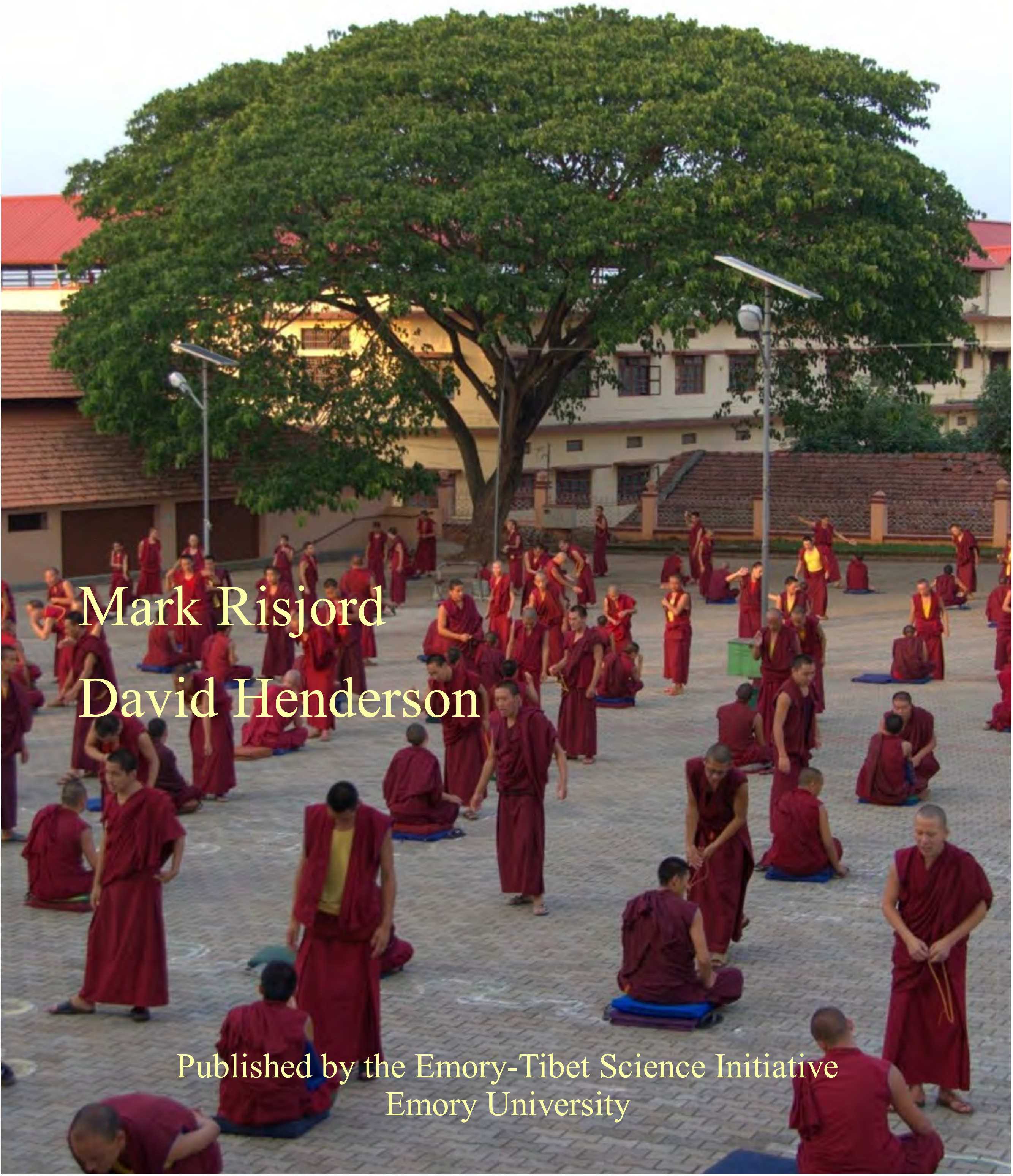
The Dharma of Science

Philosophy of Science for Buddhist Scholars



Mark Risjord
David Henderson

Published by the Emory-Tibet Science Initiative
Emory University

The Dharma of Science

Philosophy of Science for Buddhist Scholars

Mark Risjord
Emory University and the University of Hradec Králové

David Henderson
University of Nebraska, Lincoln

Published by the Emory-Tibet Science Initiative
Emory University, Atlanta Georgia, USA

Printed on November 24, 2019

THE DALAI LAMA

FOREWORD

Despite their obvious differences, science and Buddhism share several key features in common. Both are committed to empirical observation, the testing of hypotheses, avoiding blind adherence to dogma, and cultivating a spirit of openness and exploration. Most importantly, Buddhism and science share as a fundamental aim the contribution they can make to humanity's well-being. While science has developed a deep and sophisticated understanding of the material world, the Buddhist tradition has evolved a profound understanding of the inner world of the mind and emotions and ways to transform them. I have no doubt that improving collaboration, dialogue and shared research between these two traditions will help to foster a more enlightened, compassionate, and peaceful world.

I have long supported the introduction of a comprehensive science education into the curriculum of the traditional Tibetan monastic educational system. When I first heard that Emory University proposed to develop and implement such a science education program for Tibetan monks and nuns in collaboration with the Library of Tibetan Works and Archives, I thought it would take many years. When I visited Emory University in October 2007, I was genuinely surprised to be presented with the first edition of a science textbook for Tibetan monks and nuns, the result of more than a year's work by a team of dedicated scientists and translators at Emory.

By extending the opportunities for genuine dialogue between science and spirituality, and by training individuals well versed in both scientific and Buddhist traditions, the Emory-Tibet Science Initiative has the potential to be of great meaning and significance to the world at large. Once more, the creation of this primer series, presented in both Tibetan and English, is a clear tribute to the commitment and dedication of all those involved in this project. With the preparation having been done with such care, I am confident that the long-term prospects for this project are bright.

I congratulate my friend Dr. James Wagner, President of Emory University, the science faculty and translators of the Emory-Tibet Science Initiative, and everyone who has lent their support to this project for achieving so much in such a short time and offer you all my sincere thanks.

4 October 2010

EMORY

Office of the President

Education is one of the most potent tools we have for ensuring a better world for ourselves and for generations to come. To be truly effective, however, education must be used responsibly and in service to others. This ideal of an education that molds character as well as intellect is the vision on which Emory University was founded, and the challenges of our time show that the need for such education is as great as ever.

This vision is one that His Holiness the Dalai Lama shares deeply, and it is the reason for the close relationship that has emerged between His Holiness and Emory over the past two decades. On October 22, 2007, it was my pleasure and privilege to welcome His Holiness to Emory to be installed as Presidential Distinguished Professor and to join our community as a most distinguished member of our faculty.

The interdisciplinary and international nature of the Emory-Tibet Science Initiative, the most recent and ambitious project of the Emory-Tibet Partnership, is an example of Emory University's commitment to courageous leadership for positive transformation in the world. This far-reaching initiative seeks to effect a quiet revolution in education. By introducing comprehensive science instruction into the Tibetan monastic curriculum, it will lay a solid foundation for integrating insights of the Tibetan tradition with modern science and modern teaching, through genuine collaboration and mutual respect. The result, we trust, will be a more robust education of both heart and mind and a better life for coming generations.

The Emory-Tibet Partnership was established at Emory in 1998 to bring together the western and Tibetan traditions of knowledge for their cross-fertilization and the discovery of new knowledge for the benefit of humanity. This primer and its three companion primers are splendid examples of what can be accomplished by the interface of these two rich traditions. We at Emory University remain deeply committed to the Emory-Tibet Science Initiative and to our collaboration with His Holiness and Tibetan institutions of higher learning.

To the monastic students who will benefit from these books, I wish you great success in your studies and future endeavors.

James W. Wagner

James W. Wagner
President

Emory University
Atlanta, Georgia 30322
An equal opportunity, affirmative action university

PROLOGUE

It is most fulfilling to be able to bring out this primer as part of the complete series of core science textbooks along with reference materials for a six-year program in modern science developed for Tibetan monks and nuns. This collaborative undertaking between Emory University and the Library of Tibetan Works and Archives, which is both unprecedented and highly challenging, took birth under the aegis of His Holiness the Dalai Lama and is known as the Robert A. Paul Emory-Tibet Science Initiative (ETSI). To ensure the successful implementation of the ETSI, we have been working to create bilingual materials catered to the Tibetan monastic community and make them available in print forms.

It is a great honor for the two of us to play a part in overseeing this project. While each of us finds great inspiration for this project and the promise it holds, the full scope of its vision lies with His Holiness the Dalai Lama. For several decades, His Holiness has had the dream of introducing science education as a crucial component of the Tibetan monastic curriculum. While this is a bold step, His Holiness sees far-reaching benefits in such an undertaking. The integration of science into Tibetan monastic study will serve as a model and a trailblazer for constructive collaborations between religious and scientific traditions. It will help to inspire a paradigm shift in modern education as we know it, by providing resources for integrating the training of both heart and intellect to create a balanced education of the whole person. Furthermore, it will create a new science literature in the Tibetan language, thereby enriching the already extensive Tibetan literary tradition and helping to preserve the endangered Tibetan culture. This project represents a significant step towards a genuine convergence of science and spirituality. This convergence, which would enable us to tap into the combined resources of knowledge of the external world and knowledge of the inner world of the mind, could prove crucial for our future survival.

We are deeply honored, grateful, and humbled by the trust and confidence His Holiness has shown in us by entrusting us with this project, so dear to his heart. We thank him for his constant guidance, vision, and support at every step of the way. Furthermore, we thank all those who have made the Emory-Tibet Science Initiative possible. In particular, Dr. James W. Wagner, President of Emory University, has provided critical institutional support, without which none of this would be possible. Our role has simply been to oversee ETSI, but its actualization is due to many others, most notably the tireless and selfless ETSI faculty, our dedicated team of translators both at Emory University and at the Library of Tibetan Works and Archives, and the administrators and staff of Emory and LTWA, who have supported this ambitious undertaking in countless ways. Crucially, this project has depended upon generous financial support from Emory University, the Dalai Lama Trust, the Joni Winston Fund, the John Templeton Foundation, and a number of key donors: including the Judith McBean Foundation, the Lostand Foundation, Jaynn Kushner, and Drepung Loseling Monastery, Inc. Atlanta. To all these supporters, we would like to express humbly our deepest gratitude and thanks.

Geshe Lhakdor
Director, Library of Tibetan Works and Archives

Geshe Lobsang Tenzin Negi
Director, Emory-Tibet Science Initiative

ACKNOWLEDGEMENTS

The Robert A. Paul Emory-Tibet Science Initiative (ETSI) grew out of the longstanding vision of His Holiness the Dalai Lama and is sustained by His Holiness's continued guidance and support at every step of the way. Not only has His Holiness provided annual operational funds, but he has also provided $1 million towards the ETSI endowment fund thereby ensuring the long-term sustainability of the program. The ETSI also owes its existence to the patronage of Dr. James W. Wagner, President of Emory University, who has allocated considerable funding on behalf of Emory University and from his personal discretionary fund.

The Emory-Tibet Partnership (ETP) was established in 1998 in the presence of His Holiness the Dalai Lama through the collaborative vision and work of Dr. Robert Paul and Geshe Lobsang Tenzin Negi. ETSI is the most ambitious project to grow out of the Emory-Tibet Partnership, and in 2010 ETSI was renamed the Robert A. Paul Emory-Tibet Science Initiative in honor of Dr. Paul's visionary leadership and guidance. We express our heartfelt thanks to both these individuals for helping to establish the many programs of the Emory-Tibet Partnership, including ETSI.

We gratefully acknowledge Geshe Lhakdor, Director of the Library of Tibetan Works and Archives, Dharamsala, India, and Dr. Preetha Ram, former Associate Dean of Science Education at Emory University, both of whose leadership has been invaluable to the establishment and development of this initiative.

The project would also not have been possible without the support of Dr. Gary Hauk, Vice President and Deputy to the President at Emory University, who has guided ETP from the beginning and continues to be one of ETSI's strongest supporters. Additionally, ETSI is greatly indebted to Dr. Robin Forman, Dean of Emory College of Arts and Sciences, for providing critical resources and faculty of Emory College, which houses this initiative, towards the ongoing development and implementation of the ETSI.

We thank also the ETSI science faculty, who have worked tirelessly to develop the science textbooks and who have traveled to India each summer to teach the science intensives, and the ETSI science translators who have given of their skills and time to contribute an entirely new scientific vocabulary to the Tibetan literary tradition and lexicon. In particular, Drs Carol Worthman, Arri Eisen, and John Malko, team leaders for neuroscience, biology, and physics, respectively, oversee all of the curricular aspects of the ETSI and have been integral to any success experienced by the ETSI. Additionally, the principal ETSI translators, Geshe Dadul Namgyal and Tsondue Samphel oversee the entire translation and production of all ETSI printed materials and video lectures. Without this dedicated team of exceptional faculty and translators, the ETSI would not be where it is today.

We also thank the hard-working staff of the Emory-Tibet Partnership, who have labored far beyond the call of duty, showing time and again that their efforts are not only work, but also an act of love.

We thank all those who have contributed the financial support needed to operate ETSI and ensure its long-term sustainability. We are particularly indebted to Joni Winston for her long-term generous support to ETSI and for her unwavering conviction in the worth of this endeavor. Funding for ETSI has also come from Emory University and Emory College, including the Science and Society Program and the Office of International Affairs.

Generous support has also come from:

- The Dalai Lama Trust
- The Joni Winston Fund
- The John Templeton Foundation
- Judith McBean Foundation
- Lostand Foundation
- Jaynn Kushner
- Drepung Loseling Monastery, Inc., Atlanta, Georgia

We also thank the Advisory Board of the Emory-Tibet Science Initiative for their guidance and advice:

- Dr. Gary Hauk, Vice President and Deputy to the President, Emory University
- Geshe Thupten Jinpa, Principal English Language Translator for H.H. the Dalai Lama and President, Institute of Tibetan Classics
- Geshe Lhakdor, Director, Library of Tibetan Works and Archives
- Dr. Alan Wallace, President, Santa Barbara Institute
- Dr. Preetha Ram, former Associate Dean for Pre-Health and Science Education, Emory University
- Dr. Arthur Zajonc, President, Mind and Life Institute
- Dr. Richard Davidson, Director, Center for Investigating Healthy Minds, University of Wisconsin, Madison
- Dr. Robert A. Paul, Charles Howard Candler Professor of Anthropology and Interdisciplinary, Emory University
- Geshe Lobsang Tenzin Negi, Director of Emory-Tibet Partnership, Emory University

We would like to thank the venerable abbots and the administration of the Tibetan monastic institutions for embracing the ETSI curriculum and incorporating this material into the Tibetan monastic core curriculum. Lastly, we thank the highly dedicated monastic students of the Emory-Tibet Science Initiative, who are not only beneficiaries, but also essential collaborators in the success of this program. May the knowledge that they gain through this program and these materials benefit them greatly, and through them, all of humankind.

Contents

Chapter 1

Introduction: Buddhism and Western Philosophy of Science

At the urging of His Holiness the 14^{th} Dalai Lama, Buddhist scholars have begun seriously engaging contemporary science. Learning science poses two kinds of philosophical challenge for Buddhist scholars. First, they need to understand not only *what* contemporary science is saying, but *why*. Understanding why permits Buddhist scholars to both properly follow scientific research and critically interpret scientific results. The second challenge is to understand contemporary science philosophically, coming to grips with both its philosophical presuppositions and its philosophical consequences. This book aims to help Buddhist scholars address both challenges.

To critically engage any author, school, or intellectual tradition, one must understand the character of the reasoning: what counts as grounds for accepting a thesis, what counts as reasons for rejecting it. Similarly, understanding science requires understanding scientific reasoning. The pramana discipline of Buddhism is home to a robust and nuanced understanding of logic and epistemology. This is both a advantage and a disadvantage when beginning to learn science. Contemporary science has assimilated many ideas from the Western philosophical tradition. And science has often led philosophy, making advances in methodology or metaphysics that are later subject to philosophical scrutiny. This intertwining of philosophy and science in the western tradition means that the difference between Buddhist and Western epistemology can be a barrier to critical understanding of science. To critically engage contemporary science, Buddhist scholars need to understand the ways in which scientific reasoning is similar to their own logic and epistemology as well as those ways in which it is different.

As a first step toward understanding contemporary science, Buddhist scholars need an outline of the main features of the logical, epistemological, and metaphysical commitments of science. This includes the distinctive characteristics of scientific methodology as well as central concepts of scientific inquiry, such as observation, measurement, correlation, and cause. The first half of this book provides such an introduction to scientific reasoning. The aims are to provide an understanding of scientific reasoning to facilitate both understanding and criticism of scientific results.

Now, we must pause to note that there is no single, univocal "scientific" concept of cause or observation. And there is no single, agreed upon canon of scientific reasoning. Science is a rich and complex enterprise. Science proceeds in many ways and there are many interpretations of its results. Recognizing the complexity, the material in this book tries to present a consensus view. It seeks to present key features of science in ways that most scientists and philosophers of science would agree with. It is intended to provide a basis for understanding scientific explanations, initiating scientific investigations, and critically engaging scientific results. The later chapters of this book begin—but only begin!—a deeper, more philosophical engagement with science. Here the disputes are deeper, and there is less consensus. We will discuss different ways to conceptualize observation and causality. We will ask whether science aims at a correct account of "reality," what role of ethics plays in science, and whether there are limits to scientific inquiry. As Buddhist scholars well know, understanding such deeper issues requires that the student appreciate different perspectives. While the chapters will not try to hide the views preferred by the authors, the later chapters will exhibit some of the debate Western philosophers have had about these issues.

Many of the arguments in this text are directly relevant to traditional Buddhist concerns: *e.g.*, whether inductive arguments are sufficient grounds for knowledge, whether causality is a substantive (or real) relationship, or whether science provides knowledge of conventional or ultimate reality. While the arguments discussed in the latter part of this book may not be ones that Buddhist scholars ultimately want to accept, they are offered here as an impetus to further reflection. Throughout the history of philosophy in Europe, science has been a source of inspiration and motivation to refine metaphysical and epistemological theories. This book is written in the hope that science will provide the same stimulus for Buddhist philosophy.

The material in this book was developed by a talented group of philosophers over many years. In 2006 and 2007, Stephen (Pii) Dominick and Mark Risjord piloted an early version of the material in this text for a select group of monks and nuns. The feedback provided by these Buddhist scholars, as well as their instruction in Buddhist logic and epistemology, were invaluable as the project developed. When the formal program began at Drepung, Sera, and Gaden monasteries in 2014, Risjord assembled a team composed of David Henderson, Julie Haas, Jessica Locke, David Peña-Guzman, and Karsten Steuber. This group developed and taught the first full course of lectures. We were joined in subsequent years by Andrew Culbreth, Luke Elwonger, Ben Henke, Hamner Hill, Karl Schmid, and Preston Stovall, all of whom contributed to revisions of the program and ultimately to the shape of this text. While teaching with David Henderson in 2016 and 2017, Mark Risjord began writing down the content of the lectures in textbook form.

Mark Risjord's work on this book during the 2018–2019 academic year was supported by the project "International mobilities for research activities of the University of Hradec Králové", CZ.02.2.69/0.0/0.0/16_027/0008487."

The project would have been impossible but for the superb efforts of the translators and science program administrators at Drepung, Sera, and Gaden, as well as those at ETSI in Atlanta. Tsondue Samphel and Geshe Dadul Namgyal deserve special mention. Their patient explanations of the subtle differences between Buddhist and Western philosophical ideas were crucial to our understanding. Overall, the topics covered and the approach to them in this book is the product of what we learned from our monastic interlocutors: from the monks and nuns, the translators, and the science coordinators. We have been honored to learn from and to teach these Buddhist scholars.

Chapter 2

What is Science?

2.1 Characteristics of Scientific Inquiry

Western science, especially in the last two to three hundred years, has developed as a powerful form of knowledge production. What are the distinctive characteristics of scientific inquiry? In this chapter, we will argue that for the following conception of scientific inquiry.

Scientific inquiry: Science is a process with three distinctive features:

1. Scientific inquiry uses observation to systematically test theory.
2. Scientific inquiry is self-correcting.
3. Scientific inquiry begins with a problem or question, and devises theory to answer the question.

The ideas emphasized in the first feature—"systematic" and "observation"—are both philosophically important. Indeed, this entire book is an attempt to unpack these two ideas. As a first pass, we can say that scientific theories are systematic insofar as they permit specific predictions. Such predictions are observable in the sense that we can observe the conditions that make them true or false. We can determine their truth value by observation. The two examples of this chapter illustrate different ways in which scientists have used observation to systematically test theory. Scientific theories are also systematic in the sense that different theories fit together. As we will see in the examples below, astronomical theories were supported by theories of motion (physics). The systematic character of science thus supports the distinctive epistemology of science.

The second feature, self correction, involves the ongoing search for sources of error. Testing a theory by making predictions, as we have just discussed, is one way to look for error. We test to determine whether our theories are correct. If they are incorrect, scientists replace their theories. As new scientific theories replace older theories, the limitations and mistakes of older theories are often corrected. While testing is important, there are several other ways in which science is self correcting too. We test theories with observation, but an observation may be erroneous. One way to detect the error in an observation is make other observations of a different kind. Scientific observation often involves instruments, such as telescopes or x-ray machines. Observations from different sources can be compared, and erroneous observations identified and corrected. (We will discuss the character of scientific observation in Chapter 8.) Of course, the instruments or process of making the observation might be flawed as well. For this reason, scientific methods—ways of organizing or structuring the investigation—must be closely examined to determine whether they are reliable. The self-correcting character of science involves scientists learning how to learn, and passing these lessons on in the training of new scientists.

The capacity to be self-correcting requires scientists to have particular personal attitudes. A scientist needs to work hard to develop a theory, but at the same time, a scientist must always look for ways in which the theory might be wrong. This requires an attitude of humility, both epistemically and personally. Epistemic humility is the recognition that it is always possible to be wrong. This means that, personally, one cannot be too attached to one's theory. One must have the humility to bow before the stronger argument. Changing one's mind because of new evidence is not a sign of weakness, but of intellectual strength.

In addition to personal attributes, the self-correcting character of science requires that scientists as a group should be organized in a particular way. It is not enough for scientists to look for flaws in their own theories. Errors are difficult to find, and we may have beliefs or presuppositions that make errors difficult to see. For this reason, debate among different scientists is a crucial element of the self correcting character of science. Scientists should not think of themselves as independent thinkers who invent theories. A good scientist must think of him or herself as a part of a community of scientists. And a good community of scientists is one that freely exchanges and debates ideas. This means that a scientific community needs institutions that support the open criticism of ideas. No theory should be held immune from criticism because it is accepted by powerful or influential persons.

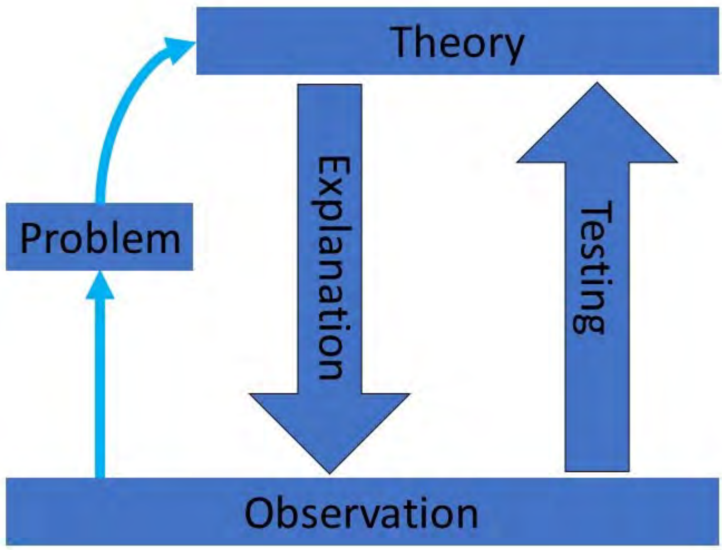


Figure 2.1: The Process of Scientific Inquiry[1]

When we learn science, especially at the beginning of our studies, there is an emphasis on theories. It is easy to lose sight of the fact that theories come from somewhere. Theories are a phase in a larger process of scientific inquiry; this is the import of the third feature of scientific inquiry. As illustrated in Figure 2.1, scientists develop theories in response to some kind of *problem*. These problems might be very practical, such as the challenge of growing crops. They might be a response to a surprising phenomenon, such as the eclipse of the sun. In general[2] scientific problems arise from something we observe or experience. Scientific theories propose ways to resolve the problem or answer the question. They do so by explaining why we make the observations we do, or by permitting us to predict what will happen. Once a theory has been proposed, it must be subject to the systematic testing by observation we have already discussed.

The distinctive features of science cannot be understood in the abstract. Science is a diverse and multifaceted enterprise. Systematic testing by observation, self-correction, and the process of inquiry will thus appear differently in the various fields. In this chapter we will discuss two examples, one taken from astronomy and another taken from biology. Our goal here is not to explain all of the science to you, so do not get tangled in the scientific details—these are the subject of your science classes. Rather, these examples will begin to show how criteria (1)–(3), above, make a difference to the practice of good science. They will therefore illustrate what is distinctive about scientific knowledge.

[1] Image by the authors

[2] In addition, some scientific problems or questions arise from existing theories. Einstein's theory, for instance, responded to problems raised by the existing theory of how light moved through space. To keep matters simple, we will set this kind of case aside for now.

2.2 Astronomy: A Story of Revolutionary Change

2.2.1 Ptolemy and Ancient Greek Astronomy

Most ancient societies studied the motions of the stars, the sun and the moon, and the planets. These observations were important for predicting seasonal activities, such as when to plant and harvest, when to hunt or fish, and when to conduct religious rituals. Ancient peoples around the world developed calendars that recorded how the annual change in seasons correlated with the position of the sunrise and the phases of the moon. Calendars describe observations in a *systematic* way: they cover then entire annual cycle and strive for accuracy. By being systematic in this way, they permit the prediction of regular annual changes, such as when the rains would come. While the creation of calendars is a form of systematic observation, it is only the beginning of the process of scientific inquiry. The ancient Chinese, Indian, and Greek civilizations began to do science when they asked *why* the sun, moon, stars, and planets moved as they did.

The sun, moon, and the stars all exhibit regular motions. Each day, they rise in the east, move in an arc across the sky, and they set in the west. On a series of evenings, the position of the sunset and moonrise changes slowly, moving toward the north in the summer and then toward the south in the winter. The moon has a slightly different position against the background of stars each night. The simplest theory is that the sun, moon, and stars move in circles around the earth.

The planets (Mercury, Venus, Mars, Saturn, and Jupiter were known to ancient peoples) are distinctive among bodies visible at night because they do *not* exhibit these regular motions. The planets move in a particular way. Of course, on a nightly basis, these objects arise and pass overhead just as the stars do. However, each successive night one finds these planets to have moved to slightly different positions relative to the background of stars. Over the course of several days, a planet—say, Mars—will appear slightly farther to the west relative to the stars. Then it will seem to stop and reverse course. Now it moves to the east, relative to the stars, over the course of several weeks. Then it stops and reverses course again, continuing to the west. Compared to the sun and the moon, this difference is quite striking. This retrograde motion is illustrated in Figure 2.2.

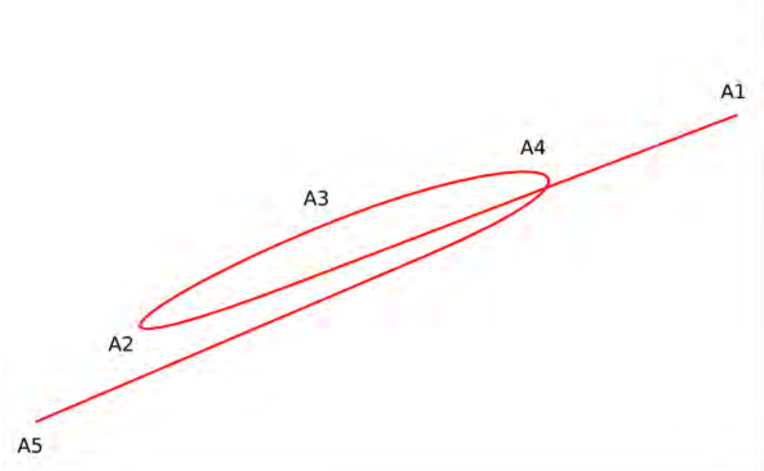


Figure 2.2: Retrograde motion of Mars[3]

The Greek astronomer Ptolemy (100-170 CE) devised a theory that explained the retrograde motion of Mars and the other planets. According to this theory, the sun, the moon, and all of the planets moved in circular orbits around the earth. The planets were different from the sun and moon insofar as their motion was understood as being determined by a circle on a circle, like a wheel spinning on the edge of a disk. These additional circle was called an "epicycle." On Ptolemy's theory, the planets actually moved in a loop, just as we see in the sky.

