚麼）都沒有看見。過了

1　熙寧，宋神宗趙頊的年號，公元1068至1077年。

2　唐末及宋代時，契丹曾以國號遼，在中國北方，建立王朝。沈括出使契丹，應為熙寧八年，公元1075年。（《續資治通鑑長編・卷二百五十五》）

3　黑水境，可能指沈括出使路經的黑河州。黑河州由遼穆宗所設置，沈括出使路經時應已廢置。永安山則應指馬盂山，處於遼國中京大定府西，即位處今內蒙古寧城縣西。

4　卓帳，卓指高，本為形容詞，活用作動詞，作立，直立解。卓帳，即使帳篷豎起來。

5　雨霽，即雨下剛止，天方轉晴。

6　扣，在這裏指貼近，靠近。扣本指牽着的意思，例如扣馬，即指牽着那一匹馬，此處與澗搭配，即指人貼近溪澗。

7　綃縠，即生絲織成的薄紗、薄絹。

8　鑠，字亦作爍，可解作銷鎔，此處指發光貌，即耀目。

9　孫彥先，字思恭，登州（今山東蓬來縣）人。曾任宛丘令，與轉運使為護民爭執，棄官而去，吳奎以其學行，推薦擔任中央官職，曾任王府侍講、天章閣侍制等。孫彥先，精擅象數之學，特別擅長關氏易（後魏關朗撰），對其中以術數談論天地化生的篇章 —— 大衍，尤見精熟。他也曾經修整天文苑的渾天儀，著有「堯年至熙寧長歷」。《宋史・列傳》云：「近世歷數之學，未有能及之者。」

很久，（彩虹）慢慢（向）正東，踰越山嶺而消失。第二天走了一段路，又再看見彩虹。**孫彥先說：「彩虹是（下雨時）雨水中太陽的影子，太陽照着雨點就會有彩虹（出現）。」**

卷二十四 雜誌一

430. 海陸變遷

予奉使河北[1]，遵太行而北，山崖之間，往往銜螺蚌殼及石子如鳥卵者，橫亙[2]石壁如帶。此乃昔之海濱，今東距海已近千里。所謂大陸[3]者，皆濁泥[4]所湮[5]耳。堯殛鯀于羽山[6]，舊說在東海中，今乃在平陸。凡大河[7]、漳水[8]、滹沱[9]、涿水[10]、桑乾[11]之類[12]，悉是濁流。今關、陝以西，水行地中，不減百餘尺[13]，其泥歲東流，皆為大陸之土，此理必然。

【譯文】

430. 海陸變遷

我奉命任察訪使（出訪）河北，沿着太行山北（行），（發現）山崖

1 奉使河北。熙寧七年，即公元1074年，朝廷命沈括為河北西路察訪使，到河北視察地理及軍事形勢。(《續資治通鑑長編・卷二百五十五》)
2 亙，即連貫，首至尾一直相接。
3 大陸，應指華北平原。
4 濁泥，指由河水沖積來的濕潤沙土。
5 湮，即沉沒，埋沒，壅塞，這裏主要指沉沒。
6 帝堯時洪水泛濫，堯接受四岳的推薦，命鯀治水，鯀以湮塞之法治水九年，不能成功，堯殛鯀於羽山。一說堯命祝融殛鯀，又一說為堯流放鯀於羽山，鯀為舜所殛。(《尚書・洪範》、《尚書・堯典》、《山海經・海內經》、《淮南子・原道訓》)
7 大河，即黃河。
8 漳水，即漳河，屬衛河支流，發源於山西，而流經河北省。
9 滹沱，即滹沱河，在河北省西部流經山西省繁峙縣泰戲山，東流入河北平原，在獻縣與滏陽河滙合成子牙河。
10 涿水，相信應指源出河北省涿鹿山，今稱拒馬河的河流。
11 桑乾，即桑乾河，屬永定河的上游，位於河北省西北部，山西省朔州朔城區河灣一帶。
12 這些都是流經河北省西部的河流。
13 水行地中，不減百餘尺，即指水流沖擊，切割地表，形成深谷，所以出現河道處於地表（水平線）下逾百尺處的情況。

之間，往往混銜螺蚌殼和像鳥卵一樣的石頭，橫貫在石壁中仿如一條帶子。這裏屬以前的海濱，現在離開東面的大海已經接近千里。所謂的大陸，都（由）河水沖帶下來的沙土所沉積（而成）。堯殺鯀於羽山，過去傳說在東海之中，如今已在陸地上。黃河、漳河、滹沲河、拒馬河、桑乾河等等，都是（滿含）泥沙的河流。現在潼關、陝西以西，水都行走於地表下，不少於一百尺之處。水中的沙泥年年向東流，全都成為大陸的泥土[14]，這是理所必然。

437. 指南針

方家[1]以磁石磨針鋒，則能指南，然常微偏東，不全南也[2]。水浮多蕩搖。指爪[3]及盌脣[4]上皆可為之，運轉尤速，但堅滑易墜，不若縷[5]懸為最善。其法取新纊[6]中獨璽[7]縷，以芥子[8]許蠟，綴於針腰，無風處懸之，則針常指南。其中有磨而指北者。予家指南、北者皆有之。磁石之指南，猶柏之指西，莫可原其理。

【譯文】
437. 指南針

方術者以磁石摩擦針尖，就能（令針尖）指向南方，然而經常稍稍偏向東方，不全然指向正南方。（磁針）浮在水上常蕩漾搖晃。指甲及碗邊上都可以放置磁針，運轉尤其快速，但是（這些物品的表面）又硬又滑（磁針）容易墜落，不如以絲線懸掛最好。其方法（是）抽取新絲中（由）單一個

14 指沙泥沉積成為新的土地。

1 方家，古代稱各種技術，諸如工程、醫藥、堪輿等為方術，擅某術者，稱方家或術者。
2 地球是一個大磁場，指南針受地磁力作用影響，一端指向地磁南極，一端會指向地磁北極。然而地理的南極與北極並不與地磁的南極與北極完全重疊，所以就出現了指南針不能準確指出真正南方的情況。沈括所發現的角度差異，就是地磁南極線與真正南方的角度差異，今天稱為磁偏角現象（現在主要以磁北線與真北線的夾角來說明磁偏角現象）。沈括的發現，比西歐早了逾四百年，英國人諾曼（Robert Norman），在1581年，於*The New Attractive* 一書中，描述了這個現象，並稱之為磁偏角（magnetic dip）。
3 指爪，即指甲。
4 盌脣，即碗的邊緣。盌亦作椀、碗，音義相同；脣，根據說文解字是唇的本字。
5 縷，即絲線。
6 纊，即新的絲綿。
7 璽，同繭，指蠶繭。
8 芥子，即芥菜籽。

蠶繭（繅出來）的絲，以芥菜籽大小的臘，（把絲）黏貼於針腰，在沒有風的地方掛起它，則針尖經常指向南方。其中（也）有磨擦後（針尖）指向北方的。我家中指向南方、北方的針都有。磁石指向南方，好比松柏指向西方，不能夠推原其中的道理。

Text 10b

from

Brush Talks from Dream Brook

by Shen Kua

Total Number of Possible Situations in Chess*

The story-tellers say that I-Hsing once calculated the total number of possible situations (*chü*[1]) in chess,[2] and found he could exhaust them all. I thought about this a good deal, and came to the conclusion that it is quite easy. But the numbers involved cannot be expressed in the commonly used terms for numbers. I will only briefly mention the large numbers which have to be used. With two rows (*fang erh lu*[3]) and four pieces (*tzu*[4]) the number of probable situations will be of 81 different kinds. With three rows and nine pieces the number will be 19,683. Using four rows and sixteen pieces the number will be 43,046,721. For five rows and twenty-five

* Headings are added by this publication.

1 局

2 Presumably *wei chhi* 圍棊(棋). This 'chess' is of course a different game from that known in the West, and must go back at least to the +3rd century (*Chhou Fen Chuan*, ch. 1). See Culin (1), p. 868, where it can be seen that the modern board has 19 rows. Shen Kua says that the oldest board had 17 rows, making 289 positions on the grid, and that there were 150 white and 150 black pieces. The 'situations' referred to mean therefore either 'white-occupied', 'black-occupied' or 'empty'. The board is empty at the beginning of the game. See also Volpicelli (1); H. A. Giles (6).

3 方二路

4 子

pieces, the number will be 847,288,609,443. . . . Above seven rows we do not have any names for the large numbers involved. When the whole 361 places are used, the number will come to some figure (of the order of) $10{,}000^{52}$ (*lien shu wan tzu wu shih erh*[5]). . . .

Moveable Type Printing

During the reign of Chhing-li [+1041–48]** Pi Sheng, a man of unofficial position, made moveable type. His method was as follows: he took sticky clay and cut in it characters as thin as the edge of a coin. Each character formed, as it were, a single type. He baked them in the fire to make them hard. He had previously prepared an iron plate and he had covered his plate with a mixture of pine resin, wax, and paper ashes. When he wished to print, he took an iron frame and set it on the iron plate. In this he placed the types, set close together. When the frame was full, the whole made one solid block of type. He then placed it near the fire to warm it. When the paste [at the back] was slightly melted, he took a smooth board and pressed it over the surface, so that the block of type became as even as a whetstone.