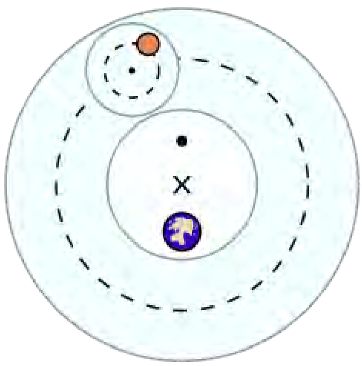


Figure 2.3: Epicycle[4]

Ptolemy's theory permitted him to predict where the planets would be in the night sky in the future. During Ptolemy's time, Greek mathematics had developed a very powerful account of geometry. Ptolemy used geometry to construct an account of the orbits of the planets around the earth, one that understood them as

[3]Image in the public domain. Source: Wikimedia Commons. Notations A1 through A5 correspond to positions in Figure 2.4

[4]Image in the public domain. Source: Wikimedia Commons

moving along epicycles. This theory afforded significant success in accounting for the apparent position of the planets. However, the predictive power of Ptolemy's theory was limited. Predictions in the near future were accurate, but predictions about where planets would be several years from now were not. To make his theory more accurate, Ptolemy had to suppose that the earth was not exactly at the center of the orbit. The orbit was slightly offset. This is illustrated in Figure 2.3 by the small $\times$ in the center of the figure (the blue circle represents the earth).

Ptolemy's theory exhibits several of the features we have proposed as being characteristic of science. First, it illustrates the scientific process. Ptolemy's theory is developed as an answer to several questions: why do the sun, moon, planets, and stars move in the way they do? And why do the planets have retrograde motions, while the sun, moon, and stars do not? Ptolemy could see the position of these points of light on different nights, and notice that their relative positions changed. To answer the questions, Ptolemy had to propose that the planets move in ways that were not directly observable. In particular, the planets move along epicycles on circular paths, where the center of the circle is offset. The circular paths and the epicycles are postulated by the theory. The theory *explains* the observations; it explains why we see what we do when we look into the night sky. In addition, the observations *test* the theory. The theory successfully predicts where the planets will be observed in the future.

The general study of how and why objects move is a branch of physics known as "mechanics." Because Ptolemy was proposing that the planets moved in a particular way, his theory had to be consistent with the science of mechanics. And so it was. Greek physics proposed that motions in the heavens and motions on earth followed different principles. Heavenly motion had a kind of perfection, and circles were perfect. The natural motion of stars and planets was to move in circles. Ptolemy's theory thus followed the dictates of Greek physics.

2.2.2 The Copernican Revolution

The Ptolemaic astronomical theory was the dominant view among scholars in the middle east and Europe for more than 1000 years. There were alternative theories, even some proposing that the earth orbited around the sun. None became widely accepted, partly because of the difficulty of explaining retrograde motion. Copernicus (1473-1543 CE) proposed an alternative to Ptolemy's theory that explained retrograde motion.

According to Copernicus, all of the planets, including the earth (but except the moon) orbit the sun. Retrograde motion was explained as a kind of illusion. The illusion is similar to an illusion familiar from riding in a car or train. When one car in which I am riding passes another, the other car seems to move backwards. This is an illusion because both cars are really moving forward. The illusion is created by my car moving faster than the other car. Similarly, when planets "pass" each other, the other planet seems to stop, move backwards, and then continue.

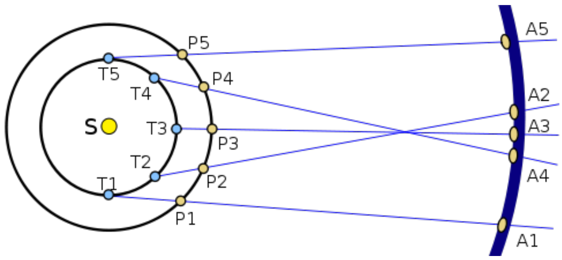


Figure 2.4: Copernican Explanation of Retrograde Motion[5]

Copenicus' theory was very different from Ptolemy's theory. Copernicus put the sun at the center of the solar system, and he postulated that the earth moved. Interestingly, both Copernicus's and Ptolemy's theories were supported by observation of the motion of the planets. They made nearly identical predictions about where the planets would be seen in the night sky, and therefore they were equally well supported by observation. This leaves us with an interesting philosophical problem: how do we choose between theories when they make the same predictions?

In the case of astronomy, the problem of choosing the best theory was facilitated by the *systematic* character of science. Both Copernicus' and Ptolemy's theories provided a detailed account of why the planets moved. The difference between their theories was used by scientists to find a prediction that would show which theory was superior. The difference lay in the way that the two theories treated the planet Venus.

Venus is always observed to be close to the sun. That is, Venus is only viewed in the early morning and early evening. While the other planets will be seen in the middle of the night, when the sun is on the other side of the Earth, Venus is never seen in this position For this reason, both Ptolemy and Copernicus theorized that Venus' orbit was close to the sun.

On Ptolemy's theory, Venus orbited the Earth like the Sun, moon, and other planets. However, since Venus is never seen on the opposite side of the Earth from the sun, its orbit had to stay in line with the sun. Figure 2.5 shows how the sun, Venus and Mercury all orbited together. This means that the full face of Venus should never be visible. Unlike the moon, Venus should never appear "full," but only as a crescent.

[5]Image used under the terms of the GNU Free Documentation License. Source: Wikimedia Commons.

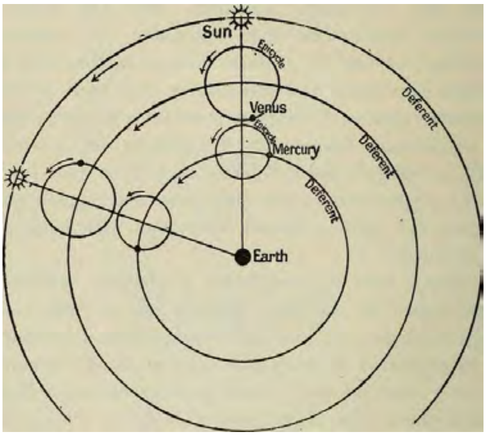


Figure 2.5: Epicycles of Venus and Mercury Stay in Line Between the Earth and the Sun[6]

According to Copernicus, by contrast, Venus orbited the sun. Sometimes, the Earth will be on one side of the Sun and Venus on the other. When this happens, Venus should appear "full" in the sense that it appears as a full disk. This, then, is a difference in the two theories that entails different observations. If Ptolemy is correct, then Venus will never be seen as a full disk; if Copernicus is right, then it will, as illustrated by Figures 2.6 and 2.7

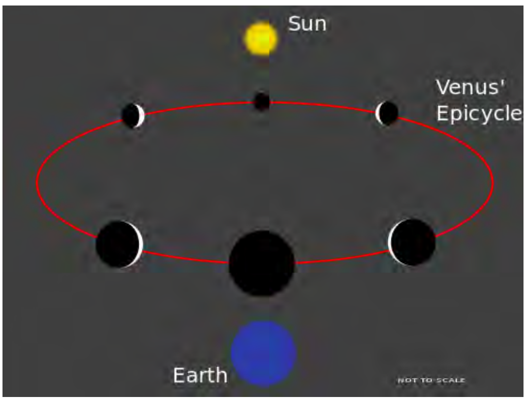


Figure 2.6: Phases of Venus as Predicted by Ptolemy[7]

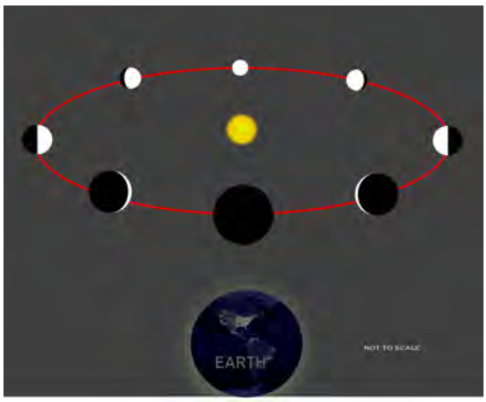


Figure 2.7: Phases of Venus as Predicted by Copernicus[8]

While the two theories made different predictions, the difference is very difficult to observe. Indeed, the difference between the crescent and the full face of Venus is not visible with the naked eye. For this reason, the predictions of Ptolemy's theory could not be tested until the invention of the telescope. Galileo

[6]Evershed, M. A. *Dante and the Early Astronomers*. Gall & Inglis, 1913, p 177. Image in the public domain, obtained from Internet Archive Book Images.

[7]Image used under the Creative Commons CC0 1.0 Universal Public Domain Dedication. Source: Wikimedia Commons

[8]Image in the public domain. Source: Wikimedia Commons

(1564-1642 CE), who was one of the first scientists to use the telescope for scientific observations, observed the phases of Venus, and thereby confirmed Copernicus's theory.

In spite of this success, Copernicus' theory had defects. Like Ptolemy's theory, its predictive accuracy was limited. It could predict the position of the planets in the near future, but longer range predictions were inaccurate. Copernicus, like Ptolemy, proposed that orbits were circular. The astronomer Johannes Kepler (1571-1630 CE) had the idea to propose that planets moved in elliptical orbits, not circular, as illustrated in Figure 2.8.

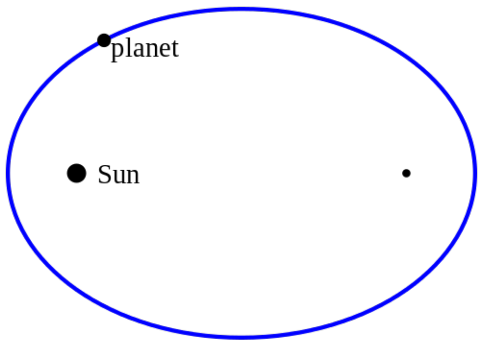


Figure 2.8: Kepler hypothesized that planets moved in eliptical orbits[9]

Kepler's modification of Copernicus's theory made much more accurate predictions about the position of planets in the night sky. We noted earlier that for Greek physics, celestial motions had to be circular. Such perfect, unchanging, circular motion had been thought to be the only form of motion fitting to the heavens, which had been thought to be distinctly different from motion on earth. Kepler's hypothesis thus broke with Greek physics, and it thus demanded a different kind of mechanics to explain planetary motion.

The new theory of motion was provided by Issac Newton (1642-1746 CE). Newton proposed that all motion was governed by three laws.

First Law: Bodies at rest tend to stay at rest, and bodies in motion tend to stay in motion (inertia).

Second Law: $F = ma$ (The force on an object equals its mass times acceleration)

Third Law: For every action there is an equal and opposite reaction

Newton's first law was an important change from Greek physics. While natural motions (in the heavens) was circular for the Greeks, for Newton natural motion is in a straight line. This is the importance of the First Law. When objects travel along a curved path, some force must be acting on them. The planets follow curved

[9] Image used under Creative Commons 3.0 license. Source: Wikipedia Commons.

paths, since they move in elliptical orbits. Newton postulated that the force of gravity was acting on the planets and causing them to follow elliptical orbits. Newton showed that the elliptical orbits of the planets could be mathematically derived from his three laws. Copernicus' idea that the sun is at the center of the solar system was finally made consistent with a physical account of motion.

Notice how this episode in the history of Western science illustrates the distinctive features of science identified in Section 2.1. The systematic character of the theories lies in the detailed ways in which they show how the planets move. This systematic character permits specific predictions about where the planets will be observed in the future, as well as other possible observations, such as the possibility of seeing Venus as "full." The success of the predictions made by Copernicus's theory, as developed by Kepler, was the reason scientists came to prefer it over the Ptolemaic theory.

The replacement of Ptolemy's theory by Copernicus's (and later Kepler's) theory also illustrates the self-correcting character of science. Ptolemy's theory was accepted and used by scientists for over 1000 years. It was very successful at explanation and prediction. Nonetheless, scientists continued to search for alternative explanations. Even before Copernicus, astronomers in India and the Middle East looked for errors in Ptolemy's theory and tried to develop better accounts. Ptolemy's theory was ultimately replaced when scientists could identify specific errors in the theory and when they had an alternative that could do better.

Astronomy and physics are an important part of scientific inquiry, but not all parts of science look like astronomy and physics. To properly see the characteristics of science, we should look at some other kinds of scientific research as well.

2.3 Biology: Curing Disease

When people began to take very long ocean voyages—voyages that were several weeks, even months long—new kinds of diseases appeared. "Scurvy" is one such disease. The symptoms included fatigue, sores on the skin, and loss of teeth. Many sailors died. This was an important problem, and many remedies were proposed. The scientific problem was to determine which of these possible remedies work. This is a general problem in medicine: how do we know whether our treatments really cure disease? James Lind (1761-1794 CE) answered this question in a way that has become a common method: he did an experiment. Lind was not the first to do experiments. Ancient Greek scientists did experiments too, and scientists before Lind's era, such as Frances Bacon (1562-1626 CE) and Galileo Galilei (1564-1642 CE) had developed an experimental approach to science. Lind added some distinctive elements to the work of his predecessors. Lind's methods are very similar to those used by scientists today.

On a particular voyage, he divided sailors who came down with scurvy into several groups. He gave each of the groups a different treatment. Some were given vinegar, others cider, still others were given oranges and limes. One group was simply given some sea water to drink. Lind chose these treatments because he theorized that scurvy was caused by a process in the body that was similar to the way food rots and decomposes. Such a process should be slowed or prevented in the same way fresh food is prevented from spoiling, such as by pickling with vinegar and salt. What made Lind's experimental method modern is that he was looking for a difference between those sailors who were treated in a particular way and those who were left alone. The sailors all had the same food and lived on the same ship. If one of they remedies works, we would expect sailors who took that remedy to get better when the untreated sailors did not.

Those sailors who were given oranges and lime juice recovered quickly and dramatically. This was evidence, but only partial evidence, for Lind's theory. He had theorized that acid or salt would prevent scurvy, but salt and some kinds of acid (like vinegar) had no better effect than nothing at all. Citrus fruits, on the other hand, had a strong effect. Lind's work led to the provision of citrus fruits to sailors, but it was not until the 20th century that the specific value of juice from oranges, limes, and lemons was understood. We now know that the juice from these fruits contains vitamin C.

Lind's research fits the picture of scientific process we sketched in Section 2.1. The problem of sailors suffering from scurvy is observable, and it was a problem that needed to be solved, if possible. The difference between Lind's work and Ptolemy's, however, is that Ptolemy had an elaborate and mathematically based theory, while Lind's was less detailed and precise. Where the astronomers' theories permitted precise predictions, Lind's theory was a general relationship between acid and scurvy. Lind's theory also had a different relationship between theory and observation than astronomy. While astronomers made predictions, then looked to the sky to see whether the predictions were confirmed or refuted, Lind conducted an experiment. He actively created a situation where his prediction could be tested.

Lind's experiment gives us another way to understand the idea that scientific theories are systematically tested by observation. As we have already mentioned, Lind's theory was not systematic in the way that Ptolemy's or Copernicus' theories were systematic. The astronomers and physicists sought a single theory that explained a wide range of phenomena. Lind had a much simpler theoretical idea. The systematicity and sophistication of Lind's work was in his method. While we will discuss the details and the epistemology of his experiment in Chapter 6, his experimental method was specifically arranged to eliminate the possibility of error. Lind's experiments thus also illustrate one of the ways in which science is self-correcting. Lind sought out possible sources of error and tried to eliminate them.

2.4 Philosophical Issues in Scientific Inquiry

The goal of this chapter has been to characterize scientific inquiry. Section 2.1 suggested three features that are distinctive of science. We have seen how these features are implicit in two episodes from the history of science. The systematic use of observation and the attempt to be self-correcting have, arguably, been responsible for the astounding success of recent science in understanding the natural, social, and psychological world.

The three features of science we have identified are important, but they remain vague and ill defined. Exactly how does scientific inquiry seek to minimize error? How does science learn from its mistakes? What makes some experiments good observational evidence for a theory while others are not? To answer these questions, we need to turn to the idea of good inference in science, and this is the subject of the next three chapters.

Chapter 3

Deductive Arguments and Theory Testing

3.1 Scientific Reasoning

In science, as in Buddhist thought, all knowledge has two sources: observation and inference. Both traditions have sophisticated methods for developing knowledge from these sources. There are important differences. Scientific observation often involves measurement and can be mediated by instruments. Scientific argumentation is supported by powerful mathematics. The presuppositions of these methods are somewhat different than the presuppositions of Buddhist epistemology. In some cases, scientific methods extend beyond the commitments of Buddhist epistemology, supplementing them with new techniques. In other cases, scientific methods for observation and inference may conflict with the tenets of Buddhist epistemology as it has been traditionally understood. It is therefore essential for a Buddhist student of science to be aware of the criteria that Western scientists use to judge the acceptability of inference and the reliability of observation.

The next several chapters will present a consensus view of the epistemology that is implicit in contemporary scientific practice. Chapters 3, 4, 5, 6, and 7 will work through contemporary western approaches to logic and scientific reasoning. Chapters 8 and 9 will focus on scientific observation and its philosophical consequences. The epistemology of science has been subject to much study in philosophy, and as with any good philosophical tradition, there are debates. This presentation will try to avoid taking sides in such debates as much as is possible. We will try to characterize scientific reasoning in a way that most scientists and philosophers would accept. Ultimately, the philosophical debates generated by this text should be Buddhist debates, not Western debates. Our goal is to put Buddhist scholars in a position to understand the reasoning put forward by scientists, and thereby learn the science needed to debate about its philosophical consequences. Of course, in order to attain an interpretive understanding, Buddhist scholars will need to explore the ideas presented here, turning them over and debating them so as to attain a mature grasp of them. This is a task for which their training will have well prepared them.

Fundamental to any study of reasoning or logic is the idea of support. Suppose I have the thought that "fire is hot." This provides a reason for the further thought that "fire will burn me." The first thought supports the second in the sense that the thought "fire will burn me" is accepted, believed, or otherwise taken to be true *because* "fire is hot" is taken to be true. Not just any association of thoughts is the kind of support that we study in logic. For example, suppose I have the thought that "A dog is in the road now," and this reminds me "A dog was in the road yesterday." This is a case of being prompted to remember, not an example of reasoning.[1] The kind of support with which we are concerned in logic thus involves truth and falsity: I take "fire will burn me" to be true because I believe that "fire is hot." Let us call the supporting thought (fire is hot) the **premise** and the supported thought (fire will burn me) the **conclusion**. Adopting the convention that a line separates the premises and the conclusion, we will represent the support relation as shown in the box labeled Argument 3.1.

Fire is hot

Fire will burn me

Argument 3.1

The relationship between premises and conclusions can be assessed or evaluated, and logic aims to articulate criteria for determining whether premises provide good or bad reasons for a conclusion. In the example above, the premise provides some support for the conclusion. Western philosophers would say that the conclusion would be more strongly supported by the addition of another premise. Argument 3.2 represents this support relation.

Fire is hot
All hot things burn me

Fire will burn me

Argument 3.2

The support relationship represented in Argument 3.2 is stronger than Argument 3.1, and both examples are stronger than Argument 3.3, below. While one might take the two premises of Argument 3.3 to be reasons for accepting the con-

[1]Of course, a person *could* treat this as a support relation. But since "A dog is in the road now" provides little or no reason to accept "a dog was in the road yesterday," it would be extremely weak. The point of the example is that the mind does many things, and not all movements of thought are cases of reasoning.

clusion, logicians in both the Buddhist and Western traditions would say that the conclusion is very weakly supported. The truth of these premises does not give us reason to think the conclusion is true. The study of logic in both Buddhist and Western philosophy has aimed to understand why some support relationships are strong and others are weak. Logic is therefore normative in the sense that it studies how humans ought to reason. It tells us what we should do if we want to reason *well*.

Fire is hot
Fire is yellow

All yellow things are hot

Argument 3.3

Premises and conclusions can be conceptualized in at least two ways. In the examples above, premises and conclusions were called "thoughts" or "beliefs." The relationship of support is treated as a movement of thought. When I come to believe that "fire will burn me" because I accepted the truth of "fire is hot," I have *inferred* that fire will burn me. When premises and conclusions are taken to be beliefs or thoughts, logic is the normative study of the psychological phenomenon of **inference**.

A different way to understand premises and conclusions is to take them to be public, linguistic objects. Premises and conclusions are statements expressed in the sentences of a language. Let us use the word **argument** for one or more statements (premises) whose truth supports the truth of another (the conclusion).

When conceptualized in this way, the relationship of support is objective in the sense that it is independent of what anyone believes. Even if I do not believe that a premise is true, I can recognize that if the statement *were* true, it would strongly (or weakly) support the truth of the conclusion. Because we can all talk about the same sentences, arguments are public objects for contemplation, rather than private matters of individual thought.

Most importantly, thinking of support relationships in terms of arguments lets us evaluate them independently of belief in the premises. This point bears emphasis. In the Western tradition of logic, it is possible for an argument to be a strong one, even if the premises are not believed. This feature of arguments lets us reason hypothetically in the sense that we may suppose a premise to be true, and see what follows from it. The concepts of validity (discussed in Section 3.2) and falsification (discussed in Section 3.4) depends on our ability to evaluate support relationships independently of belief in the premises.

Arguments and inferences are closely related. In both cases the premises and conclusion can be true or false. In both cases logic concerns evaluating the relationship of support as good or bad, strong or weak. A difference between them is that we take our beliefs to be true, but not all sentences are true. So in an inference, we take ourselves to be moving from true premises to true conclusions. By contrast, since we can evaluate an argument without believing the premises, we can recognize that some strong arguments have false premises. As we will see later in this chapter, there are some important uses of arguments where the premises either have an unknown truth value or we suspect they are false.

A further relationship between arguments and inferences is that we use arguments to express inferences. By formulating our beliefs in language, we can make them available to others. My inferences thus become available to others to either accept for themselves or to criticize. Arguments, in turn, have an influence on inference. If I believe that the premises of an argument are true, and if I take the argument to be a good one, then I ought to form a belief that corresponds to the conclusion, that is, I should make the corresponding inference.

While Western approaches to logic have thought in terms of both argument and inference, the linguistic approach of analyzing arguments, not inferences, has come to be dominant. Many of the distinctive features of Western logic arise because support relationships are treated linguistically, and premises and conclusions are treated as statements. In the remainder this book will talk of scientific reasoning in terms of arguments. When we intend to talk about psychological relationships of support, we will explicitly us the term "inference."

3.2 Deductive Validity

Arguments support their conclusions with different degrees of reliability. Some arguments guarantee the truth of their conclusion, given that the premises are true. Arguments of this kind have been very important in philosophy, mathematics, and science. Western philosophers have extensively studied arguments that guarantee the truth of their conclusions, and this study of logic is called "deductive logic."

But there are many good arguments that fall short of this high standard. In everyday life, we accept as strong many arguments that make their conclusions very likely to be true, even if they do not guarantee the truth of their conclusions. The study of support relations that do not guarantee the truth of the conclusion is known as "inductive logic." Inductively strong arguments are especially important in science. In the last several hundred years, philosophers and mathematicians have made great advances in finding criteria that identify deductively valid and inductively strong arguments. In this chapter we will discuss deductive arguments and outline their use in scientific reasoning.

If a deductive argument is to guarantee the truth of its conclusion, it must be impossible for the conclusion to be false in any case where the premises are all true. Arguments with this property are called "valid:"

Definition of Validity: An argument is valid if and only if there is no possible situation where its premises are true and its conclusion false.

Clearly, a valid argument would guarantee the truth of its conclusion in any case the premises were true. It is important to recognize that this definition characterizes a relationship between the premises and conclusion. Notice that it does not require that the premises are, in fact, true. Nor does this definition require that the conclusion is true. It only requires that the conclusion be true in any case where the premises are true. To see what this definition really involves, let us turn to some examples. Consider Argument 3.4, below.

Some monks study science.
Some monks drink tea.

Some who drink tea study science.

Argument 3.4

To determine whether Argument 3.4 is valid, we must ask whether it is possible for the premises to be true while the conclusion is false. In this case, it is easy to imagine such a situation. Suppose that there were no monks who both drink tea and study science. That is, suppose the conclusion were false. This would mean that there are separate two groups of monks—those who drink tea and those who study science—and there would be no overlap between these groups. This supposition would make the premises true. So, there is a possible situation where the premises are true and the conclusion false; hence Argument 3.3 is not valid.

Now, consider Argument 3.5, below. This argument is valid. It is impossible to find a situation where the premises are true and the conclusion false. We might think about it this way: suppose the conclusion were false. This would mean that there must be some monk who did not respect life. But if the first premise is true, then this monk must be compassionate. And if the second premise is true, then he must also respect life. So supposing that the premises are true and the conclusion false leads to a contradiction.

All monks are compassionate
All compassionate persons respect life

All monks respect life.

Argument 3.5

One might worry that identifying valid arguments by imagining possible examples is not very reliable. What if we simply failed to think of the situation where the premises are true and the conclusion false? Could we mis-identify the validity of an argument? Philosophers have developed several techniques for definitively identifying the validity of an argument. They depend on the form of the argument, not on our imaginations (the idea of "form" will be discussed below in Section 3.3) We will demonstrate one here.

Arguments 3.4 and 3.5 relate terms that refer to groups or classes of individuals. Argument 3.4 relates the terms "monk," "study science," and "drink tea," while Argument 3.5 relates the terms "monk," "compassion," and "respects life." Each sentence in the arguments says that some or all members of one group are members of the other. We will represent each term with a circle. One can think of the circle as encompassing all of the individuals who are monks, who are compassionate, and so on.

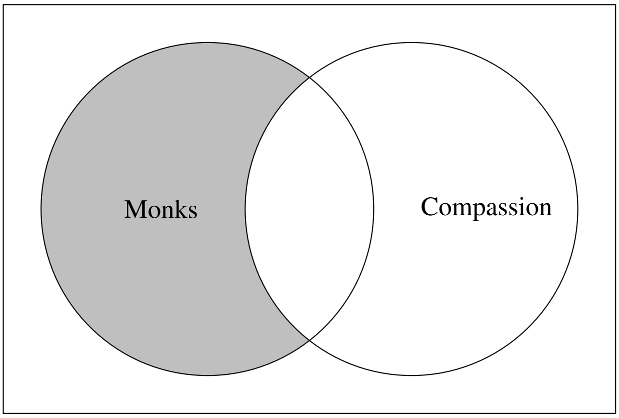


Figure 3.6

Figure 3.6 represents the sentence "All monks are compassionate." The circle on the left encompasses the monks, and the circle on the right encompasses compassionate beings. Since all monks are compassionate, any individual inside the monk circle must also be inside the compassionate circle. This means that the left side of the monk circle must be empty. We indicate this situation by making the left part of the circle gray; the darkened part of the circle has nothing in it. Of course, there may be individuals who are compassionate, but who are not monks. Hence the right side of the compassionate circle remains white.

To represent Argument 3.5, we need to add a circle for the term "respects life." Figure 3.7 does so. The second premise of Argument 3.5 says that all compassionate persons respect life. Just as before, to represent this statement, we must darken that part of the circle that is empty. So, in Figure 3.7, that part of the compassion circle that is outside of the respects life circle has been darkened.

Inspecting the diagram of Figure 3.7 makes it certain that the conclusion must be true in any situation where the premises are true. The only part of the monk circle that remains unshaded overlaps with the circle of those who respect life. Hence anything that is a monk must be something that respect life. The argument is valid.

This example has demonstrated a particular technique for assessing the valid-

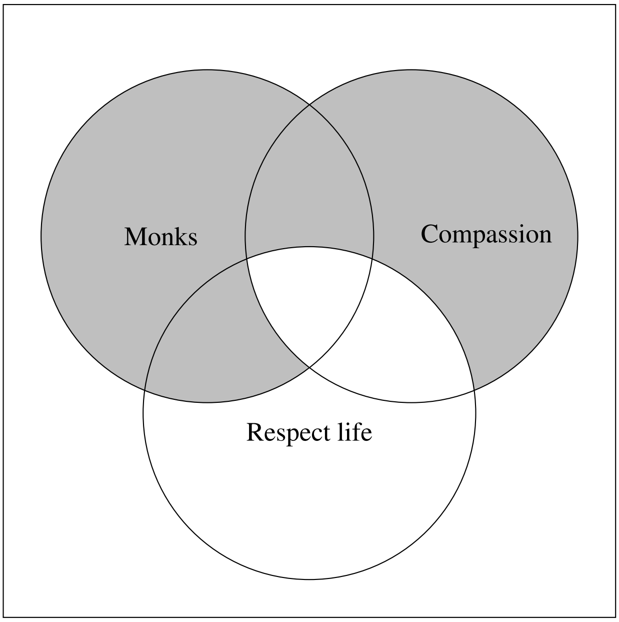


Figure 3.7

ity of arguments. Western philosophers and mathematicians have extended this technique, and others like it, to clearly delimit the class of valid arguments. It is beyond our scope to discuss them all; we are interested in the consequences of this conception of a valid argument for scientific reasoning. (For those who are curious about how these logical techniques can be extended, the Appendix at the conclusion of this book provides a brief introduction.)

3.3 Validity, Soundness, and Form

An important consequence of the definition of validity is that an argument can be valid even if the premises are false. Consider Argument 3.8. Figure 3.9 diagrams the argument just as we did in the last section. Again, the diagram clearly shows that any murder must also be in the respects life circle. Hence, any situation in which the premises are true is one where the conclusion is true. Hence, perhaps surprisingly, Argument 3.8 is valid.

All murderers are compassionate
All compassionate persons respect life

All murderers respect life.

Argument 3.8

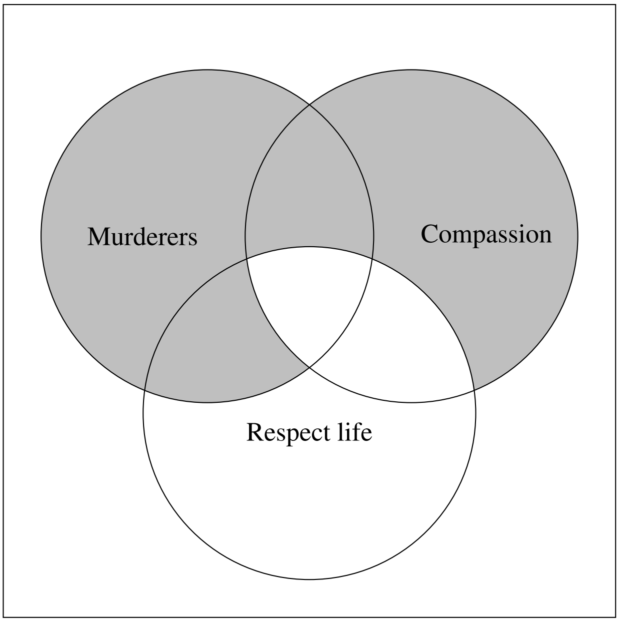


Figure 3.9

In spite of its validity, Argument 3.8 is not a good argument. Obviously, the premise “all murderers are compassionate” is false. Hence, it is possible for valid arguments to have false premises. The possibility of valid arguments with false premises means that there are two ways to evaluate arguments. We will distinguish these as “valid” and “sound” arguments.

Soundness: An argument is sound if and only if it is valid and all of its premises are true.

This definition requires that all sound arguments are also valid. But as Argument 3.8 shows, not all valid arguments are sound. Evaluating an argument as a basis for inference, then, has two parts. We need to determine whether the argument has a valid form. That is, are the premises and conclusions related in a way that will guarantee the truth of the conclusion *if the premises were true*? If the argument is invalid, then it is not giving us a definitive proof of the conclusion.[2] The second part of evaluating an argument as a basis for inference is to evaluate whether the premises are true. If the form of the argument is valid, and the premises are true (that is, if the argument is sound), then we have a very powerful reason for making the inference and believing the conclusion. If we find the premises to be false (as in Argument 3.8), the we have not been given a sufficient reason to accept the conclusion, even if the form is valid.

[2] While this is a point of criticism, we should be careful not to reject all invalid arguments. Inductively strong arguments are invalid—it is possible for their premises to be true while the conclusion is false—but the premises nonetheless provide strong support for the conclusion. So while invalid arguments like Arguments (3.3) and (3.4) provide little or no support for their conclusions, the inductive arguments we will discuss in the next chapter do strongly support their conclusions.

Comparing Figures 3.7 and 3.9 demonstrates another important feature of valid arguments. Both images are shaded in the same way. Indeed, the only difference is that "murderer" has been substituted for "monk." (This is true of Arguments 3.5 and 3.8 as well.) We could have substituted any words for "monk," "compassion" and "respects life." Figure 3.10 illustrates this by using the letters A, B, and C as place holders for words.

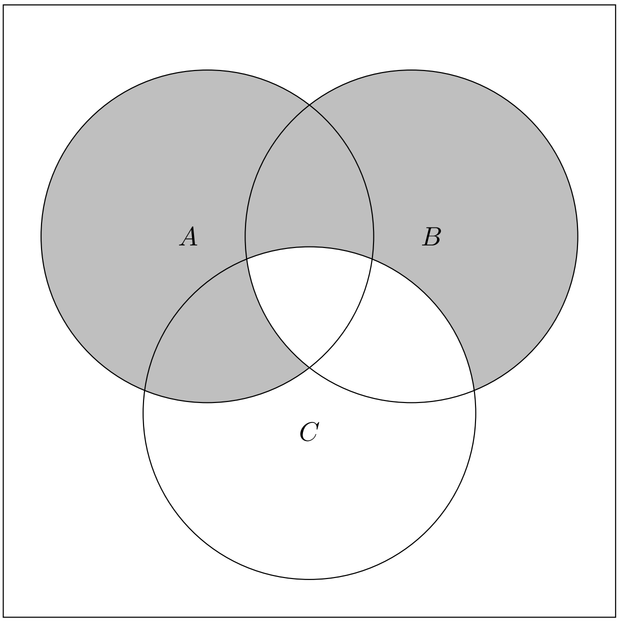


Figure 3.10

Validity is a product by the *relationship* among the words, not the particular words themselves. In arguments 3.5 and 3.8, the relationship is expressed by the other words in the phrases "All ... are ...." We can represent this by replacing the specific words of the sentences with variables A, B, and C, as in Argument 3.11, below. The diagram in Figure 3.10 demonstrates the validity of the form expressed by 3.11. This is what we mean when we say that validity is a matter of form. Any argument that shares this form—that is, any substitution of words for the variables A, B, and C—must be valid. The form in Argument 3.11 is only one of many argument forms that are valid.

All A are B
All B are C

All A are C

Argument 3.11

We call the letters A, B, and C "variables," and they are like variables in mathematics. While mathematical variables are replaced by numbers, the variables in argument forms are replaced by words that pick out properties or groups of individuals. By treating validity as a result of form in this way, contemporary logic is closely related to mathematics.

$a + b = c$

$a = 2$

$b = 3$

$c = 5$

Argument 3.12

Consider the calculation in Argument 3.12. The variables in this calculation a, b, and c can be replaced by any numbers. The first premise is a simple mathematical formula, while the second and third premises specify the values of this formula. This, too, is a deductively valid argument, as is all mathematical calculation. Contemporary western logic has the capacity to represent and demonstrate the validity of mathematical argumentation. Therefore, when formulas, like $F = ma$, are used in science to make calculations, science is using deductive logic. And if the calculations are not in error, then the arguments are valid.

This section has discussed two of the most important features of Western logic: the distinction between validity and soundness, and the relationship between form and validity. The two notions are intimately related. An argument is valid because of the relationship among the terms, not because of what the terms mean. That is what it means to say that an argument is valid in virtue of its form. However, since the meaning of the terms does not matter for validity, the actual truth or falsity of the premises does not matter for validity either. Hence, arguments can be valid without being sound.

3.4 Falsification

The last section distinguished between valid and sound arguments. One might wonder what the point of identifying valid, but unsound, arguments would be. Knowledge is the goal of inquiry, and if our arguments do not produce true conclusions, then they do not produce knowledge. However, while it may seem surprising, we can learn quite a lot from valid but unsound arguments.

An argument is valid when there is no possible situation where its premises are true and its conclusion false. If the conclusion of a valid argument is false, then we can be sure that at least one of the premises is false. Recall Argument 3.8. The conclusion, "All murderers respect life," is false. But the argument is valid, as demonstrated by Figure 3.9. Since it is valid, if both premises were true, the conclusion would have to be true too. Since the conclusion is false, one of the premises *must* be false; this we know.

Let us turn to a scientific example. In the last chapter (Section 2.2) we discussed the arguments used by scientists to show that the Copernican theory was better than the Ptolemaic theory. The Ptolemaic theory holds that Venus orbits around the earth, and that Venus is closer to the Earth than the Sun. A valid deductive argument—sketched as Argument 3.13, below, shows that Venus must always be observed as a crescent, never with its full face.

> Venus is always between the Sun and the Earth.
>
> If Venus is always between the Sun and the Earth, then Venus will always appear as a crescent when viewed from the Earth.
>
> ---
>
> Venus will aways appear as a crescent when viewed from the Earth.
>
> Argument 3.13

The truth or falsity of the conclusion of this argument can be determined by observation: we can watch Venus carefully over the course of a year to see whether it always has a crescent shape. If we see Venus as round at any time, then the conclusion must be false. The argument is valid, so we know that *at least* one of the premises must be false. Scientists rejected Ptolemy's theory because it had been falsified.

Falsification: A theory is falsified if and only if

1. a hypothesis has been validly deduced from the theory, and
2. the hypothesis has been observed to be false

This definition of falsification implicitly gives us a definition of a **hypothesis** as well: a hypothesis is a statement that has been validly deduced from a theory and can be observed to be either true or false . The conclusion of Argument 3.13 is a hypothesis. In Lind's experiment on scurvy (discussed in Section 2.3) his hypothesis was that sailors who added acid to their diet would recover from scurvy. Lind's experiment did *not* falsify his theory because, unlike Ptolemy, the hypothesis was observed to be true. Linds work also introduces some other complexities, and will talk more about it in the next Chapter.

Chapter 2 characterized science as "systematically" testing theories by observation. The process of falsification is one of the ways in which science systematically tests its theories. It is crucial that scientific theories have observational consequences that can be used to test the theory. If a theory is to be testable in this way, it must be detailed and specific enough to support valid deductions of observational statements. The creation of theories that can be falsified is one of the ways in which science is systematic.

Falsification is a somewhat more complicated process than what we have described so far. When the conclusion of a valid argument is false, then *at least* one of its premises must be false. Arguments typically have many premises, and the deduction of a hypothesis depends on all of them. A falsification does not tell us *which* of these premises is false, only that at least one is false. Since Argument 3.8 has a valid form, the falsity of "All murderers respect life" means that either (or both) "all murderers are compassionate," or "all compassionate persons respect life" must be false. Logic will not tell us which of these premises is false; we have to make that judgment on other grounds.

In the example of Argument 3.8, it is clear that the premise "all murderers are compassionate" is false. Murder is an act that lacks compassion. When scientific theories are tested, matters are more complicated. Ptolemy's theory makes many claims, and the deduction about how Venus will appear depends on more premises than are represented in Argument 3.13. It might be some other part of Ptolemy's theory that is false, not the claim that the earth is at the center of the planetary system. An important part of scientific work, then, is to determine which part of a theory needs revision when a hypothesis has been falsified. We will return to this idea in Section 8.4 and explore some of its consequences.

Falsification is the basis for a very powerful method of systematically testing a theory by observation. The first step is to validly derive a hypothesis from the theory. Again, a hypothesis is a statement that can be observed to be true or false. We then make observations to determine the truth of the hypothesis. Since valid arguments have a form that makes it impossible for the conclusion to be false when the premises are true, when a hypothesis is observed to be false, we know that some part of the theory must be revised. Indeed, as long as the deduction of the hypothesis is valid and we are confident of the observation, we can be quite sure that the theory needs revision. But again, hypothesis testing alone does not tell us how to revise the theory.

While falsification provides strong reason for thinking that a theory needs revision, scientists do not simply throw out a theory when a hypothesis is observed to be false. We are typically unsure exactly how to revise a theory. It takes time and further testing to find the best revision. Also, even if a theory is flawed, there may be no better alternative. Hence it makes sense to keep trying to perfect the current theory. The test of Ptolemy's theory by the observation of the phases of Venus is particularly dramatic because, not only was there an alternative available (Copernicus), the two theories made opposite predictions. Falsification of Ptolemy meant confirmation for Copernicus.[3]

What happens if a hypothesis is observed to be true? It may be surprising, but

[3]And again, the situation is even more complicated than we are presenting here. As we discussed in Section 2.2, Copernicus' theory was more consistent with the new physics that was emerging during this period. We will discuss the consequences of this point in Section 8.4.

such verification is not as epistemically powerful as a falsification. Just as valid arguments can have false premises and true conclusions, false theories can entail true hypotheses.

In Section 2.1, we mentioned that the practice of science requires some personal habits and attitudes. Falsification illustrates one of these attitudes: humility. If a scientist is to test a theory by deriving hypotheses, he or she must try to show that it is false. This means that a good scientist cannot become too attached to a theory. The self-correcting character of science requires that theories be modified when they are contradicted by observation. Theories are created, then, only to be replaced by better theories. The scientist who seeks to protect his or her favorite theory from falsification is a poor scientist.

Chapter 4

Good Inductive Practice

Arguments and inferences come in a range of strengths, depending on the support relationship between the premises and conclusions. Valid arguments provide the strongest form of support, since the truth of the premises guarantees the truth of the conclusions. In this chapter, we turn our attention to arguments where the premises strongly support the conclusion, but do not guarantee that the conclusion is true when the premises are true. These arguments are known as *inductive* arguments.

Historically, inductive reasoning has been viewed with skepticism by both Western and Buddhist philosophers. David Hume (1711–1776 CE) famously argued that inductive inferences failed to provide rational justification for their conclusions. On his view, our belief in the conclusions of inductive arguments was nothing more than a habit of the mind. Inductive reasoning has also been viewed with suspicion by Dharmakirti (c 7th Century CE). Dharmakirti considers Arguments 4.1 and 4.2, below.

Three mangoes from this tree are ripe.

All mangoes on this tree are ripe.

Argument 4.1

One spoonful of rice from this pot is cooked.

All of the rice in this pot is cooked.

Argument 4.2

Dharmakirti thought that Arguments (4.1) and (4.2) did not yield valid cognition of their conclusions. Similarly, from the perspective of Western logic, if we approach these inferences[1] from the perspective of deductive validity, they seem too weak to ground knowledge. It is entirely possible that the conclusions could be false, while the premises are true. Yet, at the same time, inductive inference has been crucial to the success of contemporary science. How can such an apparently weak form of argument produce such striking results?

[1]As we did in Chapter 3, we will distinguish inferences from arguments. Hume and Dharmakirti were skeptical of inductive inference. In this chapter we will be primarily concerned with arguments. Later we will turn to the question of whether such arguments are sufficient for belief, and ultimately, scientific knowledge.

4.1 Inductive Support

An argument was defined on page 19 as "one or more statements (premises) whose truth supports the truth of another (the conclusion)." We noted there that not all arguments were valid, yet the premises of some invalid arguments provide support—sometimes very strong support—for their conclusions. This gives us the definition of an inductive argument:

Inductive Argument: An argument is inductive if and only if it is not valid, and the premises provide support for the conclusion.

Since the premises of inductive arguments do not guarantee the truth of the conclusion, it is possible for the premises to be true and the conclusion false. This means that inductive arguments are invalid. Nonetheless, contemporary western scientists and philosophers hold that it is still possible for premises to support a conclusion, even if that support falls short of the guarantee provided by deductive validity.

In the evaluation of inductive arguments, the question is how well a set of premises supports the conclusion. Stronger arguments provide more support, weaker arguments provide less. The key idea, then, is *inductive strength*.

Inductive Strength: The strength of an inductive argument is the degree of likelihood that the premises confer on the conclusion.

An inductively strong argument makes the conclusion very likely to be true, while an inductively weak argument does not make the conclusion very likely to be true. "Weak" and "strong" form a scale, from those that provide almost no support for the conclusion to those that make the conclusion almost certain.

The key idea in the conception of strong and weak inductive arguments is the idea of "likelihood." As a first pass at this idea consider the following example. Suppose I have bought a new lamp for my room and inserted a new bulb in the socket. I plug it in to a socket that has been working all day, notice that the lights are on elsewhere in the building. I am ready to flip the switch for the first time. Under these conditions, I expect the lamp to light; I would be surprised if the lamp did not do so. This expectation is not capricious. New lamps and bulbs made by reputable manufacturers usually work, that is it is *likely* that the lamp will work. The inference, in this circumstance, is inductively strong.

Let us now represent my inference as an argument:

There is electrical power to the circuit.
The light bulb and lamp are new.

The light will go on when the switch is flipped.

Argument 4.3

This argument is inductively strong under the same conditions as my inference was inductively strong. If the manufacturer is reliable and the premises are true, the conclusion is likely to be true. Indeed, we could measure the strength of the argument by measuring the reliability of the manufacturer. How often does this manufacturer produce working bulbs or lamps? If the answer is "quite often," then the truth of the premises makes the conclusion likely to be true and the argument is strong. If the answer is "rarely" or "only sometimes," then the argument is weak.

The example of the light bulb might give you the impression that the conclusions of inductive arguments are true often or most of the time, but this would be a mistake. The conclusion must be likely *given that* the premises are true. It is also possible to draw conclusions about events that are unusual, even rare. Understanding the informal idea of likelihood requires developing the mathematical idea of *probability*. The mathematical investigation of probability and statistics has provided a deep understanding of inductive support. As a result, statistics is the basis for many strong inductive inferences. Chapter 5 will provide an introduction to some of the most important concepts.

Inductive strength is importantly different from deductive validity in several ways. Likelihood (or probability) is a matter of degree, so inductive strength is a matter of degree as well. Deductive validity, by contrast, is an all-or-nothing affair. Second, while deductive validity can be demonstrated by analyzing the form of the argument, inductive arguments cannot be evaluated in this way. Strong inductive arguments do not share a form analogous to that found among deductive arguments. This means that the criteria for evaluation do not arise simply from the relationships among terms in an argument, as it does for deductive logic. The strength or weakness of inductive arguments depends on what the premises and conclusion say in a particular context. This means that we must attend carefully to the meaning of the terms in the premises and conclusion and to the details of the context. Third, it follows from the lack of a form shared by inductive arguments that we cannot evaluate inductive arguments by looking at the premises and conclusions in isolation. We must use other information available to us. In the example above, the reliability of the lamp's manufacturer was an element of the context that determined the strength of the argument. So-called "background" information is crucial for assessing the strength of an inductive argument.

The goal of the next three chapters is to explain how arguments can be inductively strong and to articulate some criteria for distinguishing between strong and weak inductive arguments. Arguments like those about the mangos (Argument 4.1) and the rice (Argument 4.2) are not the sort of inductive arguments used by scientists. Scientific inductions take place in the context of theory development and testing. In the remainder of this chapter we will discuss examples of good inductive practice, showing the variety of strong inductive arguments in science. Chapter 5 will then begin to develop criteria for evaluating inductive arguments for their strength. With those criteria in hand, we will return to Arguments 4.1 and 4.2, to see whether Dharmakirti's and Hume's concerns have been addressed.

4.2 The Strange Habits of Grizzly Bears

Grizzly Bears are native to the mountains of western North America. They are large (adults can weigh more than 1000kg) and eat a very wide variety of foods. They move among different food sources depending on season, eating both meat and plants as they are available. Grizzly Bears have been observed migrating in the early spring to high mountain tops. These mountain tops are rocky and barren, too high to support either plant or animal life. There seemed to be no food source or other attraction for the bears. Biologists were puzzled by this Grizzly Bear behavior. Why would the bears spend so much effort to climb the mountains? Do they meditate? Do they enjoy the view?

Biologists think that animals direct most of their energy on activities that support their lives: finding food, building shelters, finding mates, raising offspring. Activities that do not function to enhance fitness tend to be weeded out by evolution. It seems likely, then, that the Grizzly Bears were doing something that helped them survive. Biologists observed the bears turning over rocks. When the biologists looked under these rocks, they found insects: larvae that would later turn into moths. These fat worms provide a very high energy food. This, the scientists inferred, was the explanation of the strange Grizzly Bear behavior. They climbed the mountain to find a particularly desirable food.

The biologists concluded that all Grizzly Bears who climb to barren mountain tops in this area do so to find moth larvae. Notice that this conclusion goes beyond anything that the biologists observed. They only observed *some* bears; they did not observe *all* the bears. They thus made an argument—Argument 4.4—that is rather like Arguments 4.1 and 4.2.

Observed Grizzly Bears climb to barren mountain tops to eat moth larvae

---

All Grizzly Bears (in a particular area) climb to barren mountain tops to eat moth larvae

Argument 4.4

The premise of Argument 4.4, like Arguments 4.1 and 4.2, refers to a small group of observed individuals. We will call this a "sample." The conclusion refers to a larger group of individuals, what we will call a "population." After observing the sample, we conclude that the whole population shares some of the same properties as the sample. In the next Chapter, we will discuss some of the characteristics that can make an argument like this strong. One of the argument's features is worthy of immediate comment. Notice how the background understanding of biology made the argument more plausible. Because of our background knowledge of animal behavior, we expected the bears to be looking for food. This is something that we expect all bears to do. When we observe the bears in the sample foraging for a kind of food, it is plausible that all bears do so. We will see in this Chapter that such uses background knowledge are pervasive in strong inductive inference.

4.3 Lind's Experiment as an Inductive Inference

Lind, who we encountered in Section 2.3, was seeking a solution to the widespread problem of scurvy. Sailors on long voyages would suffer from the disease. Lind was looking for a way to cure this disease, and that means that he sought a causal relationship. He was seeking something that would cause the sailors to recover from their disease. In his scientific investigation, he used twelve sailors to test his theory. On the basis of his success with just three of these sailors, he concluded that citrus juice (the juice of lemons, limes, oranges, and similar fruits) could cure scurvy for all sailors on all long voyages. If we put Lind's inference into the form of an argument, it would have superficially similar form to Arguments 4.1 and 4.2. When we look at the details of Lind's proceedure, we find some very significant differences. Most importantly, neither Argument 4.1 nor 4.2 tries to establish a causal relationship.

Arguments with causal conclusions are subject to particular sources of error. An effect might come about through several different causes. For example, when we give a patient a potion and he or she recovers, it is entirely possible that the patient recovered for some other reason than our purported cure. Responses to illness are varied. Some who contract a disease get better without treatment, while others do not. If we simply administer a treatment to some people who are sick, we cannot distinguish between two possibilities. It could be that the treatment was ineffective and they simply got better on their own. On the other hand, the treatment might have been the cause of their recovery. To establish a causal conclusion, then, we need a procedure that will distinguish between those cases where the treatment was the cause of the recovery and those cases where it is not.

Lind's experimental method is designed to reduce the possibility of an erroneous conclusion in a causal argument. He divided the sailors into groups, some of whom got treated and others who did not. If the sailors were getting better for some reason other than the treatment, then all groups should show the same rate of recovery. If a treatment is causally effective, the group who received the treatment should do better than those who did not. By comparing those who were treated with those who were not, Lind was able to establish that sailors did not recover on their own, and thus that the citrus juice made a difference.

Lind made an inference on the basis of his experiment. If we expressed that inference as an argument, it would not have the same form as Arguments 4.1, 4.2, or 4.4. Because the experimental procedure is an important part of the support for Lind's conclusion, we would have to describe the procedure in the premises. A slightly simplified presentation of Lind's argument might be Argument 4.5, below.

Scientists take arguments like Argument 4.5 to strongly support their conclu-

A sample of sailors was divided into two groups.

The only difference between the groups was that one received a daily measure of citrus juice.

Those who received citrus juice recovered from scurvy, while the others did not.

Therefore, citrus juice cures scurvy.

Argument 4.5

sions. Developing experimental procedures has been one of the distinctive features of western science, and they have been an important part of scientific success. It is therefore important to explore the epistemological and logical features of such arguments to determine why they are strong or weak. Chapters 5 and 6 will do so.

4.4 Newton's Mechanics as an Inductive Inference

Newton's laws of mechanics are extremely well supported by the evidence.[2] The reason why Newton's theory is so well supported is more complicated than in the earlier examples. Newton's laws are not simple inductive generalizations from a sample, as was the argument concerning Grizzly Bears (Argument 4.4). And while experiments have been a very important source of support for Newton's theory, the laws were not directly tested in experiments. Where, then, does the support for Newton's theory come from?

Newton's theory is supported by a mixture of inductively strong and deductively valid arguments. We noted in our earlier discussion that Newton showed Kepler's laws of planetary motion to be consequences of his theory. We can now say it this way: Newton made valid deductions (in the form of mathematical calculations) of Kepler's laws from his laws. That is, he showed that if his laws were true, Kepler's laws would have to be true too. Kepler's laws, of course, were already believed to be true on the basis of an inductively strong argument. By Kepler's time, astronomers had made very precise observations of the relative positions of the planets and stars. Kepler's laws described these motions very accurately. The observed motions of the planets, however, are only a portion of all the observed motions. They are like the observations of Grizzly Bears discussed in Section 4.2. There, scientists observed the behavior of some bears and drew a conclusion about all bears. In astronomy, we have only observed some planetary motions. In particular, we have only observed those in the past. Kepler's laws describe the motions of planets in the past, present, *and future.* Support for Kepler's laws thus

[2]You may know that Newton's theory was replaced by Einstein's. However, Einstein's theory replaced Newton's in a different way than Copernicus replaced Ptolemy. If Einstein's theory of special relativity is true, then Newton's laws are true as long as the speeds of the objects are small relative to the speed of light. Newton's laws thus remain true within a limited domain.

relied on an inductively strong argument analogous to Argument 4.4.

By showing how Kepler's laws of planetary motion were a consequence of his more general laws of motion, Newton was able to explain why Kepler's laws were true. The planets move as they do because they are subject to the same constraints, like inertia and gravity, to which all objects are subject. The mere fact that Newton's laws could deductively entail Kepler's laws, however, is not a strong argument in their favor. Every conclusion can follow from more than one set of premises. So, the fact that a set of premises (like Newton's laws) can validly entail a conclusion known to be true is little reason to suppose that the premises are true. So, the deduction of Kepler's laws—while very striking at the time—is not by itself strong support for them. Stronger support for Newton's theory came from the additional fact that Newton could explain a wide variety of phenomena and unify them within a single framework.

In Newton's time a number of regularities about the motion of projectiles—such as an arrow or cannon ball—were known by generalizing from a sample of observations. Since ancient times it has been known that there is an optimal angle for projectiles that will produce the most distance (given a force, whether this be a person's arm, bowstring, or an explosion of gunpowder). Angles steeper than 45 degrees will fall shorter, as will lower angles (see Figure 4.6 for a diagram). This regularity could be calculated (validly deduced) from Newton's laws as well. And his laws not only showed that 45 degrees was optimal, it accurately predicted how far *any* projectile would go, given its mass, the force that propelled it, and the angle of initial flight.