If one were to print only two or three copies, this method would be neither simple nor easy. But for printing hundreds or thousands of copies, it was marvelously quick. As a rule he kept two formes going. While the impression was being made from the one forme, the type was being put in place on the other. When the printing of the one forme was finished, the other was then ready. In this way the two formes alternated and the printing was done with great rapidity.

For each character there were several types, and for certain common characters there were twenty or more types each, in order to be prepared for the repetition of characters on the same page. When the characters were not in use, he had them arranged with paper labels, one label for words of each rhyme-group, and kept them in wooden cases. If any rare character appeared that had not been prepared in advance, it was cut as needed and baked with a fire of straw. In a moment it was finished.

The reason why he did not use wood is because the tissue of wood is sometimes coarse and sometimes fine, and wood also absorbs moisture, so that the forme when

5 連書萬字五十二
** All brackets are original from the source texts.—*Ed.*

set up would be uneven. Also the wood would have stuck in the paste and could not readily have been pulled out. So it was better to use burnt earthenware. When the printing was finished, the forme was again brought near the fire to allow the paste to melt, and then cleansed with the hand, so that the types fell off of themselves and were not in the least soiled.

When Pi Sheng died, his font of type passed into the possession of my nephews, and up to this time it has been kept as a precious possession.[6]

The Rainbow[+7]

Traditionally it is said that the rainbow can enter a watercourse and drink the water. This is quite true.

In the Splendid Peace era, when I was ambassador to the Qidan state [1075], I went to its far north, the Black Water 黑水 area. We set up our camp below Mt. Yongan 永安山. At the time it was just clearing after a storm. I could see a

---

6 *Meng Chhi Pi Than* (*TSHCC*), ch.18, p.117; cf. tr. Carter (1), pp. 212–13. Part of the original text was cited in *Huang Chhao Shih Shih Lei Yüan* 皇朝事實類苑 (*SFSTK*), compiled by Chiang Shao-Yü in 1145.

\+ Translated by Nathan Sivin, unpublished draft.—*Ed.*

7 The Qidan were a people who set themselves up in the north as the Liao dynasty, a rival of the Song state, from 916 to 1125.

Sun Sigong was another polymath, approximately of Shen's time. His life and work have been little studied. He has a brief biography in *Sung Shi* 宋史, 322: 10446–47. It notes that Sun, who served as a provincial and central official, was skilled in the Book of Changes, mathematics, and chronology, and renovated an armillary sphere in the astronomical bureau; "in the study of astronomical computation in recent years, no one could measure up to him."

The last two sentences of this translation appear in Hu Daojing's edition as one of Shen's notes. In some earlier editions they are part of the text. Here I include it because of its inherent interest. It is not, however, equivalent to the modern physical explanation, in which what forms the rainbow is refraction of sunlight in raindrops, which Sun does not mention.

Because the distinction between singular and plural is not built into Chinese syntax, and authors do not always specify which they mean, a translator must sometimes be arbitrary. For instance, in this jotting it is not clear

1. how many were in Shen's group. Since he was an imperial envoy, one can surmise that he was traveling with a group of colleagues;
2. whether there was one tent or a camp;
3. how many of Sun's party crossed the stream; the text does not specify.

As for the author's note, Shen clearly does not recall the time of day, but the fact that the sun was to the west, and subsequently moved further west, indicates late afternoon.

rainbow suspended in front of the camp in a stream. My colleagues and I approached the stream and watched [this phenomenon]. The two ends of the rainbow hung in the stream. I sent someone across the stream to stand facing us on the other side of the rainbow, dozens of yards away. It looked as if we were separated by gauze. Looking eastward from the west, [the rainbow] was visible,[8] but if we stood to the east of the stream and looked westward, we were dazzled by the sun and could see nothing. After a long while [the rainbow] gradually moved due eastward, crossed the mountain and disappeared. The next day, after a day's travel, we saw it again.

Sun Sigong 孫彥先=思恭 (before 1080) has said that the rainbow [forms from] images of the sun in the rain. The sun shines on the rain, and there it is.

Transformation of the Land and the Sea

When I went to Hopei on official duties I saw that in the northern cliffs of the Thai-Hang Shan[9] mountain-range, there were belts (strata) containing whelk-like animals, oyster-shells, and stones like the shells of birds' eggs (fossil echinoids). So this place, though now a thousand *li* west of the sea, must once have been a shore. Thus what we call the 'continent' (*ta lu*[10]) must have been made of mud and sediment which was once below the water. The Yü Mountain,[11] where Yao[12] killed Kun, was, according to ancient tradition, by the side of the Eastern Sea, but now it is far inland.

Now the Great (i.e. the Yellow) River, the Chang Shui,[13] the Hu Tho,[14] the Cho Shui[15] and the Sang Chhien[16] are all muddy silt-bearing rivers. In the west of Shensi and Shansi the waters run through gorges as deep as a hundred feet. Naturally mud and silt will be carried eastwards by these streams year after year,

---

8 It seems that it was a rainbow near sunset.
9 太行山
10 大陸
11 羽山
12 Legendary emperor; cf. vol. 2, p. 117, for Kun, the father of Yü the Great.
13 漳水
14 滹沱
15 涿水
16 桑乾

and in this way the substance of the whole continent must have been laid down. These principles must certainly be true.[17]

The Compass

Magicians‡ rub the point of a needle with the lodestone; then it is able to point to the south. But it always inclines slightly to the east, and does not point directly at the south (*jan chhang wei phien tung, pu chhüan nan yeh*[18]).[19] (It may be made to) float on the surface of water, but it is then rather unsteady. It may be balanced on the finger-nail, or on the rim of a cup, where it can be made to turn more easily, but these supports being hard and smooth, it is liable to fall off. It is best to suspend it by a single cocoon fibre of new silk attached to the centre of the needle by a piece of wax the size of a mustard-seed—then, hanging in a windless place, it will always point to the south.

Among such needles there are some which, after being rubbed, point to the north. I have needles of both kinds by me. The south-pointing property of the lodestone is like the habit of cypress-trees of always pointing to the west. No one can explain the principles of these things.[20]

17 *MCPT*, ch. 24, para. 11, tr. auct. Cf. Hu Tao-Ching (*1*), vol. 2, p. 756.

‡ Sivin suggests that "Magicians" is an error for "Technicians".—*Ed.*

18 然常微偏東不全南也

19 I shall never forget the excitement which I experienced when I first read these words. If any one text stimulated the writing of this book more than any other, this was it.

20 Ch. 24, para. 18; tr. auct. adjuv. Klaproth (1); Biot (14); Wylie (11); Hirth (3). Cf. Hu Tao-Ching (*1*), vol. 2, pp. 768ff.

Text 11a

from

The Mathematical Universe

by William Dunham

CHAPTER G
GREEK GEOMETRY

Geometry, once the cornerstone of mathematics, was introduced in Chapter C. In this 1
and the following two chapters, we take a deeper look at this ancient and beautiful subject. And what better place to start than with geometry's finest practitioners, the mathematicians of classical Greece?

Greek geometry ranks as a major achievement of the human intellect for 2
reasons both mathematical and historical, both practical and aesthetic. Its golden age stretched from Thales of Miletus of around 600 B.C. down to the second century B.C. work of Eratosthenes, Apollonius, and the unparalleled Archimedes of Syracuse. From there a somewhat less distinguished "silver age" continued to the time of Pappus around A.D. 300. These individuals, and many others, developed geometry from a practical method of measuring land (*geo* = earth, *metria* = measure) to a vast body of abstract theorems and constructions laced together by the uncompromising rules of logic. Greek geometry stands among the great intellectual / artistic movements of Western civilization and in this sense has much in common with Elizabethan drama or French impressionism. Like the impressionists, the Greek geometers shared a general philosophy and style, and

though there were as many variants among the Greeks as among the French artists, the deeper, unifying characteristics of an impressionist painting or a Greek theorem are instantly recognizable.

3 What are these characteristics? Historian Ivor Thomas, in his comprehensive *Greek Mathematical Works*, singles out (1) the impressive logical rigor the Greeks employed for proving theorems, (2) the purely geometric—as opposed to numerical—nature of their mathematics, and (3) their skillful organization in presenting and developing mathematical propositions.

4 To these characteristics, we add two others. One is their recognition of geometry as an unsurpassed exercise in pure thought, a subject at once ideal, immaterial, and eternal. In the *Republic*, Plato noted that although geometers draw tangible figures as an aid to their investigations,

> they are not thinking about these figures but of those things which the figures represent; thus it is the square in itself and the diameter in itself which are the matter of their arguments, not that which they draw; similarly, when they model or draw objects, which may themselves have images in shadows or in water, they use them in turn as images, endeavoring to see those absolute objects which cannot be seen otherwise than by thought.

Such a view, of course, fits nicely with the Platonic concept of an ideal existence apart from the human experience, and geometrical considerations surely played a role in shaping his philosophy. The Greek thinkers—seeking the perfect, the logical, and the utterly rational—could look to geometry as the embodiment of this ideal.

5 Less earthshaking, but nonetheless central to much of Greek mathematics, was a reliance on compass and straightedge for geometric constructions. On the one hand, these were two readily available tools for drawing the tangible figures of which Plato talked. But in a more abstract sense, these tools enshrined the straight line (via the straightedge) and the circle (via the compass) as the keys to geometric existence. With the unwavering precision of the ideal straight line and the perfect symmetry of the ideal circle, the Greeks created their geometric figures and, from there, their geometric theorems. And although we today have extended mathematics beyond the limitations of lines and circles, their supremacy for the Greek mathematicians was utterly in character.

There is no question that geometric ideas predated the Greeks. The civilizations 6 of Egypt and Mesopotamia, for instance, used geometry to partition fields and erect pyramids, and we shall return to this topic in Chapter O. But it is to the Greeks that we look for the first geometric theorems, the first propositions proved with logical rigor.

According to tradition, Greece's earliest mathematician, not to mention its 7 earliest astronomer and philosopher, was Thales (pronounced THAY-leez) who grew up along the eastern Aegean shores watching the swimming whales (*not* pronounced WHAY-leez). According to the later commentator Proclus, "Thales was the first to go to Egypt and bring back to Greece this study; he discovered many propositions and disclosed the underlying principles of many others to his successors."

Legend holds that it was Thales who first proved that the base angles of an 8 isosceles triangle are congruent and that any angle inscribed in a semicircle is a right angle (the latter is sometimes given the tongue-twisting name Thales' theorem). Unfortunately, legend is all we have to go on, for his actual proofs disappeared long ago. Still, the ancients held him in very high regard, classifying him as one of the "seven wise men of antiquity." (There is no truth to the rumor that the other six were Grumpy, Happy, Dopey, Sneezy, Doc, and Bashful.)

With Thales, Greek geometry was underway. To trace its subsequent 9 developments, its successes and failures, would occupy countless chapters, if not countless volumes. So here we limit ourselves to two specific geometric issues: how Euclid did geometry with a collapsing compass and why the Epicurean philosophers accused him of being no smarter than an ass. Although these choices may seem a bit odd, they give an accurate sense of the mathematical temperament of the times.

We begin around the year 300 B.C. with Euclid of Alexandria. Although 10 responsible for a number of mathematical treatises, he is best remembered for the *Elements*, a systematic development of much of Greek mathematics up to that point. The work is divided into 13 books and contains 465 propositions on plane and solid geometry and number theory. Aptly called the greatest mathematics textbook of all time, the work has been studied, edited, and revered from its appearance in ancient Greece to the present day.

What makes the *Elements* so important is its logical development from 11 basic principles to sophisticated consequences. Euclid began Book I with a list of

23 definitions so the reader would know precisely what his terms meant. He introduced ***point*** as “that which has no part” (one of his less illuminating definitions); ***equilateral triangle*** as a triangle “which has its three sides equal”; and ***isosceles triangle*** as a triangle “which has two of its sides alone equal.”

12 With terms defined, Euclid presented five postulates to serve as the foundations of his geometry, the starting points from which all would follow. These were given without proof or justification; they were simply to be accepted. Fortunately, such acceptance was not difficult because the postulates appeared to Euclid’s contemporaries, and indeed to most of us today, as utterly innocuous. For our purposes in this chapter, we need only the first three:

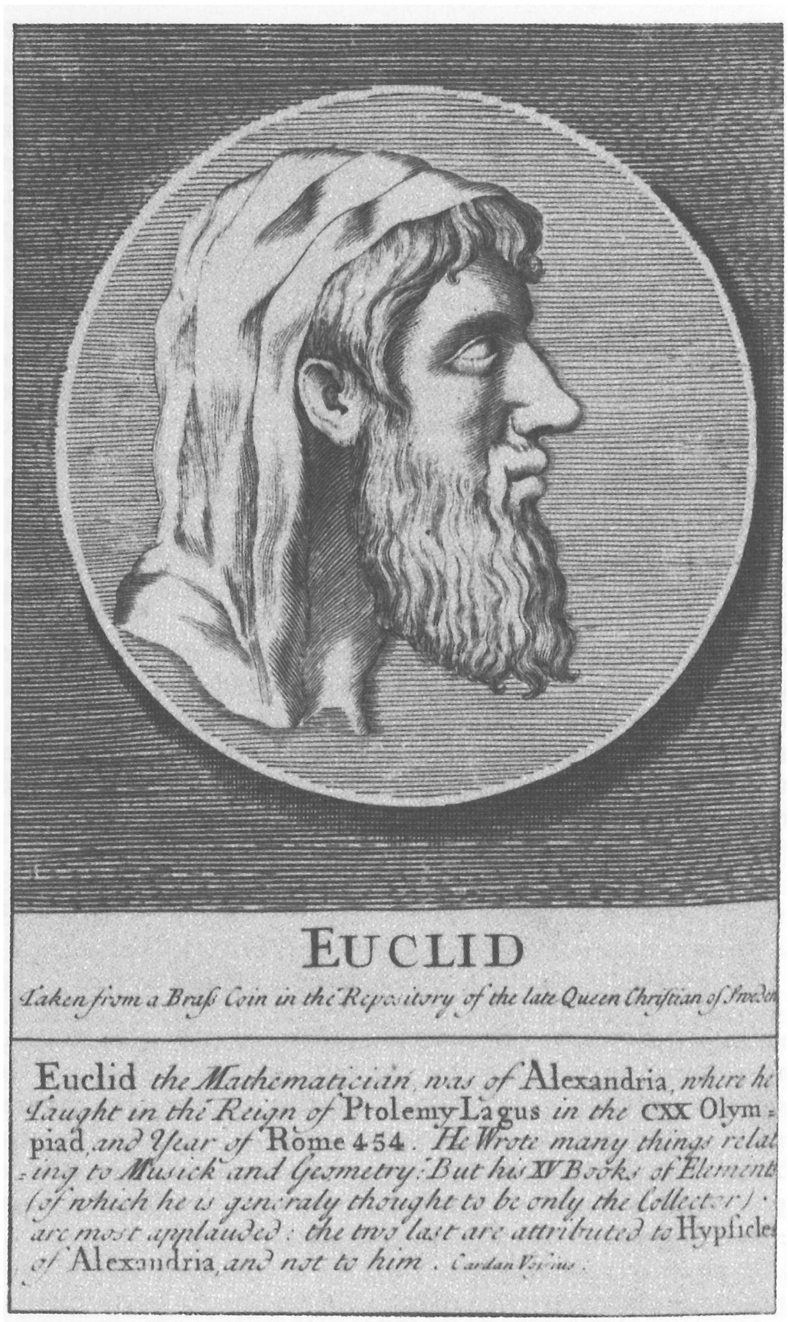
1. [lt is possible] to draw a straight line from any point to any point.
2. [lt is possible] to produce a finite straight line continuously in a straight line.
3. [lt is possible] to describe a circle with any center and distance.

13 These seem quite simple, quite self-evident. The first two legitimize the use of the unmarked straightedge in geometric constructions, for they allow us to connect two points with a straight line (postulate l) or to take an existing line and extend it (postulate 2). This is precisely what straightedges are good for. The third postulate authorizes what we do with a compass: draw a circle with a given point as center and with a predetermined length as radius. Thus, it appears that the first three postulates provide logical support for the operation of the geometric tools.

14 But the reader may think back to her or his own geometry classes and recall another operation performed with a compass: the transfer of a length from one part of the plane to another. This is easily done. We put the points of the compass at either end of the line segment whose length is to be transferred, lock the compass into place, pick it up, move it rigidly, then set it down at its desired destination. It is a process at once simple and necessary in many geometric constructions.

15 Yet Euclid included no postulate to justify transferring lengths in this fashion. Where we expect an axiomatic dispensation allowing this procedure, we find nothing. Although his compass could draw a circle, he did not explicitly permit it to be locked into position and moved about. Euclid’s is thus somewhat facetiously

called a "collapsible compass," a device that falls shut the instant it is lifted from the page.



16 This raises a serious question of logic: Did the esteemed Greek geometer forget to include a "transfer of lengths" postulate? Do we have here a Euclidean blunder?

17 Not at all. As we shall see in a moment, Euclid had his reasons—reasons logically sound and very Greek in nature—for *not* including such a postulate. Rather than a blunder, this omission is evidence of his geometric acumen and organizational ability.

18 With postulates in place, Euclid introduced a few "common notions": self-evident statements of a more general and less geometric nature. For instance, here we accept without proof that "Things which are equal to the same thing are also equal to one other," that "If equals be added to equals, the wholes are equal," and that "The whole is greater than the part." These are statements with which few will quibble.

19 Then he was ready to plunge in. With a huge body of geometry to deduce from a tiny collection of definitions, postulates, and common notions, where does one start? This is the sort of initial challenge known to freeze mathematicians (and authors) in their tracks. But if, as the Chinese tell us, a journey of a thousand miles begins with a single step, then Euclid's journey through geometry began with an equilateral triangle. The very first proposition of the *Elements* was the construction, upon a given line segment, of just such a figure.

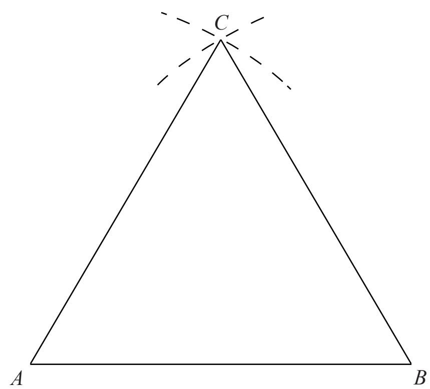


Figure G.1

20 The argument is easy. Starting with given segment AB in Figure G.l, we construct a circle with center at A and radius AB as permitted by postulate 3. Next, with center B and radius AB, we invoke the same postulate to construct a second circle.