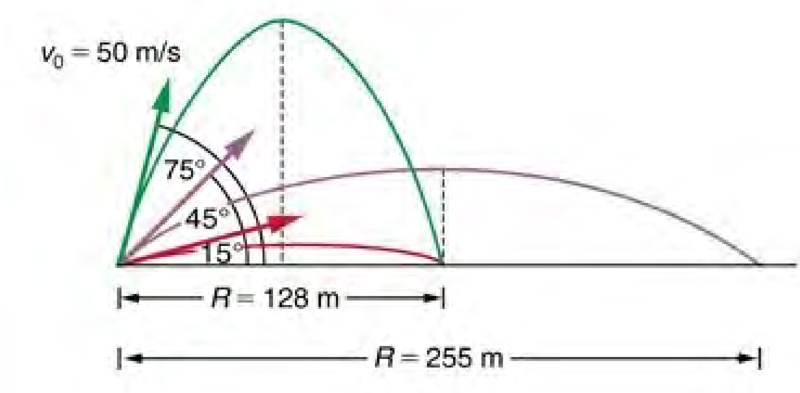


Figure 4.6: Given a constant velocity, $v_0 = 50$ m/s, the range, *R*, depends on launch angle[3]

Newton's laws could successfully explain other phenomena too. The pendulum is an important physical system for understanding acceleration. A pendulum is a simple device. A weight, called a "bob," is suspended by a string or rod. When

[3]Image used under Creative Commons Attribution 4.0 International License. Source: College Physics, OpenStax https://pressbooks.bccampus.ca/collegephysics/chapter/projectile-motion/

pulled to one side and released,[4] it swings back and forth in a regular way (see Figure 4.7). Indeed, the motion is so repetitive that very accurate clocks can be constructed from pendulums. Surprisingly, the time it takes for a pendulum to swing out and back (the "period" of the pendulum) is independent of the weight of the bob. If two pendulums of the same length are constructed, one with a heavy bob and the other with a light bob, they will have exactly the same period.

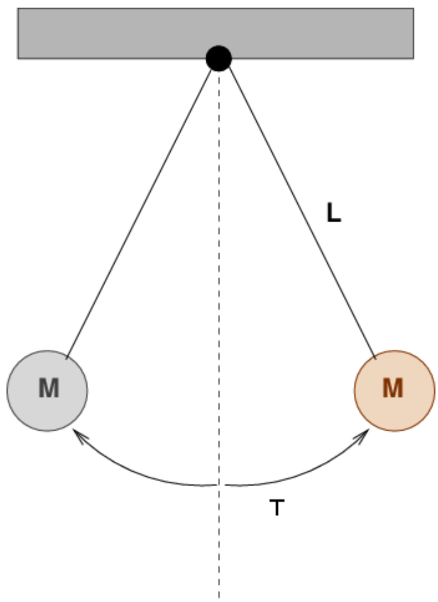


Figure 4.7: The period T of a pendulum depends on the length L of the rod, not on the mass M of the bob.[5]

The physicist Galileo studied the pendulum by constructing them with different lengths and measuring their periods. He found that shorter pendulums will swing faster and longer pendulums will swing more slowly. He determined that the period of a pendulum is proportional to the square root of the length. Expressed mathematically, this relation is:

$$T \propto \sqrt{L}$$

Notice that this law of pendulum motion was inductively supported in just the way Kepler's laws and the laws of projectile motion were supported. The regularity, however, was inexact. The symbol $\propto$ in this equation indicated that the value on the left is *proportional* to the value on the right. In other words, longer pendulums

[4] This description of pendulum behavior assumes that the pendulum is not pulled too far. Galileo's formula, below, breaks down for very large swings, roughly when the bob is raised above its pivot.

[5] Image used under Creative Commons License 2.0. Source: Michael Richmond, http://spiff.rit.edu/classes/phys207/lectures/pend_theory/pend_theory.html. Image modified by the author.

have longer periods. But the equation does not let us calculate a precise value for the period, given a pendulum's length. Newton improved on Galileo's inductive generalization by giving a new analysis of the motion. According to Newton, the force of gravity acted on the pendulum bob, and the acceleration due to this force, g, needed to figure into the equation. As a result, Galileo 's proportion was turned into an identity:[6]

$$T = 2\pi\sqrt{L/g}$$

Newton's laws, then, explained an extremely wide range of known phenomena about motion. And it did so with remarkable—indeed unprecedented—rigor and precision. As we have seen, this episode of science depended on both deductively valid and inductively strong arguments. As illustrated in Figure 4.8, below, Newton was able to show that his laws deductively entailed several known regularities. These regularities, in turn, were established by inductively strong arguments similar to Argument 4.4, where the premise is a sample and the conclusion is a generalization.

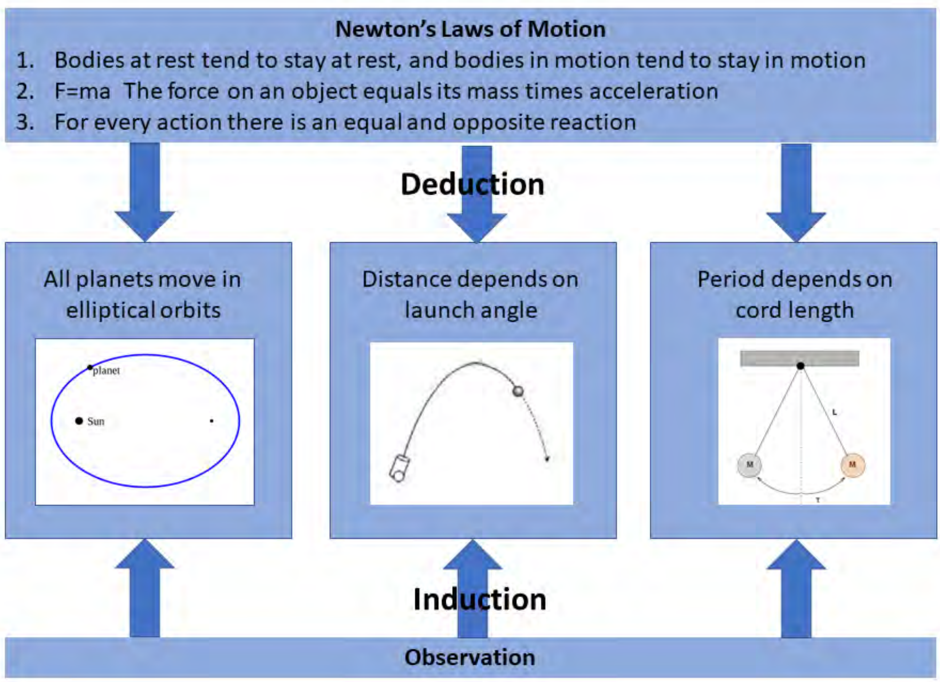


Figure 4.8: Newton's Laws are supported by both deductive and inductive arguments[7]

[6] And the 2π factor adjusts for the fact that the bob is moving in a circular path.
[7] Image by the author.

As we noticed above, the fact that Newton's laws could deductively entail one known regularity, like Kepler's laws, provides little reason to believe that the laws are true. However, Newton's laws deductively entailed a wide variety of phenomena (as illustrated in Figure 4.8). Its broad explanatory power makes Newton's laws likely to be true. While there are many sets of premises that can deductively entail a given conclusion, it is surprising to find a set of premises that entail a variety of conclusions known to be true. Newton's laws *unified* a wide variety of phenomena that were already known. Its explanatory power and capacity to show that all motions followed the same pattern provides strong inductive support for Newton's laws of motion. We will discuss this idea further in Section 8.4.

4.5 Conclusion: Forms of Induction in Scientific Practice

Each of the examples in this chapter has illustrated an inductively strong argument as it was used in scientific practice. In all three cases, the characteristics of the arguments that made them convincing were closely related to the particulars of the subject matter and the state of scientific knowledge at the time. There were also recognizable patterns or forms of reasoning that recurred. The example of Grizzly Bears generalized from a sample. The goal was simply to understand a pattern of behavior, and its strength was drawn partly from background knowledge of Grizzly Bears in particular and animal behavior in general. Lind's experiment also was based on sample, but it sought to identify a particular causal relationship. This meant that it took the form of an experiment. While it was a similar argument to the generalization from a sample, the experiment added several important premises to the inductive argument. The support for Newton's laws of mechanics looked quite different, since it involved a mix of deductively valid and inductively strong arguments. Unlike the first two examples, the overall support for Newton's laws came from its capacity to explain a wide variety of phenomena. Yet at the same time, generalizations from samples figured prominently.

So far, we have only sought to illustrate some of the reasoning used in scientific practice. Many other inductive arguments are used as well, though the three surveyed here are quite common. While we have called these "strong" inductive arguments, we have not provided criteria for strong arguments. Nor, indeed, have we provided a case that inductive arguments, arguments which do not guarantee the truth of their conclusions, can be the basis of scientific *knowledge*. It is to these matters that we now turn.

Chapter 5

Inductive Generalizations

In the previous chapter we saw three apparently different forms of inductive support used in scientific reasoning. In all three of these examples, arguments were used that had samples as their premises and generalizations as their conclusion. This sort of inductive argument is very common in science. In this Chapter, we will look more carefully at how these arguments work and provide criteria by which their strength can be evaluated.

5.1 Populations, Samples, and Distributions

In the foregoing examples, we spoke of "populations" and "samples." Let us begin by being more precise about these notions.

Population: A population is a number of objects among which we expect to find a distribution of a property of interest.

Sample: A number of individuals who are members of the population and who are observed for the property of interest.

The character of a population and the "property of interest" sometimes depends on the practical problems or questions that gave rise to the scientific inquiry. Lind was trying to find a cure for scurvy. Since scurvy affects all humans, the population for his generalizations was all human beings. His question was whether different liquids would cure scurvy. The "property of interest," then, was the variation among the liquids, e.g. that the sailors who received lime juice got better while those who drank seawater did not. Kepler was interested in predicting the motions of the planets. His population was all of the bodies that orbited the sun. The property of interest was their recurring relative positions and motions. A population is often all objects of a particular kind (humans, planets), but it may be more limited. In the Grizzly Bear example, you might have noticed that the subject was Grizzly Bears in a particular area of the western United States. In this case, we do not expect Grizzly Bears everywhere to behave in the same way. The availability of different kinds of food in different regions, and the bears' ability to learn this and pass it on to their offspring, means that bears in different areas will have distinct habits. This is why the population was limited to bears in a particular region.

It is important to recognize that the boundaries of the population are not definitively known in advance. Indeed, it has often turned out during scientific inquiry that what we thought was a homogeneous group of individuals is dissimilar in deep and important ways. Advances in scientific understanding can change what is included in our populations. While no other planetary systems were known in Kepler's time, we now think that most stars are surrounded by planets. The original population for Kepler's generalization was only our own solar system. Now, it is all systems of orbiting bodies. Re-thinking what is and is not included in the populations about which we are generalizing is one of the ways that science identifies and corrects errors.

In an inductive generalization, the premises describe the sample. The sample is only a part of the population, since we use this form of argument when the population cannot be observed in its entirety. (After all, if we could observe the entire population, there would be no need for an inductive argument.) The sample is inspected or measured to identify the presence of a property that will figure in the generalization. In the case of the Grizzly Bears, a small number of bears were observed to eat moth larvae. Many of the features that make inductive generalizations strong or weak depend on the characteristics of the sample and the way it is obtained. We will discuss samples extensively below.

In many of the examples so far, the conclusions of inductive generalizations say that that the whole population had the same property (for example, that all planets have elliptical orbits). However, this is not always the case. Many scientific inquiries are concerned with the way in which a property *varies* among members of a population. For example, the height of adult monks varies in the sense that some monks are taller and some are shorter. If it were important to know the heights of monks, we would generalize from a sample. We would measure the height of just a few monks. The conclusion of our inductive generalization would not be that all monks have the same height, but that their heights vary within a particular range.

Things will get more complicated as we go on, as scientists are also concerned with how various properties vary together. Height varies with age, for example. From our experience we know that during the early part of their lives, older monks are taller than younger monks. Such a conclusion compares two properties (age and height). This will be important when thinking about causation. Recall that in Lind's small study, it was significant that two properties were found in the same individuals: the consumption of citrus juice and the recovery from scurvy. We will discuss these more complicated cases of inductive generalization soon enough. For now, let us focus on the simplest case, one in which we are concerned with the variation of one property within a population.

When thinking about how some property varies across a sample or population, it is useful to use two notions: that of a distribution of a property, and that of a proportion. The property of concern is found to be "distributed in our sample" in a certain way, and we generalize that it is likely distributed in a very similar way in the population. To understand the similarity envisioned—the respect in which one should think of the distributions in sample and in population to be similar, we must understand the prior concept of a proportion.

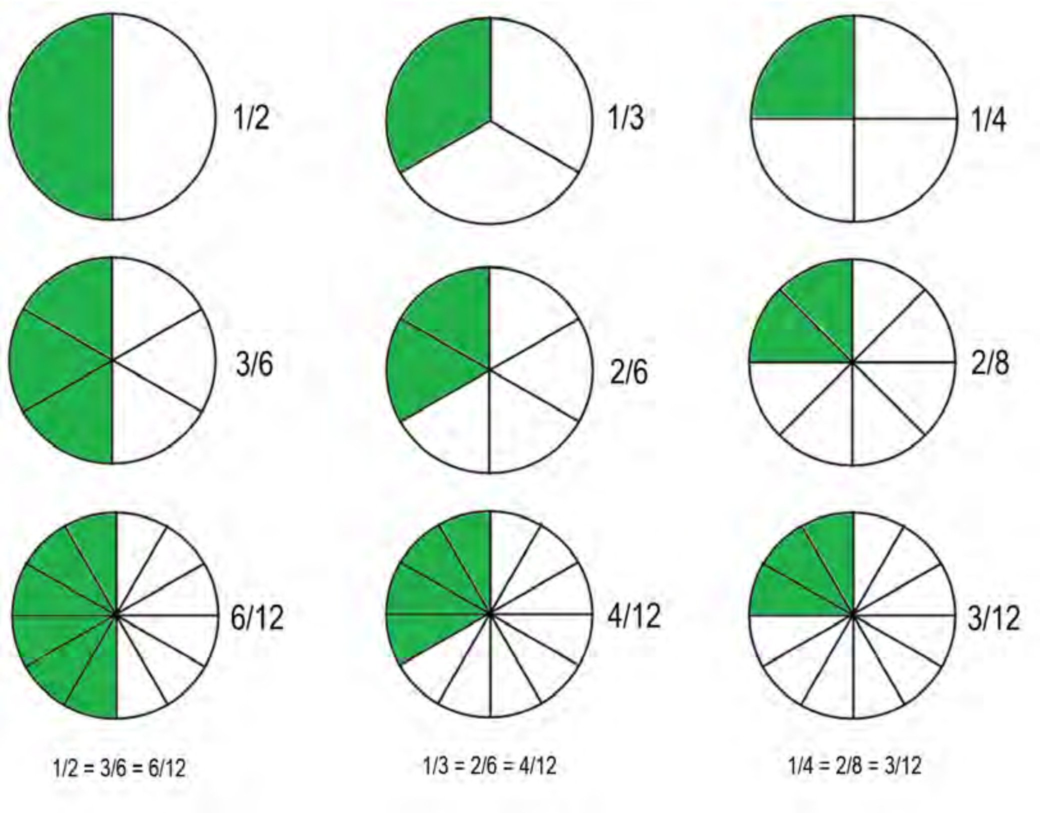


Figure 5.1: Equal fractions[1]

A **proportion** is a part of a larger whole, and it is measured by comparing its size to the size of the whole. For example, suppose three monks share a small loaf of bread by dividing it into three equal parts. Suppose on the next day, they share another, larger, loaf, again dividing it into equal parts. There is a sense in which the monks got the same share of bread on each day, even though each ate more bread on the second day. The monks got the *same proportion*. The usefulness of proportions is that they permit us to compare parts from different sized wholes; that is, to say whether the monks go the same share of bread, even though the size of the loaf changed. Proportions are expressed as fractions. The monks divided their loaves into three equal pieces, and each monk got one piece. Each monk, then, got one piece out of three, or one third: ⅓. Figure 5.1 shows some fractions

[1] Image used under Creative Commons License 1.0. Source: Wikimedia Commons. Modified by the author.

and their relationships.

A **distribution** is the proportion of the sample or population that has the property of interest. For example, suppose a farm has two chickens and one cow. Two out of the three farm animals, or two thirds (⅔) are chickens. This is the distribution of chickens in the population of farm animals on this farm. Suppose a larger farm has six chickens and three cows. The larger farm has the same distribution of chickens: ⅔. A farm that has only chickens has a distribution of chickens too: all members of the population are chickens.

Distributions let us link samples to populations in inductive inference. Just as we compared the distribution of chickens across two farms, above, we can compare the distribution of chickens in a sample and a larger population. Suppose we sample farm animals in a certain region, and find 200 chickens in our sample of 300 animals. In a strong inductive inference, the sample and population should be the same with respect to the property of interest. Of course a sample is always made up of a smaller number of individuals than the population. So sample and population cannot be the same in the sense that they have the same number of chickens. We do not expect that there are just 200 chickens in the much larger population of, lets say, 100,000 animals. Rather, we expect something more like ⅔ of the farm animals to be chickens—something in the neighborhood 66,000 chickens. If the inductive inference is strong, then the sample and population will have nearly the same distribution of chickens, expressed as the proportion of chickens to all farm animals. The ability to compare the proportions across different sized groups is central to inductive arguments based on sampling.

With these definitions in hand, we can finally give a precise definition of an inductive generalization:

Inductive Generalization: An inductive generalization is an argument where:

1. the premises describe the observed distribution of a property in a sample, and
2. the conclusion says that the distribution of the property in a population is nearly the same as in the sample

An inductive generalization, then, concludes that the unobserved members of the larger population are the same—or nearly so—as those we have observed. The phrase "nearly the same" in this definition marks the fact that we do not expect the distribution of the property in the population to be exactly the same as in the sample. If the argument is strong, then we expect distribution of the property in the population to be very close to its distribution in the sample. As we will see in Section 5.3, below, we are even able to reliably estimate how close to the real distribution in the population we can get with a particular sample.

The definition of an inductive generalization in this section has not given us any criteria for evaluation. Inductive generalizations may provide strong or weak support for their conclusions. The remainder of this chapter will explain how and why inductive generalizations can be strong.

5.2 Probability

The definition of a strong inductive argument given on page 32 used the notion of likelihood or probability. Our contemporary understanding of inductive arguments depends heavily on the mathematics of probability. Scientific arguments today typically use sophisticated mathematical tools to assess the likelihood of their conclusions. A full treatment of these ideas is well beyond the scope of this text. The goal of this section is to introduce some of the basic concepts, sufficient to support the evaluative criteria for inductive arguments.

Consider a standard, six-sided die: a cube with a number on each side. When rolled, it will come to rest with one side facing up. One or more dice are often used in games where chance plays a role. There are six possible outcomes of a role, since there are six sides. If the die is evenly formed, then there is an equal chance that it will land with any one of the six sides facing up. Since there are six possibilities, and all are equal, the **probability** of the dice landing with any particular number—say the number 5—facing up is ⅙. Probability is a number assigned to the chance of an event occurring. We represent probabilities this way:

$$\mathsf{Pr}(\text{die showing the number } 1) = {}^{1}\!/_{6}$$

This notation says that the chance or probability of a die landing with the number 1 showing is one out of six.

The number representing the probability that event will occur is always a fraction or decimal between, and including, 0 and 1. Probabilities can be 0 or 1. To say that the probability of an event, under particular conditions, is 0 is to say that it never happens; to say that the probability is 1 is to say that it always happens under those conditions. Probabilities represented by fractions or decimals close to 0 are very unlikely; those close to 1 are very likely. And probabilities represented by higher numbers are always more likely than a lower number. (No probability is greater than 1 or less than 0.)

Another way to think about probability is in terms of frequency. If a process is repeated, probability is the frequency with which an event will occur. If you roll standard dice many times, each of the six different sides will show approximately the same proportion of the time, that is, ⅙. And as more rolls are made, the more closely the proportions will be to ⅙. You can try this experiment with a coin: flip it in the air and let it land, then record which side is showing. (In English, we call the side of the coin with a face the “heads” side, and the other side “tails.”) Repeat as many times as you like. While the coin might land on the same side, say heads, several times in a row, the longer the process is repeated, the closer the proportion of each side will get to ½. The probability, then, that a coin will land heads is ½ or .5.

Probabilities are closely related to distributions. If you tried the experiment with the coin, above, you would have recorded a series of events. Within this population, there is a distribution of "heads" and "tails," indeed, the distribution should be very close to ½ heads and ½ tails. The probability *is* the distribution of the property of showing heads in this population of events.

Consider a slightly different example of distributions and probabilities. Suppose that a monastery had 100 monks and 90 of them were from Tibet and 10 were born in other countries. The distribution of monks born in Tibet in this monastery, then, is nine out of ten, or 9⁄10. If one were to close one's eyes and randomly point at a list of names of the monks from this monastery, the chance of pointing to the name of a monk from Tibet would be:

$$\mathsf{Pr}(\text{monk born in Tibet}) = {}^{9}/_{10} = .9$$

When choosing randomly from a population of individuals, the probability of choosing an individual with a given property is the same as the distribution of that property in the population.

So far we have considered only the probability of one event occurring. How do we calculate the probability of more than one event occurring? There are three ways in which a pair of event frequencies can be related, and the calculations for each are different. First, they may occur together (in conjunction) with some frequency. In this case we might want to know the probability of a monk being tall *and* born in Tibet. Second, the events might be alternatives. In this kind of case we might want to know the probability that a die will show *either* a six *or* a one. Finally, the occurrence of event might change the probability of a second event. For example, wildfires are much more likely in the dry season than in the wet season. The probability of a brush fire is *conditional* on the season.

Consider alternatives first. The probability of either one of two events, call them A and B, occurring is the sum of their probabilities:[2]

$$\mathsf{Pr}(A \text{ or } B) = \mathsf{Pr}(A) + \mathsf{Pr}(B)$$

If the probability of the number 1 showing after the roll of a die is ⅙, then the probability of *either* the number 1 or the number 2 showing is 2⁄6. And as you would expect, the probability of *either* the number 1, number 2, number 3, number 4, number 5, *or* number 6 showing is 1. That is, it will always happen that *one* of the sides will show a number after a roll.

When one event is conditional on another, we need to know the probability that one event will occur, *given that* or *conditional upon* another event occurring. Consider a deck of playing cards with 52 cards, half of which are red and half of

[2]This simple version of the rule assumes that A and B are exclusive; that is, there is no situation where both A and B obtain. The more general rule for calculating the probability of a disjunction is

$$\mathsf{Pr}(A \text{ or } B) = \mathsf{Pr}(A) + \mathsf{Pr}(B) - \mathsf{Pr}(A \text{ and } B)$$

which are black. The distribution of red cards in the deck is ½, so the probability of randomly drawing a red card (assuming that the deck is well shuffled) is one out of two:

$$\mathsf{Pr}(\text{red}) = {}^{1}\!/_{2} = .5$$

Now, suppose we have drawn one red card and have not replaced it in the deck. What is the probability of drawing a red card on a second draw? It cannot be the same because by removing the red card, we have changed the distribution of red cards in the population. Where there were 26 red cards in the deck of 52, now only 25 of the remaining 51 cards are red. That means the probability of drawing a red card, *given that* one red card has already been drawn is ${}^{25}\!/_{51}$, a slightly lower probability than ${}^{26}\!/_{52}$ (which equals ½). We represent the conditional dependence of the probability of one event, A on another, B this way:

$$\mathsf{Pr}(A \mid B)$$

This is known as a **conditional probability**. The notation $\mathsf{Pr}(A \mid B)$ is read as "the probability of A given (or conditional on) B." In the example of the cards, we would represent the conditional probability of drawing a red card on the second draw, given that a red card had already been draw in this way:

$$\mathsf{Pr}(\text{red on the second draw} \mid \text{red on the first draw}) = {}^{25}\!/_{51}$$

Using the notion of a conditional probability, we can express a related notion that will be very important in our discussion of sampling, the notion of **probabilistic independence**. In the foregoing example, each card was drawn from the deck and not replaced. The probability of the second draw depended on the first because we did not replace the first card, and this changed the number of cards in the deck for the second draw. Suppose we *did* replace the card and reshuffled the deck. Now the first draw does not matter for the second; the second draw is *independent* of the first. This means that:

$$\begin{aligned}\mathsf{Pr}(\text{red on the second draw} \mid \text{red on the first draw})\\ = \mathsf{Pr}(\text{red on the second draw})\\ = {}^{1}\!/_{2}\end{aligned}$$

The fact that one got a red on the first does not change the probability that one will get a red on the second draw. (Think of the composition of the shuffled deck from which one is drawing—it is the same in the second draw as it was in the first.) The first and second draws are probabilistically independent. In general, we can say that A and B are probabilistically independent when:

$$\mathsf{Pr}(A \mid B) = \mathsf{Pr}(A)$$

The final sort of combination we need to consider is the probability that two events will *both* happen. Let us continue working with the deck of cards. Suppose we are drawing cards and replacing them, shuffling the deck between each draw. What is the probability of drawing two red cards? That is, we want to know the probability of

$$\mathsf{Pr}(\text{red on the first draw } \mathsf{and} \text{ red on the second draw})$$

The simple rule for determining the probability of a conjunction is to multiply the probabilities:

$$\mathsf{Pr}(A \text{ } \mathsf{and} \text{ } B) = \mathsf{Pr}(A) \times \mathsf{Pr}(B)$$

The simple rule for conjunctive probabilities assumes that A and B are probabilistically independent. That is, the probability that B will occur is not changed by the fact that A occurred. By replacing and shuffling the cards, we make the two events independent. So, this simplified rule would tell us the probability of drawing two red cards, when the first card is replaced and the deck is reshuffled.

$$\begin{aligned}
&\mathsf{Pr}(\text{red on the first draw } \mathsf{and} \text{ red on the second draw})\\
&= \mathsf{Pr}(\text{red on the first draw} \times \text{red on the second draw})\\
&\qquad = {}^1\!/_2 \times {}^1\!/_2\\
&\qquad = {}^1\!/_4
\end{aligned}$$

In other words, while we would expect to draw a red card from this deck one out of every two times, we would expect to draw two red cards in a row only one out of four times. And this is as it should be: the probability of drawing two red cards in a row should be smaller than the probability of drawing one red card. And the probability should continue to go down for longer sequences of red cards, since the chances of drawing, say, 16, red cards in a row is very low.

If the properties A and B are not independent, then our calculation must be different. If the occurrence of A changes the probability of B, then the probability of B is its conditional probability. Hence, the general[3] rule for conjunctive probabilities is:

$$\mathsf{Pr}(A \text{ } \mathsf{and} \text{ } B) = \mathsf{Pr}(A) \times \mathsf{Pr}(B \mid A)$$

This rule would let us calculate the probability of drawing two red cards when the first card is *not* replaced, as in a card game. The probability of drawing a red card is ½. And as we saw before, the probability of drawing a second red card *given that* a read card has already been drawn is $^{25}/_{51}$. The calculation is therefore as follows:

$$\begin{aligned}
&\mathsf{Pr}(\text{red on the first draw } \mathsf{and} \text{ red on the second draw})\\
&= \mathsf{Pr}(\text{red on the first draw}) \times \mathsf{Pr}(\text{red on the second draw} \mid \text{red on the first draw})\\
&\qquad = {}^1\!/_2 \times {}^{25}\!/_{51}\\
&\qquad = {}^{25}\!/_{102}
\end{aligned}$$

[3] The rule is more general in the sense that it makes no assumptions about whether A and B are probabilistically independent. Since B is independent of A just in case $\mathsf{Pr}(B) = \mathsf{Pr}(B \mid A)$, the general rule is the same as the special rule when B is independent of A.

The mathematical study of probability and statistics is a sophisticated and powerful branch of contemporary mathematics. We have gotten only a small glimpse of this field, but it will be enough for our purposes. The important idea of this section is the idea of *probability* and its relationship to distributions in a population. The mathematical rules for calculating probabilities will figure prominently in the argument of the next section.

5.3 Sample Size

We have assembled the concepts necessary to make the central argument of this chapter: that there are strong inductive arguments. In particular, a well constructed argument from a sample to a generalization—what we have been calling an inductive generalization—makes its conclusion likely to be true. As defined on page 44, the premises of an inductive generalization describe the observed distribution of some property in a sample. The conclusion says that in the population, the distribution of the property is nearly the same as in the sample. In this section we will see why the distribution in the population *must* be nearly the same as the sample (assuming certain conditions hold). The mathematical result that lies at the foundation of the argument is that **as the size of the sample increases, the probability that the sample is like the population increases**. Indeed, the probability increases in precisely specifiable ways that permit us to precisely characterize what it means to say that the sample is "like" or "nearly the same as" the population.