Let C be the point where the circular arcs intersect (see the chapter notes regarding the existence of such a point of intersection). We then draw lines AC and BC by postulate 1, forming ΔABC. In this triangle, AB and AC are the same length because they are radii of the first circle; and AB and BC are the same length since they are radii of the second. Because things equal to the same thing are equal to each other, all three sides are mutually equal. By Euclid's definition, the triangle is equilateral and the proof is complete.

It is critical to observe that in using the compass for this construction, Euclid never needed to move it rigidly. After each arc was drawn, the compass can fall to pieces without affecting the proof in the least. 21

But in the next two propositions of Book I, Euclid showed how a length could be transferred *even with a compass that collapses*. This means that length transfer was implied by the postulates already on the table. A new postulate to this end would have been unnecessary baggage, and Euclid was sharp enough to realize this. 22

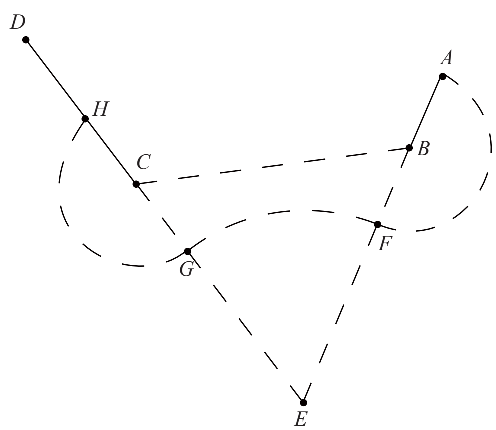


Figure G.2

His proofs—here combined into a single argument—were quite elegant. Suppose we have segment AB as shown in Figure G.2 and wish to transfer its length onto the segment CD emanating from the point C. First, use the straightedge and apply postulate 1 to draw the segment connecting B and C. Then upon segment BC construct equilateral triangle BCE; the legitimacy of this construction, of course, is exactly what was established in the previous proposition. 23

We then engage in a round of circle drawing. With center B and radius AB, construct a circle meeting BE at F (upon picking up the compass, suppose it 24 collapses). With center E and radius EF, draw a circle meeting CE at G (again, the compass may collapse as we lift it from the page). With center C and radius CG, draw a final circle meeting CD at H. All of these constructions are permitted by Euclid's third postulate; none require a rigid compass.

25 Now we merely generate a string of equalities (for ease of notation, we shall denote the *length* of segment XY by $\overline{XY}$):

$$
\begin{aligned}
\overline{AB} &= \overline{BF} && \text{because they are radii of the same circle}\\
&= \overline{BE} - \overline{EF} && \\
&= \overline{BE} - \overline{EG} && \text{because } EF \text{ and } EG \text{ are radii of the same circle}\\
&= \overline{CE} - \overline{EG} && \text{because the three sides of } \Delta BCE \text{ are equally long}\\
&= \overline{CG} && \\
&= \overline{CH} && \text{because again we have radii of the same circle}
\end{aligned}
$$

The beginning and end of this chain reveal that $\overline{AB} = \overline{CH}$. Therefore the initial length of AB has been transferred onto segment CD as required, yet nowhere did we have to pick up a compass and move it rigidly.

26 The surprising conclusion of this proof is that constructions seeming to require a noncollapsing compass can actually be accomplished with a collapsible one. As he subsequently developed his geometry, Euclid could therefore legitimately transfer a length from one place to another as if with a rigid compass, his rationale being the just-proved theorem. By getting it out of the way so early, and so simply, he was free to use it in all that followed.

27 At this point some readers may be stifling a yawn, regarding the whole business as much ado about nothing. After all, everyone knows that stationery stores sell cheap metal compasses that are made to stay open, and surely it would have done no great damage for Euclid to have included an extra postulate to that effect.

28 Adherents of this position, we believe, have not yet gotten into the spirit of formal Greek geometry. First, the real-world existence of rigid compasses has no bearing on the development of ideal concepts. Second, stationery stores had not yet been invented. Third, and most crucially, Euclid would not have wanted to add an *unnecessary* postulate to his list. Why assume something that could be derived from other assumptions? It would make his postulates less pure, less streamlined, and less perfect and thereby violate an aesthetic, not a mathematical, principle.

That aesthetic considerations were critical for the Greek mathematician is obvious. In Euclid's proof above we glimpse what Ivor Thomas meant when he wrote:

> [A] feature which cannot fail to impress a modern mathematician is the perfection of form in the work of the great Greek geometers. This perfection of form, which is another expression of the same genius that gave us the Parthenon and the plays of Sophocles, is found equally in the proof of individual propositions and in the ordering of those separate propositions into books; it reaches its height, perhaps, in the *Elements* of Euclid.

We now move deeper into Book I to see further evidence of Euclid's genius. 29
After disposing of the collapsible compass in propositions 2 and 3, Euclid proved the so-called side-angle-side, or SAS, congruence scheme in proposition 4. That is (see Figure G.3), if we have triangles ABC and DEF in which $\overline{AB}=\overline{DE}$, $\overline{AC}=\overline{DF}$, and the included angles $\alpha = \delta$, then the triangles are congruent, meaning they have precisely the same size and shape. In other words, if ΔDEF were picked up and placed atop ΔABC, the two would coincide perfectly, line for line, angle for angle, point for point.

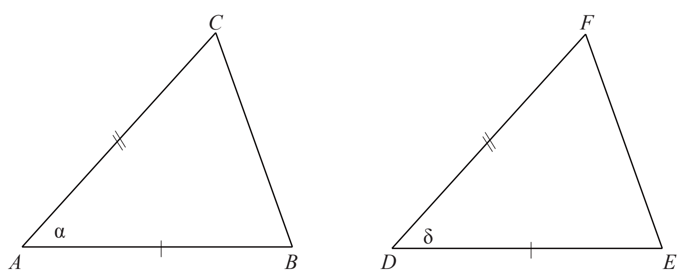


Figure G.3

In Euclid's hands, triangle congruence was *the* key to proving geometric 30
propositions. Later he established the additional congruence patterns of side-side-side (or SSS) in proposition 8, and angle-side-angle (ASA) and angle-angle-side (AAS) in proposition 26.

Proposition 5 of Book I proved that the base angles of an isosceles triangle 31
are congruent. As noted, this result has been attributed to Thales, but the proof in the *Elements* is probably Euclid's own. Although we shall not give it here, we note that it was accompanied by the diagram in Figure G.4. This configuration,

which suggests a bridge (at least to those with vivid imaginations) may account for calling proposition 5 the *pons asinorum*, or bridge of asses. According to tradition, dullards—that is, asses—find the proof beyond them and thus cannot cross this logical bridge into the geometric promised land of the *Elements*.

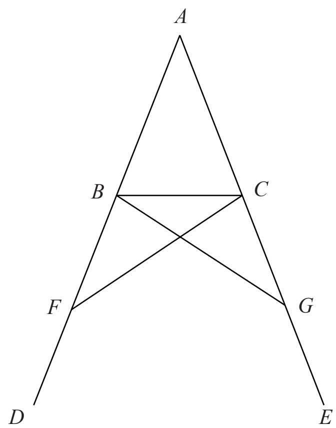


Figure G.4

32 If weak students were likened to asses, Euclid himself met a similar fate at the hands of the Epicureans for his proof of proposition 20. To see why, we must describe a few of the intervening theorems of Book I.

33 After crossing the *pons asinorum*, Euclid showed how to bisect angles and construct perpendiculars with compass and straightedge and soon arrived at one of the critical theorems of Book I, commonly known as the exterior angle theorem. This result, which appeared as proposition 16, guaranteed that the exterior angle of any triangle exceeds either of the opposite and interior angles. That is (see Figure G.5), if we begin with ΔABC and extend side BC rightward to D, then both α and β are less than $\angle ACD$.

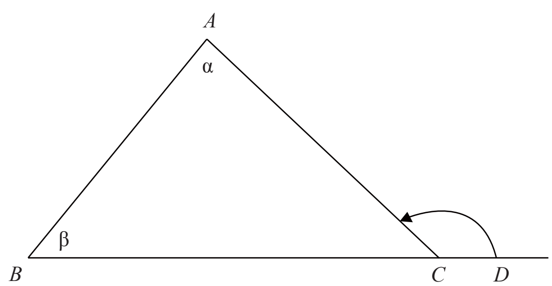


Figure G.5

The exterior angle theorem was the first geometric inequality in the *Elements*. 34
Where Euclid had previously shown sides or angles to be equal (as in the *pons asinorum*), he here proved certain angles to be unequal. This theorem would play a significant role in the remainder of Book I.

This brings us to another inequality, proposition 19, whose diagram appears 35
in Figure G.6. Euclid stated it as, "In any triangle the greater angle is subtended by the greater side," which in more modern notation becomes:

PROPOSITION 19: In ΔABC, if $\beta > \alpha$, then $\overline{AC} > \overline{BC}$ (i.e., $b > a$).
PROOF: Here we assume that $\beta > \alpha$. It is our job to prove that side AC opposite $\angle ABC$ is longer than side BC opposite $\angle BAC$.