In Section 5.2 we discussed an example of drawing cards from a deck. While we will use a different example in this section, we made some assumptions about the cards that were important and will apply to the example in this section. In the card example, we assumed that each time we drew a card, it was replaced and the deck was re-shuffled. These assumptions made each draw probabilistically independent of the next. The process of reshuffling and drawing a card at random also ensured that the each sample is independent of the others (we will characterize independence more precisely below). The argument in this section will assume that the samples are drawn in a way that ensures independence.[4]

For our example in this section, let us suppose we have a very large bin of mixed beads. The beads are either red or black, and the bin is so large that counting them all would be very difficult. The beads in the bin are our population. We will stipulate that half of the beads are red and half black. This distribution in the population means that the probability of drawing one red bead is the same as the probability of drawing one black bead: ½ or (.5). Note that we could stipulate any distribution of color in the bin. This would change the probability of drawing each color, and hence the calculations below would look different. However, different stipulations would not affect the argument; the argument goes through no matter what the actual distribution in the population.

[4]The assumption of independence permit a relatively simple presentation of the mathematical argument. However, they are not required. There are mathematical techniques that permit similar arguments to succeed when the elements of the samples are not fully independent.

The argument of this section, again, is to show that as the sample size gets *larger*, the *likelihood* that the distribution in the sample is *near* the distribution in the population increases. This conclusion is complex because three things are changing together: sample size, likelihood, and nearness. To see how and why they are changing, we will fix the range of what counts as "near," then show that as the sample size increases, the likelihood that the sample's distribution falls within that range increases. When we understand why this happens, it will be clear why the three values change together.

Let us begin with the notion that a sample is "near," "similar to," or "like" the population. We will think of this in terms of distributions. The property of interest has a distribution in both the sample and the population. In our bin of beads, the distribution is the proportion of red and black beads. Since the distribution of beads in the population is ½ red, a sample of 8 beads with 4 reds would be exactly the same as the population. If the sample had one more red bead (5 red and 3 black) or one less (3 red and 5 black), it would be natural to say that the sample was "similar to" or that its distribution was "near" the distribution of the population. Since samples can vary in size, we must use proportions to characterize similarity. In a sample of 8 beads, one bead is ⅛ of the sample. So, let us say[5] that a sample's distribution is *near* the distribution in the population if the distribution in the sample is no more than ⅛ (or 12.5%) higher or lower than the population.

What we will now do is consider four sample sizes: two, four, eight, and sixteen. We will show that as the sample size gets larger, the distribution in the sample must get nearer the distribution of the population.

In a sample of two beads, a bead is drawn from the bin, its color noted, and then it is replaced and the bin is mixed. Then a second bead is drawn, replaced, and mixed in. There are four possible combinations of colors that could be drawn, as shown in Table 5.2. The probabilities for each combination are calculated with the simple conjunction rule, since we are finding the probability of drawing the indicated color on the first draw *and* the indicated color on the second draw (and replacing and mixing makes the samples independent).

First Draw	Second Draw	Probability
Red	Red	¼
Red	Black	¼
Black	Red	¼
Black	Black	¼

Table 5.2

While there are four possible combinations that could be drawn with a sample

[5] We are again stipulating; a different criterion for "nearness" could be used here. Recall that the strategy of the argument is to fix what counts as "similar" or "near," and show that as the sample gets larger, the likelihood that its distribution is near the distribution of the sample increases. This result would hold no matter what criteria we chose for "nearness."

of two beads, there are only three possible distributions of colors in the sample: all red, half red and half black, and all black. Each of these distributions has a probability. In other words, we can ask, what is the likelihood that a sample will contain half red beads? What is the probability of all black beads? The probability of each distribution can be calculated using the rules discussed in Section 5.2. The probability of getting a distribution that is all red or all black is represented in Table 5.2. Both of these distributions have a probability of ¼ (or .25). Notice that there are two ways for the distribution of half red to arise. We could have drawn a red bead, then a black one, or drawn a black then a red. This means we need to calculate the probability of getting *either* red then black *or* black then red. The simple disjunction rule says that in this case we must add the probabilities. So the probability of getting half red beads in the sample is $1/4 + 1/4 = 1/2$ (or $.25 + .25 = .5$).

We can represent the probabilities of getting samples with different distributions in a graph. In Figure 5.2 each of the three possible distributions of red beads within our sample of two beads is represented with a vertical bar. The height of the bar represents the probability of drawing a sample with that distribution. Given our criterion for nearness, only one of these three possible distributions counts as near: the middle bar representing half red beads. The other two bars are not within ⅛ (or 12.5%) of the population's distribution. This means that half the time, with a sample of two, we would draw a sample with a distribution near to the population (indeed, exactly the same as the population!). Being correct half the time may seem good, but notice that the other half of the time our sample is very different—all red or all black. So, with a sample of two we have equal chances of being exactly right and completely wrong.

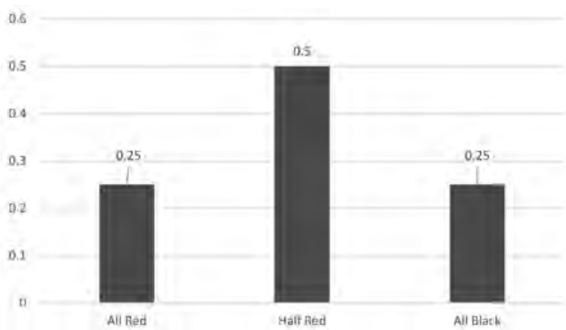


Figure 5.3: Sample of Two

Suppose we double the size of the sample from two to four beads. Now there are sixteen possibilities, as represented in Table 5.4. As in the sample of two, the probability of getting each of the possible samples is calculated with the simple conjunction rule.[6]

[6]In other words, the probabilities on each row of Table 5.4 are calculated by multiplying the probability of drawing a red or black bead (which is $1/2$) each of four times: $1/2 \times 1/2 \times 1/2 \times 1/2 = 1/16 = 0.0625$

First Draw	Second Draw	Third Draw	Fourth Draw	Probability
Red	Red	Red	Red	1/16
Red	Red	Red	Black	1/16
Red	Red	Black	Red	1/16
Red	Red	Black	Black	1/16
Red	Black	Red	Red	1/16
Red	Black	Red	Black	1/16
Red	Black	Black	Red	1/16
Red	Black	Black	Black	1/16
Red	Red	Red	Red	1/16
Black	Red	Red	Black	1/16
Black	Red	Black	Red	1/16
Black	Red	Black	Black	1/16
Black	Black	Red	Red	1/16
Black	Black	Red	Black	1/16
Black	Black	Black	Red	1/16
Black	Black	Black	Black	1/16

Table 5.4

As before, we are interested in the distribution of red and black in the sample, and many of the rows in Table 5.4 have the same distribution. There are really only five possibilities here. Figure 5.5 shows the probabilities for getting all red, three out of four red, and so on. Again the probability for each of these distributions is calculated by adding together the probability of all rows in Table 5.4 that have the same distribution.[7]

Compare Figures 5.3 and 5.5. When the sample size was two, the chance of getting a distribution in the sample that was near to the distribution in the population was ½ (or 50%). When the sample size is four (Figure 5.5), again, the only possible distribution in the sample that counts as "near" is the middle bar: half red and half black. One out of four and three out of four are too far from the distribution in the population. While the probability of getting a sample distribution that is near the population's distribution remains the most likely outcome, the probability of being "near" has actually gone down! But something else interesting and important has happened here: the chance of getting all red or all black is *much* lower with a sample of four than a sample of two. As the sample size increases, the likely sample distributions cluster together around the value of the distribution

[7] For instance, there are four rows on the table with three reds and one black. Since the probability of getting the distribution of three out of four reds is the probability of *either* reds on the first, second, or third draw and a black on the fourth *or* black on the first and reds on the rest, *or*... the simple disjunction rule for calculating probabilities applies: ($1/16 + 1/16 + 1/16 + 1/16 = 1/4 = .25$).

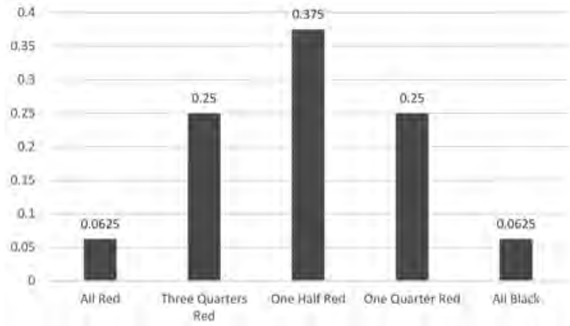


Figure 5.5: Sample of four

in the population.

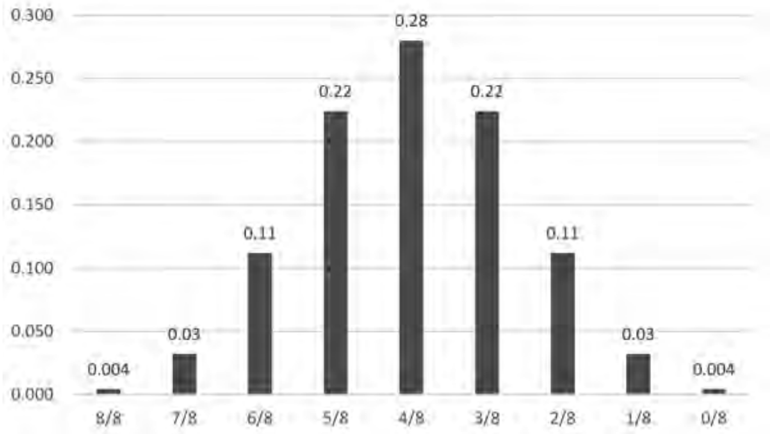


Figure 5.6: Sample of Eight

The pattern continues with larger sample sizes. Compare graphs of samples of eight (Figure 5.6) and sixteen (Figure 5.7) with the samples of two and four. With samples of these sizes, more than one possible sample distribution counts as "near" to the correct distribution. In the sample of eight, the middle three bars represent samples that either have the same distribution as the population (the middle bar), or one bead more or less. Hence, the middle three bars are all within ⅛ (or 12.5%) above or below the population's distribution, and in the sample of sixteen the middle five bars fall within the range. This means that there are three ways to get a result that is near the distribution in the population with a sample of four, and five ways with a sample of eight.

Since we have a distribution in the sample that is near the distribution in the population when we get any one of the similar alternatives, we can use the simple disjunction rule to calculate the probability of being near. In a sample of eight, the distribution of the sample will fall between ⅝ and ⅜—represented by the middle three bars—73 times out of 100 (or 73%). In the sample of sixteen, the samples

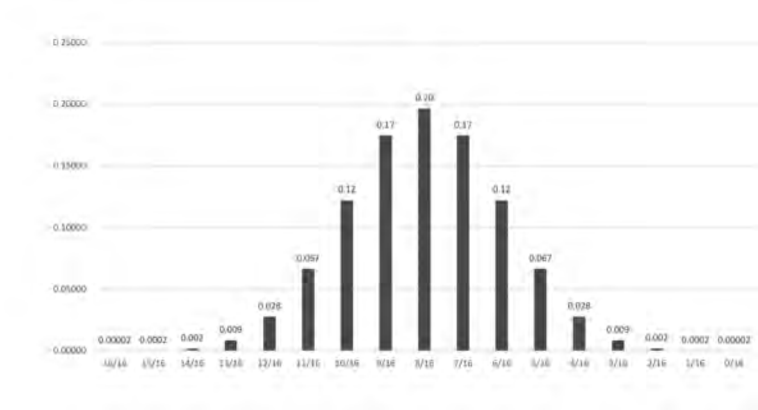


Figure 5.7: Sample of Sixteen

fall within that range 79 times out of 100 (79%). When the sample size reaches sixty, samples will be near 95 times out of 100 (95%), and when the sample size reaches 100, 99% of the samples have a distribution that is "near" the distribution in the population.

If the conclusion of an inductive argument, given the premises, was true 99% of the time, the argument would be very strong. So, we have shown that Argument 5.8 is very a strong inductive argument.

100 beads were sampled with replacement.

The distribution in the sample was within $\pm$12.5% of 50%

$\pm$12.5% is the range that counts as near.

The distribution in the population is near to 50%

Argument 5.8

We have worked through these four sample sizes in detail to show both the pattern and why it arises. As the sample size gets larger, the likely samples cluster closer to the value of the distribution in the population. The likely samples cluster because of the way that the possible distributions in the sample combine, and because the probability of drawing a bead with a particular color is the same as the distribution of that color in the population. We showed this by fixing what counts as "near" to the distribution in the population. However, given the way the calculations worked, we could have equally fixed a level of likelihood and shown that as sample size increases, the range of samples that are probable (at that level of likelihood) gets narrower. The likelihood of getting a sample with a particular distribution and the similarity between sample and population (nearness) are closely related concepts.

Contemporary scientists talk about the inter-related notions of likelihood and nearness in terms of a **confidence level** and a **confidence interval**.

Confidence Level: The probability that the distribution of a property in a population is within the *confidence interval* above or below the distribution of the property observed in the sample.

Confidence Interval: A range of possible distributions of the property of interest within which the distribution in the population is likely (at the probability of the confidence level) to fall, expressed as a proportion above or below, *e.g.* $\pm 3\%$

In contemporary scientific practice, we set the confidence level at high level—normally either 95% or 99%— and then choose a sample size large enough to determine an acceptably narrow confidence interval. This guarantees that we will have a strong inductive argument when the premises and presuppositions of the argument are true. Table 5.9 shows the approximate confidence intervals for different sample sizes at a 99% confidence level.

Sample Size	Confidence Interval
250	$\pm 8\%$
500	$\pm 6\%$
800	$\pm 4.5\%$

Table 5.9: Sample sizes and Confidence Intervals at 99% Confidence Level

This section has argued that there are some inductive generalizations where the premises provide strong—indeed, very strong—support for their conclusions. This does not mean that all inductive generalizations are strong, even when the sample sizes are large. The argument in this section made two crucial assumptions: that each draw of the sample was probabilistically independent of the others, and that the probability that a bead of a given color would be drawn was the same as the distribution of the color in the population. These assumptions may or may not be satisfied in inductive arguments used in realistic contexts. Hence, while the mathematics of sample size, confidence interval, and confidence level show how inductive generalizations work, we need criteria for evaluating such arguments when they are presented in scientific contexts.

5.4 Evaluating Inductive Arguments

Evaluating the strength of inductive arguments requires careful analysis of the inference in its context. Inductive arguments with similar forms can provide strong support for a conclusion in one context and weak support in another. Deductive validity, by contrast, is simply a matter of form, and context is irrelevant. This means that we need more elaborate criteria for judging the strength of inductive arguments than for judging validity, and that the application of these considerations is often a subtle matter.

The strength of an inductive argument that is based on a sample can be assessed by answering the following questions:

Confidence Interval: Is the sample size large enough to support an appropriate confidence interval?

Confidence Level: Does the sample size support a sufficiently high confidence level?

Independence: In the process of sampling, does the selection or observation of one element influence the selection or observation of others?

Representativeness: Is there any reason to think that the sample is unlike the population?

Each of these questions arises from a feature of the argument in Section 5.3. There we saw how larger sample sizes produce both narrower confidence intervals and increased confidence levels. The relationship between sample size, confidence interval ("nearness") and confidence level ("likelihood"), however, depended on two assumptions. First, each element of the sample (that is, each draw of a bead) was independent of the others. We satisfied this assumption by stipulating that the sampled beads were replaced and the population mixed. Second, we assumed that the probability of drawing a bead of a given color is the same as the distribution of the property in the population. This assumption would be violated if the beads were not well mixed, say, if all the reds were at the top of the bin. So we met this assumption by supposing that the bin was well mixed. These two assumptions constitute an ideal situation. The questions above try to determine how much the real-world context of an inductive argument is like the ideal. In a particular situation, the ideal may be more or less closely approximated. If the context of the argument is very like the ideal, then the argument is strong.

A well-designed study will have determined in advance the size of the sample necessary to support a particular confidence level and confidence interval. Determining whether a scientific study has appropriately used statistics is an important part of evaluating its arguments, but the mathematics of this sort of evaluation is outside of the scope of an introductory text. Rather, the first two questions, about confidence level and confidence interval, help us assess whether the results of a scientific study are reliable or useful in the contexts in which we will use the knowledge. Moreover, as readers of scientific studies, we must bear in mind that when a study is presented—especially if the presentation comes through secondary sources like a web page or a newspaper—the conclusions can be distorted. It is common for press reports to present scientific conclusions in stronger terms than are justified by the actual study. As consumers of science, it is important for us to look for information about the sample size and use it to assess the strength of the arguments presented.

The question about confidence interval asks whether the confidence interval is "appropriate." A inductive generalization with a small sample can provide very strong support for large confidence intervals. The problem with such a study is that its conclusions will often be uninformative. We can have 100% confidence that a distribution of a draw of cards falls between "all red" and "all black,', but this conclusion is useless. The question of whether the confidence interval is appropriate, then, depends on how much precision we need. Where a property is very rare, or where we need to make fine discriminations in our data, we need large samples. In contemporary physics of subatomic particles, experiments must run millions of trials because the events of interest are so rare. By contrast, if we expect the population to be uniform, or the property to be common, a small sample will do. To take another example from physics: Kepler inductively inferred his laws of planetary motion from very precise measurements of the relative positions of the planets during a relatively short period of time. These observation were not repeated many times over many years. Planetary motion is extremely regular, so a small sample of measurements was sufficient to support the generalizations.

The question about confidence level asks whether the conclusion is sufficiently likely for the purposes at hand. It is common for well-designed scientific studies to rigorously argue for a conclusion at a 95% confidence level. As a scientific matter, this result may be well-established, but there is a further question that arises from the application of this knowledge. If we are to use that study as the basis for engineering a bridge or recommending a medical treatment, is 95% high enough? In general, as the risks of an application rise, we demand more assurance of our conclusions. If little is at stake, then we can accept weaker arguments. Therefore, an argument that might be accepted as strong enough in one context might be rejected as too weak in another. The question of confidence level, then, asks us to assess the confidence level supported by the sample size to see whether it is appropriate for the context in which the argument is being used.

The third question asks about independence. Probabilistic independence was defined above as:

$$\mathsf{Pr}(A \mid B) = \mathsf{Pr}(A)$$

Again, this says that the probability of A given B is the same as the probability of A alone. In other words, the occurrence or non-occurrence of B has no influence on the probability of A. In the examples of drawing cards from a deck or beads from a bin, this independence was guaranteed by replacing the card or bead, and then shuffling or mixing. If the element was not replaced, then the population changed, thus influencing the next draw. Note that the crucial assumption here is not replacement, but the independence that it produces in that context. In many scientific studies, the population is so large that replacement becomes irrelevant: removing one grain of sand from a beach changes the distribution (of any property) in the population to such a small degree that there is no significant difference between replacing and not replacing it. And some observations can be made without disturbing the population—such as the measurement of planetary movements—so that replacement is not meaningful. However, there are many other ways in which independence can be disrupted and an inductive argument weakened.

When drawing a sample from a population, we often have to interact with the population in a significant way. In such cases, the process of observing one element of the sample can change the probability associated with the next observation. In studies of animal behavior, for example, the animals can change their behavior because of the presence of humans. Animals may become accustomed to humans, so that later observations of the animals are the result of a kind of training based on the earlier observations. Studies of humans are notorious for such problems, since our study subjects can talk to one another and thereby let earlier observations influence later ones. And in the design of questionnaires, the order of the questions matter because answering one question might change the subject's thinking in a way that changes their answers to others.

Choices by the investigator in selecting objects to observe can also disrupt the independence of observations. We are often hopeful that our investigation will turn out one way or another. If a physician was successful with a medicine in one patient, then tried it on a patients who were similar to the first (same occupation, age, gender, etc.), he or she would not be in a position to draw conclusions about all patients. The physician in a case like this may not be aware that he or she is influencing the sampling process and weakening the argument. For this reason, scientific studies carefully design the procedure by which the observations will be made. These procedures may involve things like "blinding" the physician to whether a medicine or placebo is being administered to a particular patient. They may also involve randomization (devices like coin flips or dice rolls) in the selection of patients to be observed. Designing and following procedures for sampling lets us think through and avoid possible sources of error—another significant way in which scientific observations are *systematic*.

The final question, about representativeness, looks for systematic differences between the sample and the population. Clearly, if there is reason to believe that the sample is unlike the population, then any inductive support for the conclusion will be significantly weakened. Unrepresentative samples arise when there is some kind of difference among members of the population that is relevant to the property under investigation. In our imaginary bin of beads, the sample would have been unrepresentative if all the red beads had been at the top where we could reach them. A more realistic example comes from in mid-twentieth century medical studies of heart disease. The subjects observed were all male, yet the conclusions concerned all humans. We later learned that heart disease in women is importantly different from heart disease in men. The earlier arguments were thus significantly weakened because the original samples were unrepresentative. A similar question arises in the case of the Grizzly Bears (Section 4.2, above). The bears in a particular region were observed to be eating moth larvae, but animal behavior varies in different environments. As a result, the conclusion of this argument was *not* about all Grizzly Bears (much less all bears or all animals), but about a more specific population. If the scientists had tried to draw a conclusion about all Grizzly Bears, their sample would have been unrepresentative.

Procedures for making sure that the sample is representative often overlap with procedures for guaranteeing independence of the observations. The mixing of a bin of beads ensures that earlier draws did not influence later draws—because we were not drawing the same bead over and over again—and also guaranteed that the probability of drawing a bead of a particular color was the same as the distribution of that color in the population. Mixing a bin and shuffling a deck of cards are ways of *randomizing* the selection of the sample. Randomization is useful, but it is not an automatic guarantee of either independence or representativeness. Moreover, it is not necessary. Picking every tenth name from a list of names is not random in the way that flipping a coin is random, but in some circumstances it might give us an unbiased sample.

One of the lessons of Chapter 4 was that determining the strength of an inductive argument relies heavily on our background knowledge about the objects being studied. We are now in a position to understand how and why this is the case. Assessing representativeness requires attending to the variety of properties found in the population and how they influence each other, just as the gender of the patient influences his or her heart disease. Assessing independence requires thinking through how our observation or sampling procedure will influence the properties which we are investigating,just as observing animal behavior might influence that behavior. Deciding whether the confidence interval is appropriate depends on our expectations of how common the property will be in the population. And determining the right confidence level depends on what the conclusions will be used for. The inductive strength of a particular argument, then, can only be assessed by bringing to bear what we already know about the population, the properties of interest, and the procedures by which we will make observations. Again, this is an important difference between the assessment of inductive strength and the assessment of deductive validity. It is also a way in which scientific knowledge is *fallible*. If our background assumptions are incorrect, we may mistakenly take a weak argument to be strong. We will return to the idea of fallibility and its consequences for how we think about science in Section 8.4.

5.5 Conclusion: Strong Inductions, Mangoes, and Rice

We are now in a position to evaluate the two arguments discussed by Dharmakirti, Arguments 4.1 and 4.2. Dharmakirti rejected these arguments, but how do they fare by the criteria we have established in this chapter?

When evaluating the argument about the mangoes, we need to use our background knowledge of mangoes and how they grow. Mangoes ripen at different rates, and their ripeness depends on the fruit's position on the tree and other factors. Moreover, at the early stage of ripening, there is much variation of mangoes on a single tree; it is very unlikely that they would all be ripe at the same time. This means that early in the mango season, when the mangoes are just coming ripe, Argument 4.1 would be very weak—here we agree with Dharmakirti's assessment. Late in the season, the argument would be stronger, though the argument is still not very strong. After all, part of the tree might have grown in the shade, while the rest of the tree is in the sun.

Consider a similar argument in a different context. Suppose we had a large grove of mango trees. We want to send the workers in to harvest only if more than half of them are ripe. We draw our sample from all parts of the orchard, and we are careful to sample from both the sunny and shady sides of the tree. Imagine that we use a randomization process (say, rolling a dice) to determine which mangoes are sampled. If we sampled sixty mangoes and found that forty of them were ripe, we would have a 95% confidence level that more than half were ripe. This would be a very strong inductive argument for the conclusion that more than half the mangoes in the orchard are ripe. Therefore, while we can agree that Argument 4.1 is weak, there are similar arguments that are quite strong.

The argument concerning rice is slightly different. In the experience of the authors of this text, testing one spoonful of rice is a quite reliable way of determining whether the whole pot is cooked. When rice is purchased from a commercial producer, as it is today in many places, the grains are extremely uniform. Moreover, the temperature of the water when it is boiling is also very uniform—the physics of phase transitions (changing from liquid to gas) guarantee that there are not big differences in temperature at the top or bottom. These two facts mean that the rice cooks in a uniform way. It is extremely unusual for some of the grains to be cooked while others are uncooked. Moreover, the boiling process mixes the grains, and my spoonful is taken at random from them. Under these conditions the small sample size does not weaken the inference, and the sampling is representative and independent. Therefore, Argument 4.2 is a strong argument.

Now, Argument 4.2 may *not* have been strong in Dharmakirti's time. While we're speculating, it could well be that Dharmakirti's rice was much less uniform, either in size or dryness, than the rice we cook. If so, then it would have been the case that, like the ripening mangoes, rice grains cooked at different rates. Having all the rice be cooked at the same time would have been very unusual, and hence a poor conclusion to draw from such a small sample. If this is plausible, then Argument 4.2 is an excellent example of an argument that is good in some contexts but not in others. Moreover, like the mango argument, there are ways to strengthen it. The sample size might be increased. If there is variation in the whole pot, a larger sample size is more likely to capture that variation. And like the test of the mango orchard, a stronger argument would be produced by testing for a distribution, rather than an all-or-nothing conclusion about whether the rice is cooked.

Contemporary scientific practice thus both agrees and disagrees with Dharmakirti about Arguments 4.1 and 4.2. In the contexts that Dharmakirti was probably considering, these arguments provide weak grounds for their conclusions. However, there are conditions under which they would be stronger, and there are very similar arguments that are very strong. This chapter has outlined four criteria by which the strength of inductive arguments can be assessed. Applying these criteria lets us distinguish between strong and weak arguments in ways that Dharmakirti did not consider, but of which we hope he would have approved.

Chapter 6

Correlations and Causes

6.1 Introduction: The Problem of Causal Inference

In both Buddhism and science, causality plays a central role in understanding of the everyday world around us as well as deeper truths. Buddhist epistemology and scientific methodology thus share an interest in the question of how we can come to know causes. The fundamental epistemological challenge of causality is that causal relations are never on the surface. We can notice events occurring, and we can notice that when an event of one type occurs, an event of another type occurs. Noticing such patterns merely leaves us with a question: do the earlier events cause the later events? Perhaps one notices that one feels sleepy many late summer afternoons (the event of coming to feel sleepy) and that rain storms commonly move in shortly thereafter (the event of rain onset). Does one of these events cause the other? Or are they commonly occurring together for reasons that do not involve a causal relationship between them? Could they each be effects of some other causes? If the one causes the other, which does what? There is a lot to sort out.

Consider a more scientific example. Suppose my doctor tells me to change my diet. I do so, and some days later I find myself feeling better. Did I get better *because* of the dietary change, or for some other reason? Anyone can see that I've changed my diet, and I can tell that I feel better. But the causal relationship is not on the surface. The true cause could have been any one of a number of other factors: a change in my exercise routine, a change in weather, or a disease having run its course.

When we say that the change in diet made me feel better we are saying more than that I changed my diet, and later I felt better. We are also saying that if I *had not* changed my diet, I *would not* have felt better. In addition, there is an implicit generalization: if others (who were in a situation like mine) changed their diet, they too would feel better. Causes are "hidden" in the sense that these stronger commitments cannot be read off of the simple evidence that I changed my diet and then got better. The fundamental epistemological issue about causality, then, is this: what observations provide convincing evidence for causal relationships? Using the framework of inductive inference discussed in Chapters 4 and 5, we can put the question in terms of arguments: are there strong inductive arguments from observational premises to causal conclusions? If so, what are their characteristics and how do we distinguish strong from weak inductive arguments with causal conclusions?