Euclid separately considered the three possible cases: $b = a$, $b < a$, and $b > a$. 36
His strategy was to show that the first two are impossible and thereby conclude that the third case must hold, as the theorem asserted. Such a technique is called a double *reductio ad absurdum*, or a proof by double contradiction. This powerful logical strategy was nowhere better employed than in Greek mathematics. Here is how Euclid handled it:

Case 1: Suppose $b = a$. 37

Referring to Figure G.6, we have $\overline{BC} = a = b = \overline{AC}$. This makes ΔABC isosceles and so, invoking the *pons asinorum*, we conclude that the base angles are themselves congruent. That is, $\angle BAC = \angle ABC$, or equivalently $\alpha = \beta$. But this contradicts the initial assumption that $\beta > \alpha$. Hence we dismiss Case l as impossible.

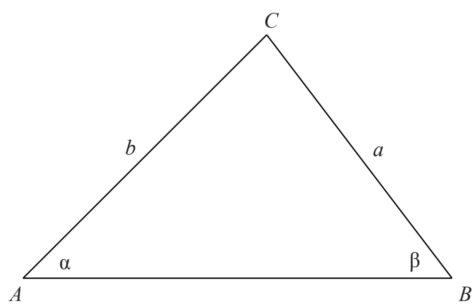


Figure G.6

38 ***Case 2:*** Suppose $b < a$.

Here we have the situation depicted in Figure G.7. Because AC is assumed to be shorter than BC, we can construct segment CD of length b where D falls within the longer side BC. Then draw AD to form ΔADC. This triangle, having two sides of length b, is isosceles and therefore has congruent base angles $\angle DAC$ and $\angle ADC$. But applying the exterior angle theorem to the narrow ΔABD, we deduce that

$$\begin{aligned} \beta &= \text{interior } \angle ABD \\ &< \text{exterior } \angle ADC && \text{by the exterior angle theorem} \\ &= \angle DAC && \text{because } \Delta DAC \text{ is isosceles} \\ &< \angle BAC && \text{because the whole is greater than the part} \\ &= \alpha \end{aligned}$$

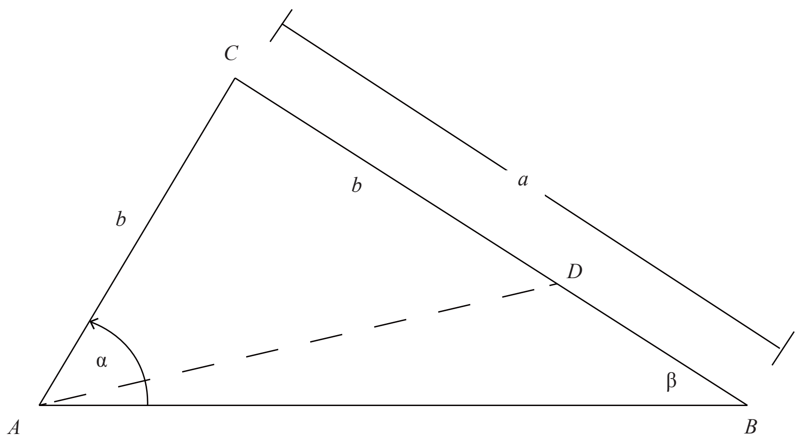


Figure G.7

In other words, $\beta < \alpha$, which contradicts the theorem's original stipulation that $\beta > \alpha$. Case 2, leading as it does to a contradiction, also bites the dust. We are left with:

39 ***Case 3:*** $b > a$.

This must be true because no alternatives remain, and the theorem is proved.■

40 We now have reached the proposition that so troubled the Epicurean philosophers. On the surface it sounds harmless enough:

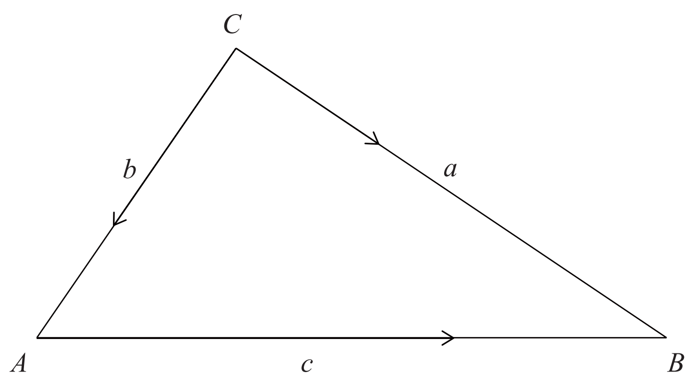
PROPOSITION 20: In any triangle two sides taken together in any manner are greater than the remaining one.

Why the controversy? Why the derision? We quote the commentator Proclus: 41

> The Epicureans are wont to ridicule this theorem, saying it is evident even to an ass and needs no proof; it is as much the mark of an ignorant man, they say, to require persuasion of evident truths as to believe what is obscure without question . . . . That the present theorem is known to an ass they make out from the observations that, if straw is placed at one extremity of the sides, an ass in quest of provender will make his way along the one side and not by way of the two others.

In short, even a dumb animal knows to take the straight route from *C* to *B* in Figure G.8 rather than go the long way around via *A*. So why, the Epicureans asked, did Euclid bother to prove something so blatantly obvious? Proclus gave an answer: 42

> It should be replied that, granting the theorem is evident to sense-perception, it is still not clear for scientific thought. Many things have this character; for example, that fire warms. This is clear to perception, but it is the task of science to find out how it warms.



$b + c > a$

Figure G.8

43 In the Euclidean spirit, the spirit that so typified Greek geometry, we must employ our faculties of reason to demonstrate that which an ass knows by instinct. Even a seemingly self-evident proposition cries out for a proof, and this Euclid was only too happy to provide. Building upon previous results, he reasoned as follows:

PROPOSITION 20: In ΔABC, $\overline{AC}+\overline{AB}>\overline{BC}$ (i.e., $b+c>a$)
PROOF: In Figure G.9, extend side BA to D so that $\overline{AD}=\overline{AC}=b$ and therefore $\overline{BD}=b+c$. This construction generates ΔDAC, which is isosceles because it has two sides of length b. Considering the large ΔBDC, we note that

$\angle BCD$	$>$	$\angle ACD$	because the whole is greater than the part
	$=$	$\angle BDC$	because these are base angles of isosceles ΔDAC

So $\angle BCD$ is greater than $\angle BDC$. As Euclid had just proved that the greater side is opposite the greater angle, it follows that $\overline{BD}>\overline{BC}$; in other words, $b+c>a$, which is exactly what was to be proved.■

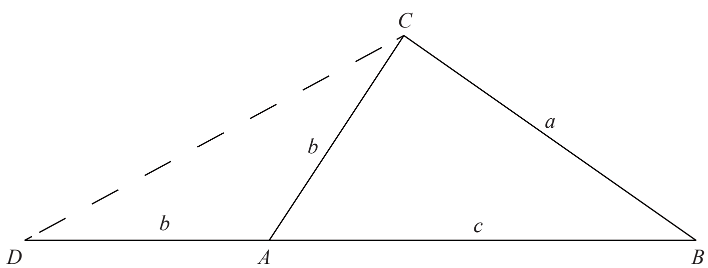


Figure G.9

44 This is a splendid little proof. It has its subtleties, its clever use of inequalities, its quiet elegance.

45 In Sir Arthur Conan Doyle's *A Study in Scarlet*, Dr. Watson described the deductive powers of Sherlock Holmes in these words: "His conclusions were as infallible as so many propositions of Euclid." Watson was not alone in his high opinion of the Greek geometer. Centuries ago, the Arabic scholar al-Qifti said of Euclid, "nay, there was no one even of later date who did not walk in his footsteps," and the incomparable Albert Einstein added his own tribute: "If Euclid failed to kindle your youthful enthusiasm, then you were not born to be a scientific thinker."

What we have considered thus far, of course, is just the tip of the iceberg, just a sampler of what historian Morris Kline calls the Greeks' "grand exercise in logic." We must leave them at this point. But in a sense, no mathematician can leave behind the legacy of the classical geometers. They gave demonstrative mathematics its start, honed its logical tools, and aimed it in directions it has traveled ever since. We end with the words of the twentieth-century British mathematician G. H. Hardy: "The Greeks . . . spoke a language which modern mathematicians can understand; as Littlewood said to me once, they are not clever schoolboys or 'scholarship candidates,' but 'Fellows of another college.' " 46

There is no arguing with Hardy when he says, "Greek mathematics is the real thing." 47

Text 11b

from

Elements

by Euclid

BOOK I

Definitions

1. A *point* is that which has no part.
2. A *line* is breadthless length.
3. The extremities of a line are points.
4. A *straight line* is a line which lies evenly with the points on itself.
5. A *surface* is that which has length and breadth only.
6. The extremities of a surface are lines.
7. A *plane surface* is a surface which lies evenly with the straight lines on itself.
8. A *plane angle* is the inclination to one another of two lines in a plane which meet one another and do not lie in a straight line.
9. And when the lines containing the angle are straight, the angle is called *rectilineal*.
10. When a straight line set up on a straight line makes the adjacent angles equal to one another, each of the equal angles is *right*, and the straight line standing on the other is called a *perpendicular* to that on which it stands.
11. An *obtuse angle* is an angle greater than a right angle.
12. An *acute angle* is an angle less than a right angle.