This chapter will begin addressing these questions. There is a strong consensus among Western philosophers and scientists that the best evidence for causal claims is the repeated association between the purported cause and effect. All by itself, the fact that I changed my diet and then felt better is very weak grounds for the claim that I felt better *because* of the dietary change. My doctor, presumably, has seen this pattern in many patients. What makes her evidence better and argument stronger? In the last chapter, we identified criteria that permitted us to distinguish strong from weak inductive arguments of a particular form: inductive generalizations. In this chapter, we will show how the criteria for strong inductive generalizations apply to causal inferences as well.

6.2 Correlations

When the doctor noticed the pattern that patients who changed their diet felt better, she noticed what we will call a "correlation." She noticed that patients (in particular circumstances, perhaps with a specific diagnosis) who changed their diet were more likely to feel better than those who did not. The scientists that we discussed in Chapter 4 noticed similar patterns. The scientists who studied Grizzly Bears noticed a correlation between the places where Grizzly Bears were found and insect larvae under the rocks. Lind discovered a correlation between drinking lemon juice and having the symptoms of scurvy disappear. In both of these cases, a correlation formed the primary premise for a causal argument. To understand the epistemology of causation, then, we must understand the concept of a correlation.

6.2.1 Identifying Correlations

In all of the examples above, the scientists observed a relationship: a relationship between diet and health, between the location of Grizzly Bears and insects, between lemon juice and scurvy symptoms. These are observable properties. A correlation is a pattern in the occurrence of two properties–the occurrence of the one is associated with the occurrence of the other. The pattern observed by the scientists in the above examples can be expressed in terms of likelihood (or probability). The presence of one property makes the other more or less likely (probable) than it otherwise would be. Sailors who drank lemon juice were much more likely to recover from scurvy than those who did not; Grizzly Bears were much more likely to be found near moth larvae than elsewhere. We can provide a formal definition of correlation as follows:

Correlation: Two properties A and B are correlated if and only if the probability of A occurring given that B has occurred is either higher or lower than the probability that A would occur in the absence of B.

Using the more precise mathematical representation developed in Chapter 5, we can define a correlation as two properties, A and B related in this way, where $\neg B$ means that B is absent or is not the case, and $\neq$ means that the two quantities are not equal (one is greater than the other):[1]

$$\mathsf{Pr}(A \mid B) \neq \mathsf{Pr}(A \mid \neg B)$$

To precisely illustrate the definition of a correlation, let us consider a simplified example. Imagine a village with 100 children who are the appropriate age to go to the village school. 50 of these children are boys and 50 are girls. However, not all of the children in the village go to school; let us suppose that only 80 of the village's school-age children are enrolled. 45 of the enrolled children are boys, which means that 5 of the boys do not attend. The remaining 35 school children are girls, which means that 15 do not attend. Now, is there a correlation between the property of being a boy and the property of going to school?

To apply the definition of a correlation to this example, we have to determine what properties fill in for A and B. Let us stipulate that A stands for the property of attending school, and that B stands for the property of being a boy. This means that to determine whether there is a correlation, we need to know:

1. the probability of a child attending school, given that the child is a boy, and
2. the probability of a child attending school, given that the child is not a boy

According to the definition, there is a correlation if and only if these to probabilities are not equal. This would mean that boys are more (or less) likely to go to school than the girls. If boys and girls were equally likely to go to school, then there would be no correlation.

In Chapter 5, we explored the relationship between distributions and probabilities. Recall the example of the large bin of red and black beads, where half of the beads were black. The *distribution* of black beads in the population was ½ (or 50%). This means that when a bead is chosen randomly from the bin, the probability of choosing a black bead is ½ (or 50%). Similarly, in our imaginary village, the probability that a child is a boy, given that he goes to school, is the probability of randomly choosing a child who attends school from among the village boys. In other words, to know the probability of attending school, given that a child is a boy, we only have to determine the distribution of the property "attends school" in the population of boys. That is, we treat the boys of the village as the population of interest and look for the distribution of school children in this population.

[1] Astute readers will notice the relationship between this definition and the definition of probabilistic independence given on page 47, above. If two properties are correlated, they are not independent, and if they are independent, they are not correlated.

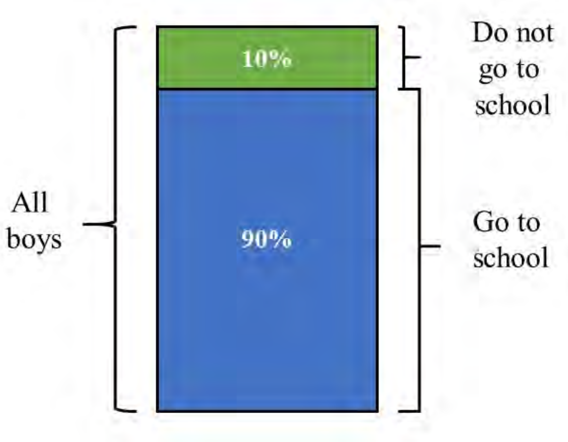


Figure 6.1: Distribution of boys in school

Given what we have stipulated about the village, the distribution of school children from among the boys is easy to work out. Of the 50 boys in the village, 45 go to school. Expressed as a proportion, this is $^{45}/_{50}$ or $^{9}/_{10}$. Expressed as a percentage, this equals 90%. This distribution is a probability so the probability that a child attends school *given that* the child is a boy is 0.9. Figure 6.1 lets us visualize this distribution.

Now consider all those children who are not boys (that is, the girls). This is another population, and wthin it there is again a distribution of childeren who do and do not go to school. of the 50 girls in the village, 35 attend school. The proportion of school atendees among the girls is therefore $^{35}/_{50}$ or $^{7}/_{10}$. Expressed as a percentage, this is 70%. The distribution among girls is also a probability: the probability that a child goes to school *given that* she is a girl (not a boy) is .7.

According to the definition of a correlation, the properties of being a boy and attending school are correlated if and only if:

$$\mathsf{Pr}(\text{In School} \mid \text{Boy}) \neq \mathsf{Pr}(\text{In School} \mid \text{not a Boy})$$

As we have seen, these probabilities are not the same. In particular, the probability of being in school, given that the child is a boy is .9 (or $^{9}/_{10}$), and this is greater than .7 (or $^{7}/_{10}$), which is the probability of being in school, given that the child is a girl. Therefore, the property of being a boy and the property of being in school are correlated. We can visualize this easily if we compare the two distributions, as illustrated in Figure 6.2.

The correlation between being a boy and being in school is a **positive correlation**: boys in this example are *more likely* than girls to go to school. In general, two properties are positively correlated when the presence of one property makes the other property *more likely* than it would otherwise be. Notice that the definition permits the probability to be lower as well. These are **negative correlations**: the presence of one property makes the other less likely. Lind noticed a negative

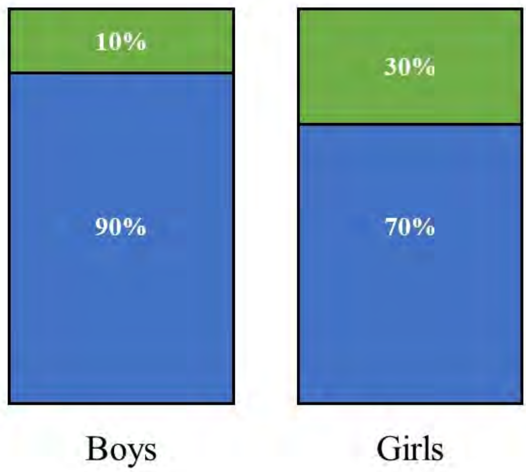


Figure 6.2: A positive correlation between being a boy and being in school

correlation between lemon juice and the symptoms of scurvy. Sailors who drank lemon juice were *less likely* to have the symptoms of scurvy. Similarly, vaccines are drugs that prevent the occurrence of diseases like measles. Once one has been vaccinated for measles, it is unlikely that one will contract it; there is a negative correlation between vaccination and contracting measles.

There are different kinds of properties used to identify correlations. Being enrolled in school is an all-or-nothing affair; one either has filled out the forms and paid the fees or not. By contrast, a property like speaking Tibetan is a matter of degree; one can speak just a little Tibetan, quite a bit, or be fluent. Many of the correlations interesting in science use properties that are measured in degrees. While we will discuss these, the mathematical details are beyond the scope of this book. To keep matters simple, when we give detailed examples, we will choose properties that are all-or-nothing, or a least can be treated as such without too much distortion.

Correlations are a relationship between probabilities, and this has two important consequences. First, probabilities can be high or low, and this gives us a way to measure the strength of a correlation. If there is a large difference between $\mathsf{Pr}(A \mid B)$ and $\mathsf{Pr}(A \mid \neg B)$, then the correlation is a **strong correlation**. On the other hand, if there is a small difference in probabilities, it is a **weak correlation**. The notions of strong and weak correlation are not precise: we have not said just how big the difference has to be before a correlation is strong. These ideas admit of degrees, and their most important usage is comparative. Imagine, for example, a second village. Here the probability of going to school, given that one is a boy is $^{51}/_{100}$, while the probability of going to school given that one is a girl is $^{50}/_{100}$. This is a very small difference in probability: boys are not much more likely to go to school than girls. In this village, the correlation between being a boy and going to school is much weaker than in the first village. Strength and weakness have limits, and we give these limits names. If *all* of the children in school were boys, there would be a **perfect correlation**. On the other hand, if the proportion of boys and girls in school were the same, then there would be **no correlation**. Figure 6.3 illustrates the different kinds of correlations.

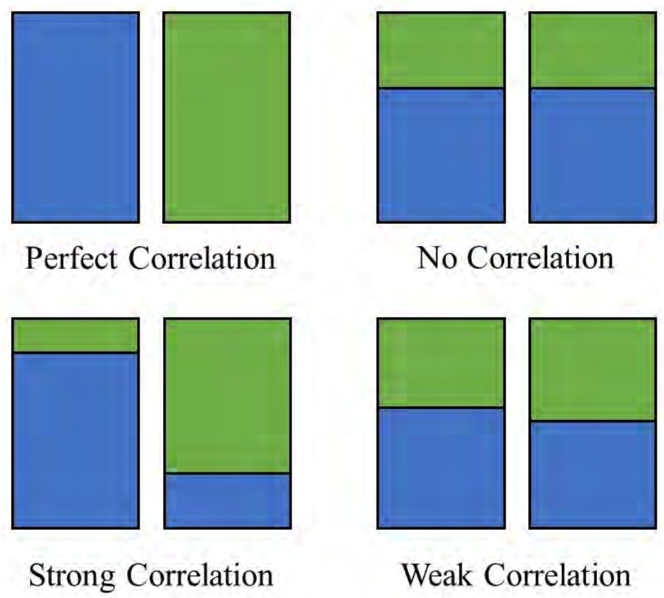


Figure 6.3: Varieties of Correlation

The second consequence of correlations being a relationship among probabilities has to do with our knowledge of correlations. In the example of a village, the number of children was so small that we could determine the correlation by simply counting the children. Suppose the number of children in the population were to large for simple counting to be practical. For example, we might want to know whether there were a world-wide correlation between gender and education. In such a case, we would have to sample and make an inductive generalization from the sample to the population. Discovering correlations is different, however, from the examples we discussed in Chapter 5. In the previous examples, there was only one distribution in the population to be discovered. Now there are two. We need to know what proportion of boys go to school, and we need to know what proportion of girls go to school. So, we will have to sample the boys and determine the distribution of school attendance, and we have to sample the girls to see their distribution of school attendance. In general, to discover whether there is a correlation between properties A and B, we only need to construct two inductive generalization arguments: one for the distribution of property A in the population of B, and another for the distribution of property A in the population of things that are *not* B (what we have represented as $\neg B$).

6.2.2 Evaluating Arguments for Correlations

Let us return to the doctor who has seen many of her patients feel better after a change in diet. Has she noticed a correlation in the population of patients? A common mistake is to conclude that there is a correlation on the basis of a high proportion of patients who get better after a change in diet. This would be like noticing that most boys go to school and concluding that there was a correlation between being a boy and school attendance. If we did not know what proportion of girls go to school, we would not know whether there was a correlation. Therefore, the physician also has to consider whether those who did *not* change their diet felt better. There are, then, two inductive arguments involved in establishing a correlation in a population. Since each argument is an inductive generalization, we can use the criteria developed in Chapter 5 to evaluate whether the physician is entitled to the conclusion that there is a correlation between change in diet and feeling better.

Suppose our physician does the following measurements. She looks back over her records and sees that she has had 20 patients with similar symptoms. 10 changed their diet, while 10 did not. Each of these groups of 10 is a sample of the larger population. Within each group she needs to know how many in each group felt better. Suppose her records show that 8 of those who changed their diet felt better, while only 4 of those who did not change felt better.

She now has the premises of two inductive generalizations, represented in Arguments 6.4 and 6.5.

Among patients who changed their diet, 80% diet felt better.

Therefore, 80 % of all patients who change their diet feel better.

Argument 6.4

Among patients who did not change their diet, 40 % felt better.

Therefore, 40 % of all patients who do not change their diet feel better.

Argument 6.5

If we accept the conclusion of these arguments, the physician would indeed have found a correlation in the population of patients. Changing diet would raise the probability of feeling better. But are these arguments strong? To evaluate them, we would have to apply the criteria developed in Chapter 5:

Confidence Interval: Is the sample size large enough to support an appropriate confidence interval?

Confidence Level: Does the sample size support a sufficiently high confidence level?

Independence: In the process of sampling, does the selection or observation of one element influence the selection or observation of others?

Representativeness: Is there any reason to think that the sample is unlike the population?

The first two criteria take on a special significance when we are evaluating arguments for a correlation. A correlation is a *difference* in probabilities. This means that the confidence interval and confidence level have to be sufficient to show that there is a difference, not just in the sample but in the larger population. As we saw in Chapter 5, confidence interval and confidence level are determined by the sample size. You will recall that a confidence interval is a range of possible distributions (see p. 5.3), and the real value is likely (at the probability of the confidence level) to fall within this range. The confidence interval is often represented in a scientific graph by an "error bar." This shows that the real value could be higher or lower, and by how much. Figure 6.6 illustrates the distributions and error bars for our physician's data.

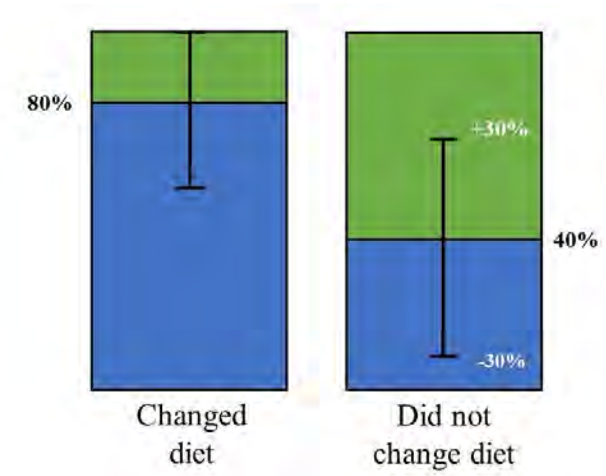


Figure 6.6: Distributions with error bars for two samples of 10 patients

Science seeks to identify and eliminate mistaken inferences. When evaluating an argument for a correlation, we need to be sure that the observed difference in probability did not arise by chance. The difference between the physician's two samples is quite large: the difference between 40% and 80%. But what is the chance that there is no correlation in the population? To address this concern, we look at the confidence intervals for the two samples. If the confidence interval for each observed distribution is so broad as to include the other, then there is a significant chance that the distribution of the property in the two populations is *the same*, in which case there is no correlation.

When we look at the physician's sample size—10 in each group—we see that it is rather small. While the calculations are beyond the scope of this book, a sample size of 10 and a confidence level of 95% yields a confidence interval of approximately ±30%. That means that 95% of the time, the proportion of those who felt better from among those who changed their diet could be as low as 50% (80 - 30). And the proportion of those who felt better without changing their diet could be as high as 70%. The chance that the two distributions are really the same, then, lies within the confidence interval. In Figure 6.7, the area of overlap is represented by the red rectangle. In other words, the possibility that there is no

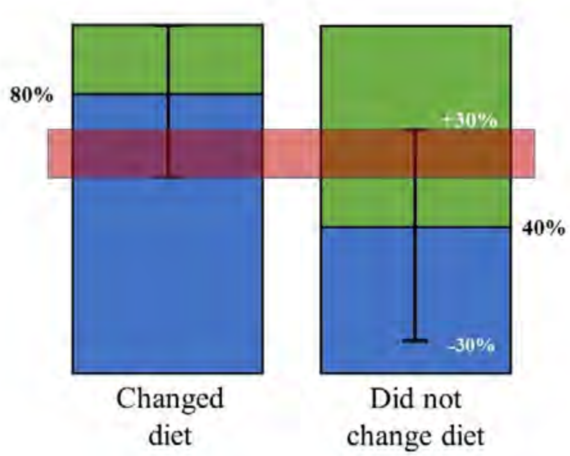


Figure 6.7: The possibility of no correlation falls within the confidence intervals

correlation lies within the range that we would expect 95% of the time. So, we have not ruled out the possibility that diet and feeling better are not correlated. Under these circumstances, therefore, we have a weak argument for the existence of a correlation in the larger population. The conclusion of the foregoing argument should be a bit surprising. After all, twice as many patients felt better once they changed their diet. This shows that reasoning with a 95% confidence level is quite demanding, indeed, it is more demanding than much ordinary reasoning about causes.

The discussion of sample size in Section 5.3 shows us what to do in order to make the argument stronger. The confidence interval depends on the sample size: as the sample size gets larger, the confidence interval gets narrower. If the physician could double the sample sizes of her two groups to 20, the confidence interval would shrink to approximately ±20%. This sample size is still too small: the proportion of those who felt better from among those who changed their diet could be as low as 60%, and the proportion among those who did not could be as high as 60%. The physician who found distributions of 80% and 40% in her samples would need each group to be about 30 patients before we could conclude with 95% confidence that she had identified a correlation. Figure 6.8 shows the distributions with error bars. Notice that they no longer overlap, as indicated by the green rectangle. The larger sample has reduced the confidence intervals.

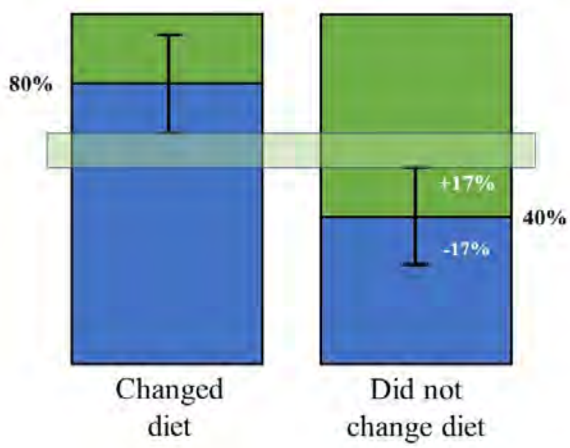


Figure 6.8: Distributions with error bars for two samples of 30 patients

Even if the arguments could be strengthened by increasing the sample, there are two more evaluative criteria to consider. The criteria of independence and representativeness also provide grounds for criticizing our imaginary physician's arguments. We imagined that she observed her own patients. This means that the observations were not entirely independent—there may be commonalities in their treatment that make the observations dependent on each other. Moreover, there may be reasons to think that her patients are different from the rest of the population, and hence that the sample was not representative. Perhaps they are wealthy enough to afford a physician, hence have different living conditions and better overall health than the rest of the population. The character of the observations over and above sample size, then, are further grounds for considering the argument weak. The argument would be strengthened by finding a way to eliminate these biases. For example, the physician might have chosen names randomly from a telephone listing or tax roll; she might have tried to make sure that her sample included people with characteristics known to affect health; young and old, men and women, vegetarians and non-vegetarians, and so on. Along with an increase in sample size, had our physician been sure to make the observations independent and the sample representative, then she would have had strong evidence for a correlation between diet and health.

While our physician has been imaginary, she has taught us some important lessons. First, identifying a correlation depends on a pair of inductive arguments with two samples. A correlation between properties A and B requires that $\mathsf{Pr}(A \mid B) \neq \mathsf{Pr}(A \mid \neg B)$, so we must sample from a population that has property B and population that lacks it; in each we must identify the proportion that exhibit property A. Second, while our physician's results looked quite dramatic, a closer look showed that the sample size was too small to support a strong argument. And there were some problems with the way in which the sample was collected. This second point shows an important aspect of the *systematic* character of scientific observation. Observations must be made so as to build strong inductive arguments. One dramatic result is not enough to demonstrate a correlation, much less a causal relationship. My own experience of changing my diet and then feeling better is completely insufficient. I am a sample of 1 person, hence the sample is far too small. Moreover, my positive experience shows only half of what needs to be demonstrated: we do not know the probability of feeling better without a change. Scientific evidence must be gathered more systematically than such casual observations about health will allow.

6.3 Indicators of Causation

While identifying correlations is a crucial part of the epistemology of causation, it cannot be the whole story. There must be more to causation than correlation. The problem is that there are many cases where A is correlated with B in a sample or population, but A is clearly not the cause of B. To make this claim however, is to presuppose some prior conception of causation. Indeed, it is impossible to discuss causal arguments without discussing the metaphysics of causation. After all, we cannot judge whether an argument is strong without knowing exactly what we are arguing *for*. This means that before we can fully articulate how inductive arguments for causal conclusions work, we need to establish the characteristics of causality.

Just as in Buddhism, the fundamental characteristics of causality have been subject to much debate in Western philosophy. Adhering to this book's goal of presenting a consensus view (insofar as possible), we will not dive into this debate. Rather, we will look first to scientific practice. Implicit in the activities of experimental design and data interpretation is a conception of how causes differ from mere correlations or other sorts of regularity. Different philosophical accounts of causality take this conception of causality and subject it to philosophical interpretation and analysis. The metaphysical disputes can be seen as different ways to interpret the notion of cause implicit in science, and they therefore take the consensus characteristics of causality as a starting point. In the next chapter, we will explore this idea of causality in more detail, what we will call the *simple* conception of causation. In this section, we remain concerned with the evidence for causal conclusions, and therefore with indicators of causation.

Correlations alone are poor evidence for causal conclusions because correlations can arise through processes that have nothing to do with the correlated variables. In the absence of a causal relationship, there are two primary ways in which a correlation between two variables can arise in a sample or in a population: chance and common cause.

What we commonly call "chance" involves events arising through causal processes that are complex and sensitive to a variety of factors. As a result, they are difficult, even impossible, to predict and they create patterns with random distributions. In each flip of a coin, for example, there are very specific details about the its acceleration, the turbulence of the air around it, its spin, its angle of impact, the characteristics of the surface it strikes, and so on that determine whether the individual coin flip resulted in heads or tails. We use a coin toss as a randomizing device because, in general,[2] these micro-causes are difficult to predict and they result in a random distribution of heads and tails. We say that the coin lands heads "by chance." More precisely, such processes are said to be *stochastic*: there is no pattern or regularity to a sequence of coin tosses. Over a long sequence of tosses, heads and tails will come up approximately the same number of times. Heads and tails can be assigned a probability (each ½), but we cannot predict that an individual coin flip will come up heads. Determination by stochastic processes is not limited to coin tosses or dice rolls. Many processes in nature are similarly sensitive to a variety of micro processes: human health and the weather are familiar examples.

Stochastic processes can give rise to short run patterns. If you flip a coin a large number of times, it is very likely that you will see a run where the coin lands heads several times in a row. We attribute this to "chance." In the example of Section 6.2.2 we supposed that the physican found that 8 out of 10 (80%) of those who changed their diet felt better, while only 4 out of 10 (40%) of those who did not change felt better. Like a "lucky" run of heads, it is possible that the patients felt better for a variety of reasons, each complex and individual, yet a short run pattern emerged. The correlation in the sample could have arisen from chance. As the sample size grows, the likelihood that the correlation in the sample arose by chance is reduced. But notice that it is never eliminated. With 30 patients in each group, there will be correlation in the population 95% of the time. Even with 95% confidence that the two variables are correlated in the population, there remains a 5% chance that they are not. If we are to infer from a correlation that a change in diet causes patients to feel better, we need evidence in addition to the correlation in a sample.

The second way in which a correlation might appear in a sample or a population concerns common cause. Consider this example. In southern India, wearing crimson monks' robes is very highly correlated with speaking Tibetan. If I see a monk on the streets of Mysore, I can be quite confident that he speaks Tibetan.

[2] The variables can be controlled, of course, and some coin tricks depend on the magician's ability to determine the path of the coin.

It is not certain, of course, but the probability is high. Note that this correlation is not just present in a sample of monks; the variables of "speaking Tibetan" and "wearing crimson monk's robes" is correlated in the population of India. In spite of the correlation, wearing monks' robes does not *cause* a person to speak Tibetan. Were I to don them, I would not find myself suddenly able to converse with Tibetan monks.

The correlation between wearing monk's robes and speaking Tibetan arises through what we call a "common cause." There is a reason why most who wear robes do so. Today, most Buddhist monks in south India have been raised in Tibetan communities. They learned their language at home and at school. Being raised in a Tibetan community also makes it more likely or more probable than it would be otherwise that a person will become a monk. There is, then, a single underlying cause—a common cause—of both properties.

When we say a correlation arose from chance or a common cause, we are explaining why the variables are correlated. We can think of causal inferences, then, as looking for the best explanation of a correlation in the sample or in the population. To find the best explanation, we must rule out alternatives, such as chance and common cause. When we have ruled out all of the alternatives for a correlation between A and B except the claim that A causes B, then we are in a position to make a strong causal inference. The simple concept of causation highlights two sources of evidence that help to narrow the range of possible explanations.

The first source of evidence concerns time. It is a commonplace about causality that causes precede their effects. So if A is to be a cause of B, then A must occur before B. Notice that there is nothing temporal about a correlation; it is simply a relationship of probabilities or distributions, and neither property need occur before the other.[3] For example, smoke is highly correlated with fire. When we see smoke, it is very likely, if not certain, that fire is present. On the other hand, in the absence of smoke, fire is unlikely. This means that there is a correlation: the probability that there is fire, given that there is smoke is greater than the probability that there is fire without smoke. Formally, we can express this as: $\mathsf{Pr}(\text{fire} \mid \text{smoke}) \neq \mathsf{Pr}(\text{fire} \mid \neg\text{smoke})$. But smoke is not a good candidate for the fire's cause because smoke happens *after* the fire has started. Therefore, even though fire is more probable in the presence smoke than it is in the absence of smoke, it would be temporally backwards to consider smoke the cause of fire. The evidence that smoke happens after the fire and not before is evidence gives us reason to exclude the possibility that smoke causes fire from our investigation.

The temporal asymmetry of causation can also provide evidence that block inferences to causal conclusions where the correlation arises from a common cause. In the case of wearing monk's robes and speaking Tibetan, it is the case that children speak Tibetan before entering a entering a monastery and wearing robes.

[3]Indeed, it can be demonstrated that correlations are symmetrical. If A is more likely in the presence of B than without B, that is $\mathsf{Pr}(A \mid B) \neq \mathsf{Pr}(A \mid \neg B)$, then B is more likely in the presence of A than without A, that is $\mathsf{Pr}(B \mid A) \neq \mathsf{Pr}(B \mid \neg A)$.

Hence, that wearing robes might cause one to speak Tibean is ruled out because the ability to speak comes first. While temporal asymmetry provides some evidence, correlation and temporal asymmetry are not enough. After all, speaking Tibetan precedes wearing monk's robes, but we would not want to conclude that speaking Tibetan is the cause of wearing monk's robes.

The second source of evidence arises from the fact that causal relationships involve a kind of dependence that is stronger than mere correlation. A commonsense point about causality is that if A is the cause of B, introducing A will bring B about. Striking a match causes its ignition; so if I want to light the match, I should strike it. Of course, this kind of causal sufficiency is not the only kind of dependency we expect from causes. Fire also requires fuel. And removing the fuel will stop the fire. Causes are typically combinations of such necessary and sufficient causal conditions: in the presence of fuel and oxygen, the spark will light the fire. Once the fire has started, removing the fuel or oxygen will stop it. The spark, the fuel, and the oxygen are together the causes of fire. With these factors, the fire will start; without them, the fire will stop (or fail to start). This is the sort of dependency between cause and effect that we expect. For the same reason, we can determine that there is no causal relationship between speaking Tibetan and wearing monk's robes. One does not gain or lose the ability to speak Tibetan by donning and removing monks' robes, nor does teaching an individual Tibetan make them put on the robes.

Like the identification of chance correlations, identifying correlations that arise from common causes is an important part of scientific practice. It is particularly important in health research. Consider our imaginary physician from Section 6.2. Suppose the dietary change she recommended was to become a vegetarian. Now, people who can adopt a particular diet and stick with it may have a number of dispositions. For instance, they may have a greater concern about their health and stronger self-discipline. These dispositions might well have other effects that lead to better health. Such people might be more likely to exercise, or less likely to smoke, for example. If so, then the change in diet may not have brought about the change in health; it might be that the underlying behavioral disposition supported both the successful change in diet and the change in health. One way to rule out such common causes, is to conduct an experiment where the one potential cause, and it alone, is manipulated. We will discuss experimentation further in Section 6.4.

So far, we have seen how two elements of the simple notion of causality serve to exclude some correlations as non-causal. We expect causes to precede their effects, and we expect that when a cause is manipulated—when we bring it about or take it away—it will make a difference to the effect. Along with correlation, then, we have found three indicators of a causal relationship between A and B:

Indicators of Causality: It is likely that A is the cause of B when:

1. A is strongly correlated with B
2. A is temporally prior to B
3. The occurrence of B depends on A

When we have evidence for all three of these indicators, we have a strong inductive argument for a causal conclusion.

6.4 The Logic of Experimentation

The previous section argued that we can have strong inductive arguments for causal conclusions when our premises capture three causal indicators. With this idea in hand, we can see why experiments are such a pervasive feature of scientific inquiry. Properly designed, experiments can give us all three kinds of evidence, and therefore experiments are very powerful ways to establish knowledge of causal relationships.

Lind's experiment with scurvy, discussed in Sections 2.3 and 4.3, illustrates the main elements of a modern scientific experiment. First, recall that Lind divided his sailors into groups. While he had several groups, we can say that there must be at least two, what we commonly call the the **control group** and the **experimental group**. The experimental group will be exposed to the potential cause, while the control group will not. In Lind's experiment, he selected twelve sailors from those who came down with scurvy. These twelve were divided in to four experimental groups, given vinegar, cider, seawater, and lime juice. Lind was comparing these experimental groups with the other sailors on the ship who had scurvy. The other sailors did not drink lime juice, vinegar, cider, or salt water; they were the control group. Except for the intervention, the control and experimental groups were the same. All of the sailors were on the same ship on the same voyage. So, they were exposed to the same weather, the same work, and the same diet. While it is difficult to achieve in practice, the ideal is to create a situation where the potential cause is the only difference between the control group and the experimental group.

In an experiment, we introduce the potential cause to the experimental group, and then we look for a change. Of course, not all experiments are successful, so there may not be one. In Lind's experiment, while there was a dramatic change for those who drank citrus juice, there was no change in the others. However, if there is a change, notice what we have found: a correlation between the purported cause and effect. In the control group, there were sailors who had scurvy but did not drink citrus juice. Since all the sailors had scurvy, the probability of having scurvy, given that they did not drink citrus juice was 100%. In the experimental group, none of the sailors had scurvy after they drank the lemon juice. So the probability of having scurvy, given the citrus juice was 0%. If you recall the earlier discussion, that is not only a strong correlation, it is a perfect correlation.

If Linds' experimental and control groups had been larger, it is likely that his correlation would not have been perfect. Even given some citrus juice, some sailors may not have recovered. And some of those who did not receive the juice may have recovered anyway. When we are dealing with very complex systems—like the human body—we expect such variation. Nonetheless, giving citrus juice to the experimental group would have changed the number of sailors with scurvy, and would thereby have produced a correlation. Notice that *any* change to the experimental group will constitute a correlation. An experiment is designed to find a difference in probability between those exposed to the purported cause and those who are not. If there is any change to the experimental group, then the definition of a correlation must be satisfied. If there is a correlation to be found, then a properly designed experiment can find it.

Notice also how a properly designed experiment gives us evidence of both a temporal difference between the purported cause and the effect and the right sort of dependence. In Lind's experiment, both the control and the experimental groups had scurvy before the experimental intervention. The change occurred afterwards, demonstrating the appropriate temporal relation. Also, the introduction of lemon juice is a kind of manipulation: we add the purported cause and look for a change. In a different experimental set up, we might also make a change on the experimental group by *removing* a purported causal factor. These manipulations reduce the possibility that the correlation discovered in the experiment could be due to a common cause. When a correlation is produced by a common cause—such as the correlation between speaking Tibetan and wearing monks robes—changing one property does not bring about change in the other.

While the experimental manipulation reduces the possibility that the correlation is due to a common cause, it does not eliminate it. If there is a common cause in an experimental set up, it would have to somehow be the cause of both the intervention and the change observed in the experimental group. The experimental setup thus narrows the range of possible common causes. These are typically called "biases" in the experiment, since they bias or make the results inclined to show a particular result. Biases can have many sources, so they are best explained by example.

In a medical experiment, the experimenter him or herself is an important source of bias. Suppose, for instance, Lind had chosen particularly young and strong sailors for the experimental group. These sailors might have been likely to recover from scurvy anyway. In this case, the correlation between drinking lime juice and recovering from scurvy would not be due to the causal efficacy of the lime juice. Rather Lind himself could be the common cause of the correlation. He chose the strong sailors to be in the experimental group, and he gave them the lime juice. So, by selecting particularly strong sailors for treatment, Lind would have biased his results and would not have been able to tell whether it was the antecedent strength of the treated sailors or their drinking citrus juice that made a difference. A crucial part of experimental design, then, is to be sure that such biases are not present.

Experiments exhibit the same logic of inference from samples to populations as we discussed Chapter 5. As a result, the same criteria for evaluation apply when we are interpreting the results of an experiment. However, the special characteristics of an experiment means that there are some complexities in the application of our criteria. First, consider sample size. The issue of sample size applies primarily to the estimation of the size of the sample needed to elicit the effect. In Section 5.3, we noted that when a phenomenon is rare, it takes a large sample to discover it. The same point goes for correlations: if an effect is very small, then it would take an experiment with a large sample groups to discover it. Lind's experimental groups had only three sailors each. This is far too small by modern standards. In Lind's defense, one might point out that the effect he found was very strong, so strong that could be detected by a small sample. Even so, a modern test of Lind's theory would have had larger experimental groups.

Representativeness is a very important evaluative criterion when we are interpreting an experiment. The control group and the experimental group are samples. A properly designed experiment can give one strong reason to believe that the experimental intervention—what was done to the members of the experimental group and not to members of the control group—is what gave rise the difference between the results then observed in the experimental and control groups. After all, if the individuals assigned to the two groups were similar in other respects, the experimental intervention will be the only salient difference in the two groups that might give rise to the different results. We then need to generalize this result to the whole population, and this is where representativeness becomes a concern. Just as in other kinds of inductive inference, we want to make sure that the sample is like the larger population in relevant ways. In the natural sciences, this issue is often relatively easy to resolve, since there is little reason to believe that, for instance, samples of gold differ from the larger population. In the life sciences, medicine, and the social sciences, things are more complicated. For example, a medical experiment conducted only on adult males may or may not generalize to women or children. A social scientific experiment conducted in a rich, industrial society may not generalize to poor, rural communities. Here we have to be very careful to determine that the sample and population are similar in relevant ways.

In addition, the special characteristics of experiments means that judging representativeness creates some unique challenges. In an experiment, we want to be sure that the only difference between the control and experimental groups is the experimental intervention (the purported cause). This often means that we have to be careful to isolate both control and experimental groups from unwanted influences. In a laboratory, this might mean special shielding. In an experiment involving humans, it may mean isolating them in a hospital. In the general population, however, the people or objects will not be protected from outside influences. It is common for a cause demonstrated in a laboratory setting to fail to arise in natural settings. Sometimes, contravening causes so overwhelm the phenomenon that the effect observed in the laboratory cannot arise. In cases where a well designed experimental result cannot be generalized to the whole population, scientists infer that the causal relation exists, but it is rarely seen "in the wild." While they may be less useful practically, knowledge of such subtle causes can be very important for our understanding of fundamental causal processes. Moreover, they may be the basis of important technologies, where the shielding necessary to produce the effect is built into the device.

6.5 Conclusion: The Observational Basis for Causal Knowledge

We began this chapter with the fundamental epistemic problem of causation: causal relationships are never directly observable, so what observations and inferences will allow us to have scientific knowledge of causation? This chapter has focused on the observational basis for causal knowledge. Correlations are one of the most important forms of evidence for causation. Indeed, it is necessary: without a correlation between A and B, there can be no causal relationship between A and B.

Identifying correlations turns out to be a subtle matter. A few instances of A and B happening together is not enough to establish a correlation. To make a claim about a correlation is to say something about a whole population. This means that we have to attend to matters of sampling and inductive generalization, as discussed in Chapter 5. In addition, if we are to use a correlation as the basis for a causal argument, we need to rule out the possibility that the correlation is due to chance and that it is due to a common cause. While we can do so without an experiment, we have seen in this chapter that a properly designed experiment will rule out common causes as an explanation of the observed correlation.

The upshot of this chapter is that some inductive arguments with causal conclusions can be strong. A properly designed experiment will establish a correlation, if there is one to be found. By design, it shows that the cause comes before the effect. And it will show that introducing the purported cause (or taking it away) makes a change in the effect. It will therefore show, with a high degree of reliability, that the purported cause really did bring about the effect. When we generalize from the experiment to the larger population, we need to evaulate this as we do any other inductive argument, and we need to attend to some special features of experimentation as well. Just as with inductive arguments from the distribution of a property in a sample to its distribution in a population, we can mount a strong inductive argument from an experiment to a population-level causal relationship.

Chapter 7

Causal Processes and Causal Modeling

The foregoing chapter drew some indicators of causality from what we called the "simple" conception of causality. Our concerns there were largely epistemological: can we make inductive inferences to causal conclusions? In this chapter, we turn to somewhat more metaphysical concerns. This chapter will begin by unpacking the simple conception of causality and illustrating several features of the contemporary scientific way of thinking about causes. We will then turn to examples of natural phenomena that do not seem to fit the simple model. In the latter sections of this chapter, we will show how the simple model of causation can be extended to more complex examples. Extending the basic understanding of causation and the arguments that support causal knowledge to these complex cases will involve *causal modeling*; the dissolution of complex phenomena into systems of simple causes.

7.1 The Simple Conception of Causality

Newton's theories of motion, including his accounts of gravity and planetary motion, projectile motion, marked a profound change in science. Suddenly, scientists had a unified and precise way to explain a wide range of phenomena. And they were doing so with a system of ideas that was deeply different than the science inherited from the Greek scientists of antiquity. It was a radical change in thinking. Because these new scientific ideas conflicted with ancient science, they provided challenges to philosophers. Philosophy in Europe has always drawn strongly on its Greek roots. The concepts of Greek science and philosophy were closely intertwined. The radical change in the scientific world view during the eighteenth century, then, presented significant challenges to the philosophers.

It was during this period of scientific and philosophical change that David Hume (1711-1776 CE) turned his attention to causality. Causality had been analyzed by Greek philosophers, and their ideas were widely accepted by European philosophers of Hume's time. Newton and his fellow scientists of the eighteenth century were deploying what they called the "mechanical philosophy." Everything was to be explained in terms of the motions of elementary particles and forces on them. The universe, on their view, was like a clock: a hidden, intricate system of interacting parts produced the changes we can observe. Hume expressed the idea of causality he saw in eighteenth century science in this way:

> an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed.[1]

Hume had his own philosophical understanding of causality, but will not linger over it here. Hume's formulation has been used by philosophers with many different points of view because it expresses the phenomenon that requires understanding. For this reason, it makes a good formulation of what we will call the simple concept of a cause.

To unpack what is implicit in the simple concept of a cause, let us consider a simple, everyday example. I rub a match across a rough surface and it lights. The match lit *because* it was struck. Here we have "one object followed by another." Causality, on this view, is a relation between two objects. But what is an "object" in this context? It is clear that by "object" Hume does not mean just the match. All by themselves, objects are not causes. The match burst into flame because something happened to it (the striking). Or again, it is not the rock, but the motion of the rock that breaks the window. It seems best to read Hume's talk of "objects" accordingly. An event of one sort was followed by an event of another sort. The head of the match sliding across the rough surface was followed by the match head igniting. Paradigmatic causes and effects are events regularly exhibiting such temporal order

Hume's formulation has two parts. The first sentence can be read as very loose characterization of a correlation: a pair of events that occur together. As we saw in Section 6.3, however, it is a mistake to identify causation with correlation. Hume agrees, and he clearly points beyond mere correlation here. The first sentence also specifies temporal order. Although the second sentence seems intended to rephrase the first by saying "in other words," it clearly expresses a kind of dependency: when A causes B, without A, B would not have happened. Reading the two sentences together, we get the suggestion that the cause, A, is what made a difference in whether B occurred, so that with A (e.g. the striking), the correlated type of event B (e.g. the ignition) happened, and B would not have happened without A having happened first in this case.

Just what is this dependency relationship wherein one event brings about another? That question leads us deep into metaphysics. It is an important question, but answering it would take us too far afield. There are many answers in Western

[1] David Hume, *An Inquiry Concerning Human Understanding*, New York: Bobbs-Merrill, 1955, p. 87. Original edition, 1777.

philosophy. Philosophers (mostly) agree that causes depend on effects in the way just outlined, and our consensus approach to the philosophy of science demands that we leave the matter here, on the surface. Nonetheless, there remain some interesting aspects of this dependency relationship that can be touched upon.

One might ask whether the dependency is such that whenever the cause occurs the effect must occur. Can there be exceptions? Certainly, our evidence for causality typically includes such exceptions. Matches do not always light when struck; windows do not always break when rocks hit them. Does this mean that causes do not always produce their effects? Answering this question requires that we reflect further on the second part of Hume's formulation.

For many simple examples of causes, Hume's statement that "if the first object had not been, the second never had existed" is false. The match may have burst into flame from other causes—perhaps it was lit by another match. In general, a kind of event can be brought about in many ways, which is to say that an effect can have many different causes. Nonetheless, there is a sense of "cause" that has the properties that Hume expresses. These are often called *necessary* causal conditions.

To understand necessary causal conditions, let us consider a different example. Instead of the match, consider the burner on a gas stove. When we bring a lit match close, the burner ignites. The match's flame is the cause of the burner's lighting. Now, if we turn the knob and stop gas, the flame will go out. This fits Hume's formulation: if the gas had not been present, the burner would not have ignited. The gas is necessary for the flame; it is required. The match is not necessary, since other sources of heat would have brought about the burner's flame. But in the presence of the gas (and some other things), it is enough or *sufficent*. Hume's formulation thus contains two important ideas:

Definition of a Sufficient Causal Condition: A is a sufficient causal condition of B if and only if when the occurrence of A brings about B.

Definition of a Necessary Causal Condition: A is a necessary causal condition for B if and only if B will not come about in the absence of A

Considering on the example of the burner, we can see that the ignition of the burner depends on both necessary and sufficient causal conditions. When the gas is on, the match will light it; without the gas, the match will not light it. In fact, the example is a little more complicated. The burner will not light without Oxygen, since the burner's fire is a reaction of oxygen with the gas. The oxygen and the gas are both necessary for the flame, and only in their presence is the match sufficient to light the burner. Whenever there is a sufficient causal condition, then, there are also a number of necessary conditions in place.

In Section 6.3, we argued that the epistemic indicators of causes drew on *three* aspects of the simple concept of causality. They were correlation, temporal order, and dependence. It is clear, we hope, how activities like experimentation use these epistemic indicators to provide evidence of causal relationships. Scientific methods often manipulate possible causes by adding them or removing them from systems. If the purported cause is a real cause, then adding it or removing it will produce some change in the effect—assuming that the appropriate necessary conditions are in place. Where manipulation is not possible, the relevant correlations can be used. If A is a sufficient causal condition of B, then we would expect the presence of A to make B more likely. Similarly, if A is a necessary causal condition of B, we would expect that taking A away would make B less likely. So both the dependence and correlation identifiers are supported by the simple conception of causality.

What about the temporal indicator? It too is supported by the simple conception of causality, though it has been left implicit in our examples. In the definition of a sufficient causal condition, above, we said that the cause "brings about" its effect. In all of the examples we have considered, the cause brings about a *later* effect. Is this necessary? Could causes bring about effects that were earlier in time? While fiction writers sometimes imagine such possibilities, scientists do not. What is implicit in the examples must be made explicit: causes always precede their effects. We now have adequate materials to fully characterize the simple conception of causation.

Definition of a Cause: A is the cause of B if and only if:

1. A is temporally prior to B, and
2. Either one of the following two conditions holds:
   (a) A is a sufficient causal condition for B,
   (b) A is a necessary causal condition for B

Notice that a cause may satisfy either 2a or 2b. As discussed above, causes are always part of a system of necessary and sufficient causal conditions. Whenever we identify something as "the" cause of an event, we are picking out just one element of a system. So, talk of "the" cause is potentially ambiguous between necessary and sufficient causal conditions. Whether we are looking for sufficient or necessary causal conditions often depends on our interests. If the lights have gone out, then I am going to look for something wrong. Some necessary part of the causal system is missing. If I want to bring something about, such as lighting a burner or breaking a window, I will look for a sufficient causal condition. The simple conception of a cause, then, comes apart into a system of elements. We will continue to follow ordinary speech, however, and talk about causes and about the simple conception of causality, leaving it to context to determine whether necessary, sufficient, or some combination of conditions is intended.

We are now in a position to answer the question asked earlier: do causes always bring about their effects? The answer is "yes," as long as all of the necessary causal conditions are in place.[2] The match will always light the burner, as long as

[2] Another qualification would be a counteracting cause. Perhaps someone blows out the match just as I bring it to the burner. The reader is invited to reflect on examples like this.

the burner is on and there is Oxygen in the room. For this reason, the probabilistic character of correlations, which are our primary form of evidence for causality, does not confict with the deterministic character of the simple conception of cause. When we introduce a cause into a population, such as giving lime juice to sailors with scurvy, the effect may not be exhibited by every member of the population. The members of the population are different, and they may not all exhibit the conditions necessary to support the action of the cause. Reality is a messy place, and so the evidence for causality is messy, even if causality is clean.

7.2 Understanding Causal Processes

While we have presented it as a consensus position, the simple conception of causality is not without its challenges. Even with the more sophisticated ideas of necessary and sufficient causal conditions at hand, the simple conception seems to apply best to straightforward mechanical relations. Lighting a match, a rock breaking a window, or a bat striking a ball, are clear examples of causal relationships, but not all causes are like this. Two kinds of causal phenomena have seemed to resist understanding in terms of simple causal relationships: causal processes and purposes.

Consider, for example, grinding stone used by the Tibetan man in Figure 7.1. As long as he pushes on the handle, the stone will turn. It seems fairly obvious

Figure 7.1: Tibetan man using a grinding stone, 1938[3]

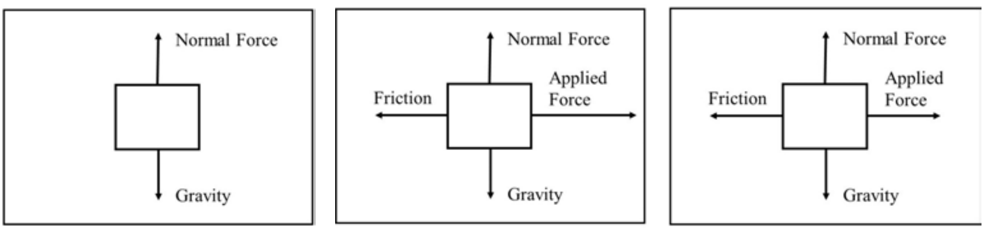
that pushing on the handle causes the stone to turn. But what corresponds to the

[3]Image used under Creative Commons 3.0 license. Source: Wikimedia Commons, Bundesarchiv, Bild 135-BB-152-11, Bruno Beger, photographer.

A and B in the definition on page 84? The apparent answer is “the push” and “the motion,” but these seem rather unlike the example of the striking the match and its ignition. In that example, there were two separate and discrete events. For the grinding stone, the pushing and the movement would seem to be simultaneous. This means, at least, that the temporal condition of the definition is apparently violated: once the stone is moving, the pushing does not happen before the movement.

Newton’s laws of motion remain the best explanation we have of motions like the grindstone. To explain the motion, we need to appeal to Newton’s First and Third laws of motion (page 13). The First law says that objects in motion tend to remain in motion unless acted upon by another force. The grinding stone does not stay in motion. If the man stops pushing the handle, the grindstone stops. Newton’s first law entails that there must be some force acting on the grindstone to make it stop.

The force that makes the grindstone stop is *friction*. Friction is a commonplace and familar phenomenon. Consider a table. If you push on it gently, even lean on it, it will not move. The table resists motion, as if it were pushing back on you. Newton conceptualized this resistance as a force. The table legs touch the floor, and the fiction of the legs against the floor create a force that resists your push. When you do manage to move the table, it will quickly come to a stop. Without the friction, it would continue moving by itself after the initial push. This is what one sees when objects are on ice or other low-friction surfaces. For a smooth object on ice, once it is in motion the low friction allows it to continue to move, as Newton’s First Law says it should, while the small amount of friction gradually slows it down.



(a) At rest (b) Under acceleration (c) Constant velocity

Figure 7.2: Free body diagrams for an object subject to friction[4]

The man with the grindstone needs to push the handle to get it to move. Newton’s Third law says that for every action, there is an equal and opposite reaction. According to this law, the force of the man’s push is opposed to the force of friction. The forces involved can be displayed using “free body diagrams” as illustrated in Figure 7.2. The rectangle represents the grindstone, and the arrows represent forces acting on it. The lengths of the arrows represent their relative

[4] Images by the authors.

strengths.

Figure 7.2a represents the grinding stone at rest. The only forces acting on it are gravity (the down arrow) and the "normal force" (the up arrow). The normal force is the force exerted by the surface on which the grinding stone rests. (Without the normal force, the object would be falling!). There are no other forces acting on the grindstone when it is at rest, so there are no other arrows. And the force of gravity and the normal force are equal, so the stone does not move up or down.

Figure 7.2b shows what happens when the man begins to push on the grindstone. The normal force and the force of gravity remain the same. However, now there is the "applied force" of the man pushing the handle. This is opposed by the force of friction. Notice that the arrow for the applied force is larger than the force of friction. This extra force is necessary to overcome the inertia (Newton's Second Law) and friction in order to get the grindstone moving. If the forces remained unbalanced, however, the larger applied force would cause the grindstone to continue to accelerate (that is, to move faster and faster). To maintain a constant velocity, the man reduces the applied force so that it balances with the frictional force, as represented in Figure 7.2c.

Newton's laws conceptualize the movement of the grindstone rather differently than common sense. What seemed like a simple relationship between a push and movement is analyzed as a complex of causes. Notice that the effect Newton explains is not simply movement, but the grindstone's acceleration and subsequent movement at a particular velocity. The cause of the velocity is not one event, but a pair of forces. The sufficient causal factor is the change in the force of the push. This must happen before the grindstone can accelerate. And when the force changes—because the man pushes harder or more gently—the velocity will change.

Newton's explanation of the process of pushing the grindstone analyzes the process into a complex interaction of forces over time. It illustrates an important strategy of contemporary science: causal modeling.

Definition of a Causal Model: A causal model analyzes a process or system into component parts, and it specifies the causal relationships among the parts.

There are three key ideas in the above definition: the system, the parts, and the relationships. The system is the overall phenomenon to be explained or understood. In the case of the grindstone, it was the fairly simple matter of the grindstone continuing to move while the handle was turned. Notice that this system has an input (the applied force) and an output (the acceleration or velocity), and many systems are like this. But not all. In some systems, the phenomenon we want to understand does not have a clear input or output. In such cases, we often want to understand how such a phenomenon could arise or continue. The example of purposes in nature, discussed in Section 7.3.1 below, is such a case.

The parts of a system can take a wide variety of forms. Earlier, we noticed the ambiguity of Hume's word "object" when talking about causes and effects.

Sometimes the parts of a causal model are objects or states of affairs, like the presence of the gas and the oxygen in the example of igniting the burner. Other times, the parts of the system are events—the ignition of the match. The parts of Newton's analysis of the grindstone are the forces of the push (applied force), friction, gravity, and the normal force. Forces, as conceptualized by Newton, are not commonsense objects, they are postulated by the theory. They illustrate the wide range of theoretical entities that might serve as as elements of a causal model.

A causal model shows how the phenomenon arises by showing how the parts are causally related. The relations may be necessary or sufficient causal conditions. In contemporary physical sciences, these relationships are often specified in terms of mathematical relationships. In your physics classes, you will measure the forces on an object as you push it and use Newton's laws to calculate the velocity. In other cases, the relationships are described qualitatively, not quantitatively. Nonetheless, in all cases, the relationships should satisfy the definition of a cause (page 84, above).

We began this section with a challenge to the simple conception of causality. A continuous process, like the movement of the grindstone, did not seem to fit the definition. Newton explained this motion by analyzing the larger process into a system of necessary and sufficient causes—the changing forces and their relationships. This not only meets the challenge for the simple conception of causality, it shows us how causal modeling can be a more general strategy of scientific understanding. Let us now turn to the other challenge to the simple conception of cause, the challenge of purposes. Can causal models adequately explain natural purposes?

7.3 Modeling Purposes

Purposes are commonplace in the human and natural world. When I go to the teashop, getting tea is my purpose. The purpose of a frog's green coloration is to protect it from being seen by predators. We use purposes to explain events and state of affairs: I went to the teashop *because* I wanted tea, and the frog is green *because* that color is protective. The use of purposes to understand and explain is a prominent feature of biology. Biologists often explain why animals have different organs, such as eyes or livers, in terms of what they do for the animal.

In the course of our discussion in this chapter, we will argue that the purposes involved in biology and in intentional action will need to be given different analyses. But, both present a similar challenge to the simple conception of cause. Explanation in terms of purposes has an interesting structure. Something happens or exists *now* for the sake of some event or state of affairs in the *future*. When I am walking to the tea shop, my drinking tea lies in the future. Similarly, when the frog is born, it has a protective color. No predators have yet seen the frog, so the protection is yet to happen. Future purposes explain present events and states of affairs. It seems to follow that purposes cannot be causes. If this is right, then purposes, whether biological or intentional, cannot be causally explained.

The conclusion that purposes cannot be causes (at least, given the simple conception of cause) has motivated quite a bit of reflection by philosophers and scientists on the character of scientific knowledge. If purposes are good for explanation and understanding, and if purposes cannot be causes, then there either must be some kinds of knowledge that does not depend on causes, or there are different kinds of causality. Indeed, as we will briefly discuss in Section 11.1, below, there was a time when scientists and philosophers thought that the phenomenon of life was outside of the grasp of science, or at least outside of the kind of science that Newton exemplified. To understand biological systems, some argued, we would have to develop something entirely new.

Science did develop something new in response to these challenges, but it was not a new kind of causality or scientific knowledge. It was a new kind of causal model.

7.3.1 Evolution and Natural Purposes

Charles Darwin's (1809–1882 CE) theory of natural selection stands as one of the most important achievements of western science. It is now foundational for the science of biology. From the point of view of this chapter, its value is that it provides an account of natural purposes. As we will see, it does so by understanding talk of purposes as pointing to a kind of a causal system that gives rise over time to the characteristics of plants and animals. The theory of natural selection provides a general way of modeling those causal systems. By doing so, it showed how the anomalous phenomenon of natural purposes could be understood in terms of simple causality. Darwin's theory has three central ideas:

Heritabilty: Offspring inherit traits from their parents

Variation: Inherited traits can take a variety of forms, and offspring are not identical to their parents or each other

Differential reproduction: Offspring with different traits reproduce at different rates

We can illustrate Darwin's theory with the example of protective coloration. Most animals have skins, fur, or plumage that helps them blend into their environment. This is the phenomenon we want to explain. On Darwin's theory, it arises through natural selection. Skin color is a heritable trait; offspring tend to look like their parents. This easily observable in both humans and animals. When a light colored dog mates with another light colored dog and has a litter of puppies, many of them will also be light colored. However, it is typical that not all of the puppies have the same coloration. Even if both parents are light colored, some of their puppies might be dark. The trait of skin color thus exhibits variation.