From Euclid, *The Thirteen Books of Euclid's Elements*, translated by Thomas L. Heath (1956), made available by the Perseus Digital Library Project. Ed. Gregory R. Crane. Tufts University.

13. A *boundary* is that which is an extremity of anything.
14. A *figure* is that which is contained by any boundary or boundaries.
15. A *circle* is a plane figure contained by one line such that all the straight lines falling upon it from one point among those lying within the figure are equal to one another;
16. And the point is called the *centre* of the circle.
17. A *diameter* of the circle is any straight line drawn through the centre and terminated in both directions by the circumference of the circle, and such a straight line also bisects the circle.
18. A *semicircle* is the figure contained by the diameter and the circumference cut off by it. And the centre of the semicircle is the same as that of the circle.
19. *Rectilineal figures* are those which are contained by straight lines, *trilateral* figures being those contained by three, *quadrilateral* those contained by four, and *multilateral* those contained by more than four straight lines.
20. Of trilateral figures, an *equilateral triangle* is that which has its three sides equal, an *isosceles triangle* that which has two of its sides alone equal, and a *scalene triangle* that which has its three sides unequal.
21. Further, of trilateral figures, a *right-angled triangle* is that which has a right angle, an *obtuse-angled triangle* that which has an obtuse angle, and an *acute-angled triangle* that which has its three angles acute.
22. Of quadrilateral figures, a *square* is that which is both equilateral and right-angled; an *oblong* that which is right-angled but not equilateral; a *rhombus* that which is equilateral but not right-angled; and a *rhomboid* that which has its opposite sides and angles equal to one another but is neither equilateral nor right-angled. And let quadrilaterals other than these be called *trapezia*.
23. *Parallel* straight lines are straight lines which, being in the same plane and being produced indefinitely in both directions, do not meet one another in either direction.

Postulates

Let the following be postulated:

1. To draw a straight line from any point to any point.
2. To produce a finite straight line continuously in a straight line.
3. To describe a circle with any centre and distance.

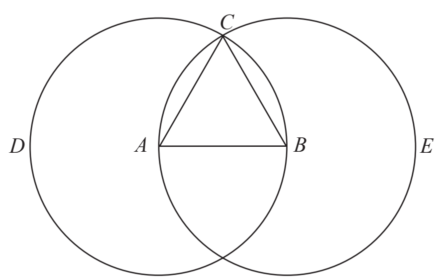
4. That all right angles are equal to one another.
5. That, if a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

Common Notions

1. Things which are equal to the same thing are also equal to one another.
2. If equals be added to equals, the wholes are equal.
3. If equals be subtracted from equals, the remainders are equal.
4. Things which coincide with one another are equal to one another.
5. The whole is greater than the part.

Propositions

Proposition 1.
On a given finite straight line to construct an equilateral triangle.



Let *AB* be the given finite straight line.

Thus it is required to construct an equilateral triangle on the straight line *AB*.

With centre *A* and distance *AB* let the circle *BCD* be described; [Post. 3]
again, with centre *B* and distance *BA* let the circle *ACE* be described; [Post. 3]
and from the point *C*, in which the circles cut one another, to the points *A*, *B* let the straight lines *CA*, *CB* be joined. [Post. 1]

Now, since the point *A* is the centre of the circle *CDB*,
AC is equal to *AB*. [Def. 15]

Again, since the point *B* is the centre of the circle *CAE*,
BC is equal to *BA*. [Def. 15]

But *CA* was also proved equal to *AB*;
therefore each of the straight lines *CA*, *CB* is equal to *AB*.

And things which are equal to the same thing are also equal to one another; [C.N. 1]

therefore *CA* is also equal to *CB*.

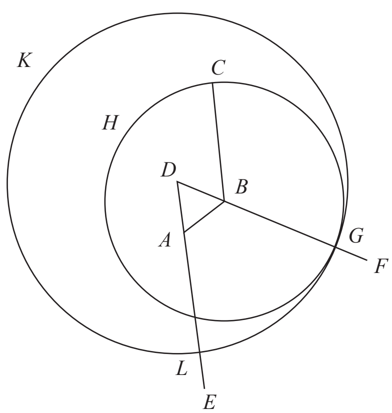
Therefore the three straight lines *CA*, *AB*, *BC* are equal to one another.

Therefore the triangle *ABC* is equilateral; and it has been constructed on the given finite straight line *AB*.

Being what it was required to do.

Proposition 2.

To place at a given point as an extremity a straight line equal to a given straight line.



Let *A* be the given point, and *BC* the given straight line.

Thus it is required to place at the point *A* as an extremity a straight line equal to the given straight line *BC*.

From the point *A* to the point *B* let the straight line *AB* be joined; [Post. 1]

and on it let the equilateral triangle *DAB* be constructed. [I. 1]

Let the straight lines *AE*, *BF* be produced in a straight line with *DA*, *DB*; [Post. 2]

with centre *B* and distance *BC* let the circle *CGH* be described; [Post. 3]

and again, with centre *D* and distance *DG* let the circle *GKL* be described. [Post. 3]

Then, since the point *B* is the centre of the circle *CGH*,

BC is equal to *BG*.

Again, since the point *D* is the centre of the circle *GKL*,

DL is equal to *DG*.

And in these *DA* is equal to *DB*;

therefore the remainder *AL* is equal to the remainder *BG*. [C.N. 3]

But *BC* was also proved equal to *BG*;

therefore each of the straight lines *AL*, *BC* is equal to *BG*.

And things which are equal to the same thing are also equal to one another; [C.N. 1]

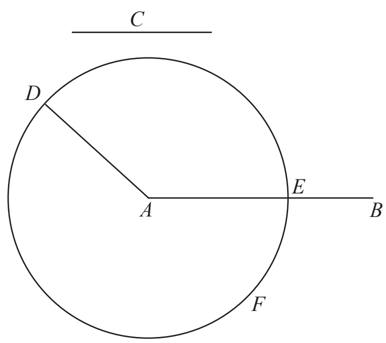
therefore *AL* is also equal to *BC*.

Therefore at the given point *A* the straight line *AL* is placed equal to the given straight line *BC*.

Being what it was required to do.

Proposition 3.

Given two unequal straight lines, to cut off from the greater a straight line equal to the less.



Let *AB*, *C* be the two given unequal straight lines, and let *AB* be the greater of them.

Thus it is required to cut off from *AB* the greater a straight line equal to *C* the less.

At the point *A* let *AD* be placed equal to the straight line *C*; [I. 2]

and with centre *A* and distance *AD* let the circle *DEF* be described. [Post. 3]

Now, since the point *A* is the centre of the circle *DEF*,

AE is equal to *AD*. [Def. 15]

But *C* is also equal to *AD*.

Therefore each of the straight lines *AE*, *C* is equal to *AD*; so that *AE* is also equal to *C*. [C.N. 1]

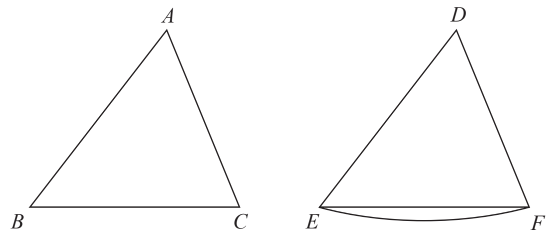
Therefore, given the two straight lines *AB*, *C*, from *AB* the greater *AE* has been cut off equal to *C* the less.

Being what it was required to do.

Proposition 4.

If two triangles have the two sides equal to two sides respectively, and have the angles contained by the equal straight lines equal, they will also have the base equal to the base, the triangle will be equal to the triangle, and the remaining angles will be equal to the remaining angles respectively, namely those which the equal sides subtend.

Let *ABC*, *DEF* be two triangles having the two sides *AB*, *AC* equal to the two sides *DE*, *DF* respectively, namely *AB* to *DE* and *AC* to *DF*, and the angle *BAC* equal to the angle *EDF*.



I say that the base *BC* is also equal to the base *EF*, the triangle *ABC* will be equal to the triangle *DEF*, and the remaining angles will be equal to the remaining angles respectively, namely those which the equal sides subtend, that is, the angle *ABC* to the angle *DEF*, and the angle *ACB* to the angle *DFE*.

For, if the triangle *ABC* be applied to the triangle *DEF*,

and if the point *A* be placed on the point *D*

and the straight line *AB* on *DE*,

then the point *B* will also coincide with *E*, because *AB* is equal to *DE.*

Again, *AB* coinciding with *DE*,

the straight line *AC* will also coincide with *DF*, because the angle *BAC* is equal to the angle *EDF*;

hence the point *C* will also coincide with the point *F*, because *AC* is again equal to *DF*.

But *B* also coincided with *E*;

hence the base *BC* will coincide with the base *EF* [. . .]

and will be equal to it. [C.N. 4]

Thus the whole triangle *ABC* will coincide with the whole triangle *DEF*,

and will be equal to it.

And the remaining angles will also coincide with the remaining angles and will be equal to them,

the angle *ABC* to the angle *DEF*,

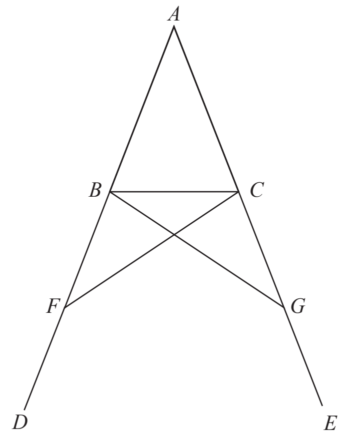
and the angle *ACB* to the angle *DFE*.

Therefore etc.

Being what it was required to prove.

Proposition 5.

In isosceles triangles the angles at the base are equal to one another, and, if the equal straight lines be produced further, the angles under the base will be equal to one another.



Let *ABC* be an isosceles triangle having the side *AB* equal to the side *AC*;
and let the straight lines *BD*, *CE* be produced further in a straight line with *AB*, *AC*. [Post. 2]

I say that the angle *ABC* is equal to the angle *ACB*, and the angle *CBD* to the angle *BCE*.

Let a point *F* be taken at random on *BD*;
from *AE* the greater let *AG* be cut off equal to *AF* the less; [I. 3]
and let the straight lines *FC*, *GB* be joined. [Post. 1]

Then, since *AF* is equal to *AG* and *AB* to *AC*,
the two sides *FA*, *AC* are equal to the two sides *GA*, *AB*, respectively;
and they contain a common angle, the angle *FAG*.
Therefore the base *FC* is equal to the base *GB*,
and the triangle *AFC* is equal to the triangle *AGB*,
and the remaining angles will be equal to the remaining angles respectively, namely those which the equal sides subtend,
that is, the angle *ACF* to the angle *ABG*,
and the angle *AFC* to the angle *AGB*. [I. 4]

And, since the whole *AF* is equal to the whole *AG*,
and in these *AB* is equal to *AC*,
the remainder *BF* is equal to the remainder *CG*.

But *FC* was also proved equal to *GB*;
therefore the two sides *BF*, *FC* are equal to the two sides *CG*, *GB* respectively;
and the angle *BFC* is equal to the angle *CGB*,
while the base *BC* is common to them;
therefore the triangle *BFC* is also equal to the triangle *CGB*,
and the remaining angles will be equal to the remaining angles respectively, namely those which the equal sides subtend;
therefore the angle *FBC* is equal to the angle *GCB*,
and the angle *BCF* to the angle *CBG*.

Accordingly, since the whole angle *ABG* was proved equal to the angle *ACF*,
and in these the angle *CBG* is equal to the angle *BCF*,
the remaining angle *ABC* is equal to the remaining angle *ACB*;

and they are at the base of the triangle *ABC*.

But the angle *FBC* was also proved equal to the angle *GCB*;

and they are under the base.

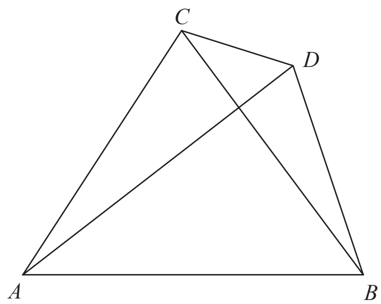
Therefore etc.

Q. E. D.

[. . .]

Proposition 7.

Given two straight lines constructed on a straight line from its extremities and meeting in a point, there cannot be constructed on the same straight line from its extremities, and on the same side of it, two other straight lines meeting in another point and equal to the former two respectively, namely each to that which has the same extremity with it.



For, if possible, given two straight lines *AC*, *CB* constructed on the straight line *AB* and meeting at the point *C*, let two other straight lines *AD*, *DB* be constructed on the same straight line *AB*, on the same side of it, meeting in another point *D* and equal to the former two respectively, namely each to that which has the same extremity with it, so that *CA* is equal to *DA* which has the same extremity *A* with it, and *CB* to *DB* which has the same extremity *B* with it; and let *CD* be joined.

Then, since *AC* is equal to *AD*,

the angle *ACD* is also equal to the angle *ADC*; [I. 5]

therefore the angle *ADC* is greater than the angle *DCB*;

therefore the angle *CDB* is much greater than the angle *DCB*.

Again, since *CB* is equal to *DB*,

the angle *CDB* is also equal to the angle *DCB*.

But it was also proved much greater than it:

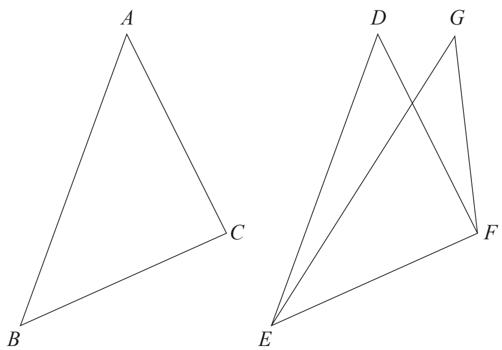
which is impossible.

Therefore etc.

Q. E. D.

Proposition 8.

If two triangles have the two sides equal to two sides respectively, and have also the base equal to the base, they will also have the angles equal which are contained by the equal straight lines.



Let *ABC*, *DEF* be two triangles having the two sides *AB*, *AC* equal to the two sides *DE*, *DF* respectively, namely *AB* to *DE*, and *AC* to *DF*; and let them have the base *BC* equal to the base *EF*;

I say that the angle *BAC* is also equal to the angle *EDF*.

For, if the triangle *ABC* be applied to the triangle *DEF*, and if the point *B* be placed on the point *E* and the straight line *BC* on *EF*,

the point *C* will also coincide with *F*,

because *BC* is equal to *EF*.

Then, *BC* coinciding with *EF*,

BA, *AC* will also coincide with *ED*, *DF*;

for, if the base *BC* coincides with the base *EF*, and the sides *BA*, *AC* do not coincide with *ED*, *DF* but fall beside them as *EG*, *GF*,

then, given two straight lines constructed on a straight line from its extremities and meeting in a point, there will have been constructed on the same straight line from its extremities, and on the same side of it, two other straight lines meeting in another point and equal to the former two respectively, namely each to that which has the same extremity with it.

But they cannot be so constructed. [I. 7]

Therefore it is not possible that, if the base *BC* be applied to the base *EF*, the sides *BA*, *AC* should not coincide with *ED*, *DF*;

they will therefore coincide,

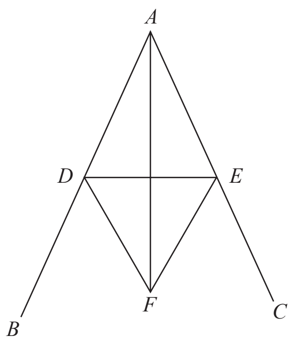
so that the angle *BAC* will also coincide with the angle *EDF*, and will be equal to it.

If therefore etc.

Q. E. D.

Proposition 9.

To bisect a given rectilineal angle.



Let the angle *BAC* be the given rectilineal angle.

Thus it is required to bisect it.

Let a point *D* be taken at random on *AB*; let *AE* be cut off from *AC* equal to *AD*; [I. 3]
let *DE* be joined, and on *DE* let the equilateral triangle *DEF* be constructed;
let *AF* be joined.

I say that the angle *BAC* has been bisected by the straight line *AF*.

For, since *AD* is equal to *AE*,
and *AF* is common,

the two sides *DA*, *AF* are equal to the two sides *EA*, *AF* respectively.

And the base *DF* is equal to the base *EF*;

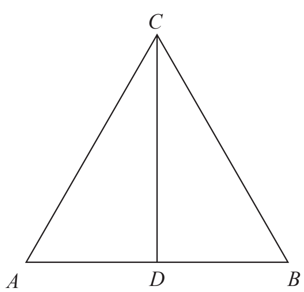
therefore the angle *DAF* is equal to the angle *EAF*. [I. 8]

Therefore the given rectilineal angle *BAC* has been bisected by the straight line *AF*.

Q. E. F.

Proposition 10.

To bisect a given finite straight line.



Let *AB* be the given finite straight line.

Thus it is required to bisect the finite straight line *AB*.

Let the equilateral triangle *ABC* be constructed on it, [I. 1]
and let the angle *ACB* be bisected by the straight line *CD*; [I. 9]

I say that the straight line *AB* has been bisected at the point *D*.

For, since *AC* is equal to *CB*, and *CD* is common,

the two sides *AC*, *CD* are equal to the two sides *BC*, *CD* respectively;

and the angle *ACD* is equal to the angle *BCD*; [I. 4]

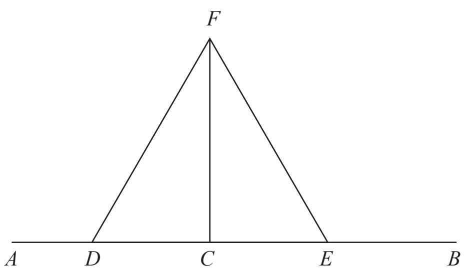
therefore the base *AD* is equal to the base *BD*.

Therefore the given finite straight line *AB* has been bisected at *D*.

Q. E. F.

Proposition 11.

To draw a straight line at right angles to a given straight line from a given point on it.



Let *AB* be the given straight line, and *C* the given point on it.

Thus it is required to draw from the point *C* a straight line at right angles to the straight line *AB*.

Let a point *D* be taken at random on *AC*;

let *CE* be made equal to *CD*; [I. 3]

on *DE* let the equilateral triangle *FDE* be constructed, [I. 1]

and let *FC* be joined;

I say that the straight line *FC* has been drawn at right angles to the given straight line *AB* from *C* the given point on it.

For, since *DC* is equal to *CE*,

and *CF* is common,

the two sides *DC*, *CF* are equal to the two sides *EC*, *CF* respectively;

and the base *DF* is equal to the base *FE*;

therefore the angle *DCF* is equal to the angle *ECF*; [I. 8]

and they are adjacent angles.

But, when a straight line set up on a straight line makes the adjacent angles equal to one another, each of the equal angles is right; [Def. 10]

therefore each of the angles *DCF*, *FCE* is right.

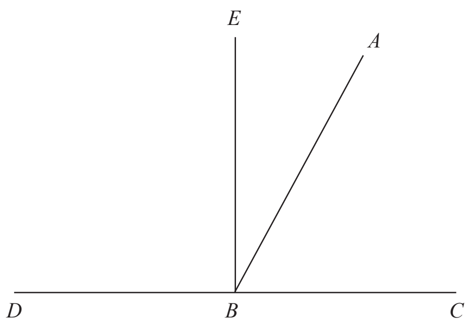
Therefore the straight line *CF* has been drawn at right angles to the given straight line *AB* from the given point *C* on it.

Q. E. F.

[. . .]

Proposition 13.

If a straight line set up on a straight line make angles, it will make either two right angles or angles equal to two right angles.



For let any straight line *AB* set up on the straight line *CD* make the angles *CBA*, *ABD*;

I say that the angles *CBA*, *ABD* are either two right angles or equal to two right angles.

Now, if the angle *CBA* is equal to the angle *ABD*,

they are two right angles. [Def. 10]

But, if not, let *BE* be drawn from the point *B* at right angles to *CD*; [I. 11]

therefore the angles *CBE*, *EBD* are two right angles.

Then, since the angle *CBE* is equal to the two angles *CBA*, *ABE*,

let the angle *EBD* be added to each;

therefore the angles *CBE*, *EBD* are equal to the three angles *CBA*, *ABE*, *EBD*. [C.N. 2]

Again, since the angle *DBA* is equal to the two angles *DBE*, *EBA*,

let the angle *ABC* be added to each;

therefore the angles *DBA*, *ABC* are equal to the three angles *DBE*, *EBA*, *ABC*. [C.N. 2]

But the angles *CBE*, *EBD* were also proved equal to the same three angles; and things which are equal to the same thing are also equal to one another; [C.N. 1]

therefore the angles *CBE*, *EBD* are also equal to the angles *DBA*, *ABC*.

But the angles *CBE*, *EBD* are two right angles;

therefore the angles *DBA*, *ABC* are also equal to two right angles.

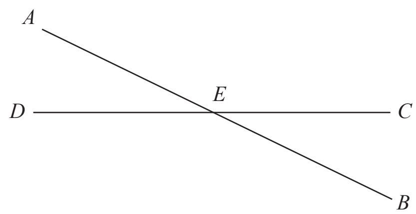
Therefore etc.

Q. E. D.

[. . .]

Proposition 15.

If two straight lines cut one another, they make the vertical angles equal to one another.



For let the straight lines *AB*, *CD* cut one another at the point *E*;
I say that the angle *AEC* is equal to the angle *DEB*,
and the angle *CEB* to the angle *AED*.

For, since the straight line *AE* stands on the straight line *CD*, making the angles *CEA*, *AED*,

the angles *CEA*, *AED* are equal to two right angles [I. 13]

Again, since the straight line *DE* stands on the straight line *AB*, making the angles *AED*, *DEB*,

the angles *AED*, *DEB* are equal to two right angles. [I. 13]

But the angles *CEA*, *AED* were also proved equal to two right angles;

therefore the angles *CEA*, *AED* are equal to the angles *AED*, *DEB*. [Post. 4 and C.N. 1]

Let the angle *AED* be subtracted from each;

therefore the remaining angle *CEA* is equal to the remaining angle *BED*. [C.N. 3]

Similarly it can be proved that the angles *CEB*, *DEA* are also equal.

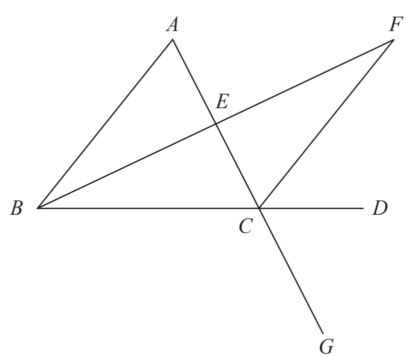
Therefore etc.

Q. E. D.

[. . .]

Proposition 16.

In any triangle, if one of the sides be produced, the exterior angle is greater than either of the interior and opposite angles.



Let *ABC* be a triangle, and let one side of it *BC* be produced to *D*;

I say that the exterior angle *ACD* is greater than either of the interior and opposite angles *CBA*, *BAC*.

Let *AC* be bisected at *E*, [I. 10]
and let *BE* be joined and produced in a straight line to *F*;

let *EF* be made equal to *BE*, [I. 3]
let *FC* be joined [Post. 1], and let *AC* be drawn through to *G*. [Post. 2]

Then, since *AE* is equal to *EC*, and *BE* to *EF*,
the two sides *AE*, *EB* are equal to the two sides *CE*, *EF* respectively;
and the angle *AEB* is equal to the angle *FEC*,
for they are vertical angles. [I. 15]
Therefore the base *AB* is equal to the base *FC*,
and the triangle *ABE* is equal to the triangle *CFE*,
and the remaining angles are equal to the remaining angles respectively, namely those which the equal sides subtend; [I. 4]
therefore the angle *BAE* is equal to the angle *ECF*.

But the angle *ECD* is greater than the angle *ECF*; [C.N. 5]
therefore the angle *ACD* is greater than the angle *BAE*.

Similarly also, if *BC* be bisected, the angle *BCG*, that is, the angle *ACD* [I. 15], can be proved greater than the angle *ABC* as well.

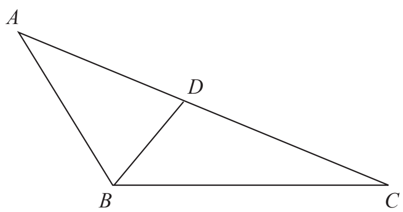
Therefore etc.

Q. E. D.

[. . .]

Proposition 18.

In any triangle the greater side subtends the greater angle.



For let *ABC* be a triangle having the side *AC* greater than *AB*;

I say that the angle *ABC* is also greater than the angle *BCA*.

For, since *AC* is greater than *AB*, let *AD* be made equal to *AB* [I. 3], and let *BD* be joined.

Then, since the angle *ADB* is an exterior angle of the triangle *BCD*,
it is greater than the interior and opposite angle *DCB*.

But the angle *ADB* is equal to the angle *ABD*, [I. 16]
since the side *AB* is equal to *AD*;

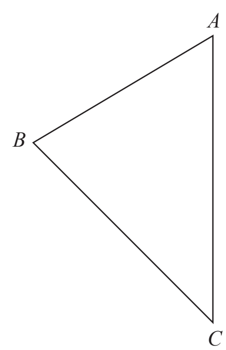
therefore the angle *ABD* is also greater than the angle *ACB*;
therefore the angle *ABC* is much greater than the angle *ACB*.

Therefore etc.

Q. E. D.

Proposition 19.

In any triangle the greater angle is subtended by the greater side.



Let *ABC* be a triangle having the angle *ABC* greater than the angle *BCA*;

I say that the side *AC* is also greater than the side *AB*.

For, if not, *AC* is either equal to *AB* or less.

Now *AC* is not equal to *AB*;
for then the angle *ABC* would also have been equal to the angle *ACB*; [I. 5]
but it is not;

therefore *AC* is not equal to *AB*.

Neither is *AC* less than *AB*,
for then the angle *ABC* would also have been less than the angle *ACB*; [I. 18]
but it is not;

therefore *AC* is not less than *AB*.

And it was proved that it is not equal either.

Therefore *AC* is greater than *AB*.

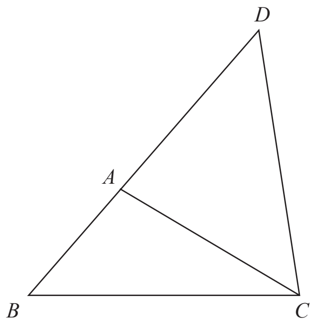
Therefore etc.

Q. E. D.

Proposition 20.

In any triangle two sides taken together in any manner are greater than the remaining one.

For let *ABC* be a triangle;
I say that in the triangle *ABC* two sides taken together in any manner are greater than the remaining one, namely



BA, *AC* greater than *BC*,
AB, *BC* greater than *AC*,
BC, *CA* greater than *AB*.

For let *BA* be drawn through to the point *D*, let *DA* be made equal to *CA*, and let *DC* be joined.

Then, since *DA* is equal to *AC*,
the angle *ADC* is also equal to the angle *ACD*; [I. 5]
therefore the angle *BCD* is greater than
the angle *ADC*. [C.N. 5]

And, since *DCB* is a triangle having the angle *BCD* greater than the angle *BDC*,

and the greater angle is subtended by the greater side, [I. 19]
therefore *DB* is greater than *BC*.

But *DA* is equal to *AC*;

therefore *BA*, *AC* are greater than *BC*.

Similarly we can prove that *AB*, *BC* are also greater than *CA*, and *BC*, *CA* than *AB*.

Therefore etc.

Q. E. D.