As we all know so well, the world can be a difficult place. Not all offspring survive long enough to reproduce. Predation is one reason why. Frogs, for example, are eaten by birds. This is where the idea of protective coloration comes into play. A frog whose skin color is more like its environment is less likely to be seen by a bird, and therefore less likely to be eaten. If it is less likely to be eaten, then it is more likely to reproduce. So, frogs who blend in to their environment are more likely to reproduce than those who do not.

A Darwinean analysis of protective coloration puts all three points together. Skin color is heritable, and it exhibits variation. Because predators can see some skin colors better than others (in a given environment), animals with skin colors that stand out to predators are less likely to survive and reproduce than those that are less visible. This means that animals with the less visible skin coloration have more offspring than those with the more visible skin color. Animals in a current population are adapted to their environment because they have traits that helped their ancestors survive.

Natural selection thus provides a causal model of protective traits. It explains the apparent purposiveness of a protective trait by analyzing the phenomenon (that is, the fact that animals have protective skin colors now) into a system several causal relationships among animals. Variation and heritability are simple causal relationships between parents and their offspring. Predation is a simple causal relationship between predators and prey. If these causes act over multiple generations, the phenomenon of protective coloration will emerge. While all the causes are simple, they are modeled in a system that lets us understand why protection from predation is the explanation for an animal's coloration.

Darwin's theory of natural selection can give similar models of many biological traits. Crucially for the purposes of this chapter, it shows how an animal can have a trait *because* that trait has a purpose. Through the use of causal modeling, simple causation can be extended to interesting and difficult cases.

7.3.2 Intentions and Goals

Darwin's theory of natural selection shows us how to analyze natural purposes, like camouflage, into systems of simple causes. What about the purposes that are part of action? For example, suppose we see a cow walking toward the creek. When it reaches the water it drinks. This is an example of purposive action: the cow was walking to the creek *in order to* get a drink. Like natural purposes, purposive action is an apparent challenge to causal understanding. The drinking explains why the cow went to the creek, and in this sense, the walking seems to depend on the drinking. But the walking happens before the drinking, so the drinking cannot be the cause. Unlike natural purposes, the purposes of action cannot be explained by evolution. While the desire for food and drink might be a product of evolution, it would be a misunderstanding to say that the cow's desire for water just now is caused by evolutionary processes. Here again, however, causal modeling resolves the difficulty.

A causal model, to repeat, analyzes a phenomenon into more elementary parts, and it postulates simple causal relationships among them. Our phenomenon is purposive action: doing something (like walking to the water) for the sake of something else (to drink water). A natural and common way to explain an animal's actions is to appeal to its desires. We might say "The cow walked to the water because it wanted to drink." The cow's desire for water is a state of its mind (or brain…an issue we will discuss in Chapter 11). Now, as we all know, a desire alone is not sufficient to cause an animal or a person to act. The person or animal must also have some information about their environment. Lobsang may desire some tea, but if he believes that the teashop is closed, he will not walk to the teashop. Similarly, the cow will walk to the creek only if thinks that there is water in the creek. The explanation of the cow's action, then, is that it is walking to the creek because it wants to drink, and it thinks that there is water in the creek. This explanation of purposive behavior is the basis for a causal model.

In a causal model, the phenomenon of purposive behavior is modeled as the product of two kinds of psychological representations: goal states and information states. An animal's goal states include wants, needs, and desires. These goal states are aspects of the creature's mind or brain. The animal also has information about its environment. The causal model of purposive action, then, analyzes an action as the effect of the goal states and the informational states. In other words, the purposive action of walking to the creek is caused by a desire for water and the belief that there is water in the creek.

Just as with evolution, the causal model of purposive action shows how a complex phenomenon arises from a system of simpler causes. The original puzzle of purposes was that the causes (drinking the water) seemed to precede the effects (walking to the creek). The causal model shows that this is an illusion. The cause of the action is not the later drinking, but the prior mental representations of goals and information.

One could object at this point that while such causal models might provide satisfactory explanations of animal behavior, they cannot be applied to humans. Human behavior is too sophisticated. Perhaps Lobsang desires tea, but also desires to pass his exam, and he believes that going for tea during the exam period is prohibited. Or perhaps he does not desire tea at all, but knows that drinking tea is required in some social situations. In the first instance, this objection is a call for more complicated models. Sophisticated causal models of human behavior have been developed in the social sciences, particularly economics, sociology, and political science. The details of these are beyond our scope, but the existence of such models illustrates the main point of this Section: that scientists seek to understand complex phenomena by analyzing them into simpler causal relationships.

The objection that causal modeling does not apply to human behavior also has a philosophical dimension. Many Western philosophers have been very uneasy about explaining human thought and behavior in causal terms. They have argued on a variety of grounds that aspects of human existence, such as conscious thought or intentional choice, do not fit within a causal system. In Chapter 11, we will explore the issue of whether there is a limit to scientific knowledge. The deeper question here is one that has fascinated Buddhist philosophers as well: what is the place of humans within the world? Such a question cannot be answered here. The lesson for us is that both scientific theories and scientific ways of knowing speak directly to the philosophical concerns of both Western and Buddhist philosophers.

7.4 Conclusion: Causation and Reductionism in Scientific Understanding

This chapter has explored the concept of causality as it appears in the practice of contemporary science. What we have called the "simple conception of causality" expresses a basic kind of dependence used in scientific explanation and understanding. We have seen how the identification of a cause typically presupposes a system of necessary and sufficient causal conditions. Moreover, by elaborating our description of such systems into causal models, we can account for subtle and complex phenomena.

There is a further ramification of causal modeling that might be of interest to Buddhist philosophers. Insofar as Buddhist philosophers espouse the doctrine of dependent arising, they may be sympathetic to the idea that complex phenomena like natural purposes can be analyzed into causal systems. For Buddhists, however, the observable world of stones and matches, frogs and cows, is not fundamental. These are subjected to analytical scrutiny and shown to be constructs. Contemporary scientists agree, and the device of causal modeling shows how such reductionism is manifested in scientific practice.

Consider again the example of striking the match and having it burst into flame. This fits the definition of a simple cause, but can this relationship between the striking and the ignition be more deeply understood? Scientists would say it can. In particular, we can treat the relationship between striking and ignition as a phenomenon to be modeled. We look for underlying micro-processes and try to understand how these parts are causally related. In the case of the match, this would be a matter of the composition of the match head, the forces of friction and the way they generate heat, and the chemical account of combustion. In this way, simple causal relationships at one level are analyzed as systems of simple causal relationships at a lower level.

The analysis of causal relationships into systems of more fundamental parts is an important aspect of contemporary scientific practice. Indeed, scientists typically think that they have not understood a phenomenon until they have understood the micro-level processes that bring it about. Is there any "bottom" or end to this process? The answer is that we don't know. Right now, the fundamental particles of physics are as far as we have been able to push the analysis. To see farther than that will take either more powerful instruments or more enlightened philosophical insight.

Chapter 8

Observation

Buddhism and western science share the idea that all knowledge is based on inference and observation. The previous chapters have been concerned with inference; now we turn to observation. Like inference, observation has been the subject of deep reflection by both Buddhist and Western philosophers. This chapter will explore two ways of conceptualizing observation. We will begin by eliciting some apparent features of scientific observation. We will see that while each of the features is desirable, they are in tension. It is difficult to find a philosophical understanding of observation that will satisfy them all together. We will then explore two different ways of thinking about scientific observation, and their consequences for scientific inquiry.

8.1 Desirable Features of Scientific Observation

Observation grounds scientific knowledge. Scientific theories are about the world around us, and it is only by consulting the world around us that we can attain scientific knowledge. Observation is thus crucial for testing theories, whether by falsification or by inductive support. What characteristics of observation support its role in scientific knowledge? Four characteristics of scientific observation arise from the use of observation to confirm or falsify theory:

Reliability. Observation should be reliable.

Neutrality. Observation should be neutral among theories.

Intersubjectivity. Observation should be intersubjective.

Predictability. Observation should be predictable from the theory it serves to confirm or falsify.

Philosophers and scientists who reflect on scientific methodology have argued that these four features of observation are all desirable. The philosophical challenge is to find a conception of observation that satisfies all four. Before turning to the different ways Western philosophers have thought about scientific observation, let us pause over these features in order to understand why they are plausible and what they are demanding.

Reliability is important for observation because the conclusion of an inference is only as strong as the premises on which it rests. Scientific knowledge requires inductive arguments and, as we have seen, the premises of many inductive argument in science characterize observations. While we can have strong (or valid) arguments without true premises, arguments with false premises give us no reason to accept the conclusion. Hence, scientific knowledge requires reliable observation. It is worth noting in this context that some historical attempts to do science have been rejected because the observations were just too unreliable. For example, the 19th century study of phrenology drew conclusions about personality and mental capacity from the shape a person's head. Phrenologists would make drawings that were supposed to reflect the head shape Unfortunately, the drawings varied enormously from observer to observer and they reflected the preconceptions of the observer more than what was to be observed about the person. The observations were unreliable, and this is one of the reasons why phrenology was never accepted as a science.

Another common idea about observation is that it serves as a neutral testing ground for competing theories. Recall the testing of the Ptolemaic and Copernican theories. The relative positions of the stars and the planets provided evidence for (and against) both theories. These observations were neutral in the sense that they gave no advantage to one or the other theory. Scientists who disagreed about which theory was correct could agree about the observations. When the phases of Venus were observed, for example, each of the disputants in the debate could agree that one theory had been falsified. Those defending the Ptolemaic view had to either accept the Copernican theory or find a way of modifying the Ptolemaic theory to accommodate the new evidence. Neutrality thus enhances the objectivity of science by eliminating bias and personal commitment to a theory from the process of confirmation or falsification.

Like neutrality, intersubjectivity supports the objectivity of science. If the theory is well supported by the evidence, anyone can see so by looking at the evidence for themselves. And if a theory is badly supported, then anyone can see that too. "Seeing the evidence for yourself" would be pointless if everyone saw different evidence. The public character of the observations that support scientific theories permits experiments to be repeated, and it allows one scientist to build on the work of another. Science is intersubjective in the sense that the reasons for accepting a theory should be available to anyone. Observation seems to give science this public character. We have already touched on the public character of science (Chapter 2), and we will do so again (Chapter 10).

Finally, if observation is to have a role in theory testing, it needs to be related to theory in the right way. In falsification, a hypothesis is derived from a theory. If the derivation is valid, and the observations show the hypothesis to be false, then the theory is falsified. Therefore, if the hypothesis is to be testable, it must be possible to observe whether it is true or false. Scientific observations, then, must be the sort of thing that correspond in some way to the sentences derivable from a theory. If, for example, Copernicus's theory predicts that Venus will be full on 12 June, then the sentence "Venus is full on 12 June" has been derived from the theory. Whatever observation is, it must be the sort of thing that can determine the truth or falsity of "Venus is full on 12 June." This is the sense in which observation must be predictable by theory.

8.2 Subjective and Objective Conceptions of Scientific Observation

Observation is a multi-faceted phenomenon, and as a result, talk of "seeing" or "observing" is ambiguous. In Western philosophical discussion, there are two broad views of scientific observation. Each draws on a different side, subjective or objective, of the commonsense notion. Consider the following example. David and Mark are walking through the woods. Mark sees a flash of blue in the trees, and says to David "I saw a bluebird[1] ." Now, suppose that there was no bluebird there at all. David, who is keenly observant, knows this and tells Mark. It would make sense for Mark to now say, "I did *not* see a bluebird." In one sense of "see," what we might call the *objective* sense, whatever Mark did, he did not see a bluebird. At the same time, Mark did have a visual experience. In another sense of "see," then Mark *did* see something. So, he might say "Well, I thought I saw a bluebird," or "I seemed to see a bluebird," or even—Mark and David are philosophers after all—"There was something going on in me that was subjectively like what goes on in me when I see a bluebird." This second, *subjective* sense of "see" is not undermined by the absence of the bluebird. Mark had the visual experience whether there was a bluebird there or not.

The little story about the bluebird highlights two aspects of observation, what we will call *content* and *correctness*. The issue of content concerns what Mark saw. Content captures the difference between the observation of a bluebird and the observation of a frog. These observations have different things as their object; they have different contents. One question about scientific observation, then, is: What is the *content* of scientific observations? Correctness is a matter of whether Mark was mistaken when he reported seeing a bluebird. Since scientists need to evaluate the reliability of an observation, any account of scientific observation must provide some grounds for evaluation. The second question about observation, then, is: What are the grounds for judging an observation to be correct or mistaken? Any philosophical account of observation will have something to say about both what constitutes the content of observation, and in virtue of what an observation is correct or incorrect.

The subjective and objective conceptions of observation have different understandings of what observation is about, as well as the criteria by which an observation can be correct or mistaken. Indeed, as we will see, answers to the content and

[1] A bluebird is a small North American bird with, as one might expect, bright blue feathers.

correctness questions are intertwined. To see this, let us characterize the subjective and objective conceptions of observation more precisely:

Subjective Conception of Scientific Observation: A scientific observation is a personal, sensory experience.

Objective Conception of Scientific Observation: An observation occurs when some thing, event, or property is detected with human senses or instruments.

On the subjective conception of observation, the content of observation is the visual experience. It is a psychological phenomenon corresponding to the sense in which Mark "seemed to see a bluebird" even when there was no bluebird to be seen. On this view, an astronomer who makes an observation is recording his or her experiences. The astronomer sees the relative position of some larger and slightly brighter spots (what we call the "planets") located against a background of a pattern of smaller white spots (what we call "the stars"). The background pattern remains the same across nights in which observations are made. However, those bright spots called planets move in the sense that, on subsequent nights, these spots appear in progressively different places relative to this background pattern. For example, a large red dot (lets call it "Mars") appears near *this* group of smaller white dots at sundown on one evening, and *that* group of dots several evenings later. If the astronomer continues watching for several years, he or she will see the series of patterns recur. In the subjective sense of observation, the scientific observations of astronomers are just the experienced patterns of bright spots. When that same astronomer uses an instrument like a telescope, the observation is the scientist's experience of the image image in the eyepiece.

The attraction of the subjective conception's account of observational content is that it passes a very high standard of correctness. In the example of the bluebird, Mark's initial judgment that he saw a bluebird was incorrect. When this was pointed out, he modified his judgment to "I seemed to see a bluebird." Epistemologically, this is safer ground. While Mark can be wrong about whether a bluebird flew by, he cannot be wrong about his own experience. There is no room for doubt about immediate experience. The subjective conception of observation thus provides an account of observational content about which we can be certain.

The certainty of observation would make it an excellent starting point for inference. As we noted above, the conclusions we draw from our arguments are only as strong as the premises with which we start. In science, the ultimate premises are observations. A starting point for science that was certain would be a clear strength. The subjective conception of observation thus claims to put science on a secure foundation, and it therefore purports to have a robust account of the *reliability* of observation.

The certainty of observation made possible by the subjective conception's account of observational content also seems to provide for *neutrality*. To see why, consider a variation on the bluebird example. Suppose both David and Mark notice a bird flying past, and Mark identifies it as bluebird. David points out that they are now in India where there are no bluebirds. So, according to David, Mark did not observe a bluebird. David is using a bit of background knowledge—that the bluebird is a North American bird—to evaluate Mark's observation. Mark might reply that "There are no bluebirds in India" is a theoretical proposition. His observation would potentially falsify the theory, so to use the theory as a reason to reject Mark's observation is circular. It would hold the theory immune from falsification in an unscientific way. The subjective conception of observational content provides a way out of this connundrum by making observation neutral. David and Mark can agree that they *seemed* to see a bluebird; that they had an experience that was subjectively like the experience of a bluebird. Since this experience is independent of theory, it can serve as a potential challenge, even when the theory is well established.

The objective conception of scientific observation regards the content of observation, not as an experience, but as the thing, event, or property detected. This conception picks up on the sense of "see" where Mark did not see a bluebird if there was no bluebird to be seen. The objective conception does not deny that seeing is an experience or psychological state. But it insists that scientific observation is more than seeing. When we speak of a scientist making an observation, we are speaking of a relationship between the scientist and the observable thing. The relationship is impersonal in the sense that anyone who positioned themselves properly would have the same (or very similar) experience. In the objective sense, scientific recording instruments also make observations. An instrument that is set up in the right way will be in a position to detect objects, events and properties. An instrumental observation is the reading on the dial, e.g. that the pressure is 150psi, not the visual experience of the numbers on the dial.

The objective conception insists that when scientists are using observation to test theories, they are talking about the objects, properties or events observed, not about their experiences. In this way, it nicely captures the way in which scientific theories predict observations. In general, when scientific theories predict an observation, they are predicting that an event will occur or an object will exhibit some property. For example, the Copernican theory predicts that, in a particular time of year, the light from the sun will reflect off the full side of Venus that faces the earth. This means that a properly positioned person—one with good vision, a cloudless night, and a properly focused telescope–could see a disk. Venus was see*able* as a disk, just as the theory predicted. The objective conception thus accomodates the *predictability* of scientific observation.

The *intersubjectivity* of observation also follows naturally from the objective conception. When Mark asks David "Did you see that bird?" he is presupposing that the observation is something they could share. They can both observe the same bird without having the same experience. Indeed, the fact that David saw it as well is evidence that Mark was not deceived by the shadows. Its intersubjective character is part of what gives us confidence in observation. Observation in science works on the same principle, but it is more systematic than everyday life. When we conduct experiments, we aim for them to be repeatable. The virtue of repeatability is that when others make the same observations, it gives us confidence that the result was not a fluke. If the experiment cannot be repeated, then the original experiment is not accepted as the basis for further research. Intersubjectivity is thus part of the objective conception's account of the reliability of observation.

We have, then, two accounts of scientific observation that differ about what the content of observation might be, and have corresponding differences about observational correctness. Each has some apparent strengths with respect to the four features of scientific observation identified in Section 8.1. The subjective conception accounts for the reliability and neutrality of observation, while the objective conception accounts for observation's predictability and intersubjectivity. Each also has apparent weaknesses. It is not clear how the subjective conception can account for the intersubjectivity of observation, while the objective conception cannot attain a level of reliability as high as certainty. Can either of these conceptions provide a satisfactory account of all four features? It is to this question that we now turn.

8.3 From Subjectivity to Objectivity in Observation

All four desirable features of scientific observation—reliability, neutrality, predictability, and intersubjectivity—arise from the role of observation in testing scientific theory. Predictability therefore holds a special place on the list. If a conception of scientific observation could not link observation to theory in a way that permitted observation to confirm or falsify theory, then it would clearly inadequate. Without an adequate account of predictability, it would not matter how well a conception accounted for the other desirable features of observation. Let us begin, then, by comparing how the subjective and objective conceptions accommodate the predictability of observation.

We have already seem that the objective conception of scientific observation has an account of predictability. At first look, the subjective conception has one too. Ptolemy's theory predicted that Venus would never appear as a full disk, but only as a crescent. Copernicus's theory predicted that it would appear as a disk. When Galileo looked, he saw the disk. So, Galileo's subjective experience of a disk confirmed Copernicus and falsified Ptolemy's theory. While this is an important argument, there are reasons to think that it is flawed.

According to the subjective conception of observation, the content of an observation is a psychological state. So, what confirms or falsifies a scientific theory is the psychological state of a human. The problem for the subjective conception is that the theories of physics and astronomy make no predictions about human psychological states. Copernicus's theory said that Venus was a sphere orbiting the sun. According to physics, the light from the sun reflects off of Venus' surface. When Venus is on the opposite side of the sun from the Earth, light from half of the sphere will reflect to Earth. When Venus is not opposite the Earth, a smaller portion of the light will reflect as a crescent. None of these predictions say what Galileo or any other individual would experience. Experience depends on many facts about the person: how good their eyesight is, whether they are tired or alert, whether they have good powers of concentration, and so on. So, contrary to the subjective conception, it is not accurate to say that a theory predicts how something will appear, if this means a prediction about experience.

To accommodate this objection, the subjective conception of observation needs to establish an epistemological link between the direct content of observation (an experience) and the properties of objects and events that scientific theory predicts. This is an issue with a long history in both the Buddhist and Western philosophical traditions. We can sidestep much of this discussion by simply pointing out that no way of connecting experience to objects will preserve the certainty of observation. It is always possible that the objects, properties, or events predicted by scientific theory are different from how we experience them. The subjective account of correctness will need to be modified so as to compensate for a variety of illusions and errors. Indeed, a defender of the subjective conception is likely to appeal to just the sort of error correction to which the objective conception appeals (see the discussion in Section 8.4, below).

The certainty of subjective experience, then, is much less of an asset than it might have seemed. While we cannot be wrong about our experience, experience is not the right sort of content to serve as the test of theory.[2] The grounds for evaluating observation need to appeal to factors outside of the subjective experience of the scientists. This point is made quite vivid when we consider the use of instruments in science. Galileo did not look at Venus with his eyes alone. The difference between the disk and crescent phases of Venus cannot be detected without a telescope. Galileo could see Venus as a disc, rather than as a crescent, only if the telescope was properly adjusted and focused. If it was not, then Galileo did not correctly or reliably see Venus as a disk, no matter what his subjective experience. The relevant grounds for evaluating Galileo's observation, then, concern is use of the telescope, not his experience.

A further disadvantage of the subjective conception of scientific observation is that it stands in rather direct contradiction to the intersubjectivity of observation. On the subjective conception, the content of observation is private in the sense that only the subject has the experience. No one but Galileo can have his experience. Observation, on the subjective view, cannot be literally intersubjective.

The subjective view can recover some intersubjectivity by pointing out that scientists can talk to each other. In this way, they can share their observational knowledge, and they can assess each other's observations. Also, anyone who is positioned in the right way will see the same thing. This approach to intersubjectivity does not support the sort of certainty that the subjective interpretation is designed to protect. Observations can be misdescribed or misidentified, hence

[2] Except, perhaps, in the science of psychology, though many of the same concerns expressed here apply to psychology as well.

there is a possibility of error. Moreover, if Galileo described what he saw to others, they would then have to accept his evidence on the basis of his testimony. The evidence for a theory is no longer the (certain) observation, it is the (uncertain) testimony of others. As with prediction, the subjective conception of observation is able to accommodate intersubjectivity only at the cost of lowering the standard it sets for reliability.

The subjective conception of scientific observation struggles with the predictability and the intersubjectivity of observation. As it accounts for these, it looses certainty in its account of observational correctness. The objective conception, by contrast, has a strong account of the predictability and the intersubjectivity of observation. Its criteria for reliability will never generate certainty. At this point, however, it should be becoming clear that certainty is an elusive goal for science. As argued above, once we admit that inductive arguments support scientific theories, we cannot hold that scientific knowledge is certain. We have seen in this Section that even if the content of scientific observation were immediate experience, and thereby certain, that reliability would have little or no force in grounding scientific knowledge. So, the consequence that the objective conception of observation loses certainty poses no additional cost.

In this section, we have made the case that the objective conception of scientific observation is at least *prima facie* superior to the subjective conception. The objective conception has a better account of the predictability and intersubjectivity of observation. The subjective conception's main strength is its account of reliability—raising the bar for correctness all the way to certainty—turns out to be an illusion. However, we have not yet looked carefully at the way in which the objective conception handles the reliability and neutrality of observation. As we will see in Section 8.4, neutrality presents some interesting challenges to the objective conception of scientific observation.

8.4 Theory Neutrality and the Reliability of Observation

Reliability and neutrality both concern the evaluation of scientific observation, that is, the judgment that an observation is correct or mistaken. The question of how the objective conception of scientific observation accommodates reliability and neutrality, then, is really the question of how correctness is determined on the objective conception.

The objective conception of scientific observation treats the content of observation as already intersubjective objects, events, and properties. That intersubjectivity goes some way toward providing grounds for correctness. One way to determine whether an observation was mistaken is to try to repeat the observation. When scientists conduct experiments, they carefully record their procedures. This has several functions. It allows other scientists to look for previously unnoticed sources of error in the experiment. It also allows the experiment to be repeated. Another team can create the same conditions and check the observation. Such checking by others is a straightforward way to determine whether an observation is reliable. Of course, no number of repetitions will guarantee that the observation is correct. Again, reliability will never amount to certainty on the objective conception.

The objective conception of scientific observation also makes it appropriate to use our background knowledge to assess an observation. Observation, on this view, is a relationship between the scientist or a detecting instrument and the objects, events, or properties detected. Using what we already know (or hypothesize) about the objects, we can determine the conditions under which they may be observed, we can identify them properly, and we can seek out the sources of possible error.

For example, consider the discovery of the planet Uranus in 1781. The planets that are further away from the sun than Jupiter, Uranus and Neptune, had not been previously identified. Uranus is visible without the aid of a telescope, but just barely, while Neptune is not. While visible, Uranus moves so slowly that its motion was not noticed. William Herschel (1738–1822 CE) observed Uranus with a telescope. He noticed that its position changed relative to the stars. This was the first indication that he was viewing something near to the sun, a planet or comet, rather than a star. He then used different magnifications on his telescope. He found that while the stars appeared as the same size in all magnifications, the apparent diameter of Uranus increased and magnification increased. This again was evidence that he was observing something relatively close to Earth. The final confirmation that he was observing a planet came when the orbit was calculated from his observations. The shape of the orbit showed that it was a planet, rather than a comet.

That Herschel observed a new planet was supported by at least three pieces of theoretically informed background information. First, planets and comets move, while the stars do not. So, observing a change in relative position indicated that he was seeing an object inside the solar system. Second, just as nearer objects appear larger than distant ones, the increase in apparent diameter relative to the stars as he increased the magnification of his telescope was also evidence that the object was close. Finally, there was the matter of identifying the observed object as a planet, rather than as a comet. This depended on the calculation of the orbit because comets have very elongated orbits compared to planets. Herschel's observation of Uranus in 1781 thus depends on a series of viewings on different nights and with different magnifications, as well as the mathematical extrapolation from those sightings to the planet's orbit.

The intersubjective character of observation and the use of background knowledge provide strong grounds on which to evaluate scientific observations. However, as we noticed earlier, appealing to background knowledge appears to sacrifice neutrality. We are using theory to characterize the objects observed and confirm that they were observed accurately. This seems to introduce a circularity into the evaluation of the observation: observations may be rejected when they do not conform to theory.

To explore this puzzle, let us consider a different scientific example. The electroscope was an instrument used in early studies of electricity to detect the presence of electric charge. It was a a specially made glass jar. Inside, two sheets of thin foil, sometimes made of gold, were suspended from a wire. The wire ran through the top of the jar, and was topped with a metal ball, as illustrated in Figure 8.1. When the ball of the jar was exposed to a charged surface, the leaves would

Figure 8.1: Electroscope[3]

spread apart. Often this was done by rubbing a glass rod with a silk cloth.

In the 18th and early 19th centuries, scientists thought that electricity was a kind of fluid flowing through wires. A positively charged object, such as a glass rod that had been rubbed with silk, contained the electrical fluid. When objects were drained of this fluid, they were negatively charged. In the following quotation, an early scientist describes an observation made with the electroscope:

> if you permit the excited glass tube to remain for some time near the ball of the electroscope, then on withdrawing it the gold leaves will first collapse and afterwards open.... Hence you will perceive that the electric fluid can be driven out of the lower extremities of the gold leaves by the repulsive action of the fluid.[4]

The "excited glass tube" had been given a positive charge and was thus full of the electrical fluid. When the rod touched the ball of the electroscope, the scientist *saw* the fluid flow into the jar and "drive out" what fluid was already present it the leaves. The scientist could see this because electricity was theorized to be a fluid.

Contemporary theories of electricity reject the idea that electricity is a fluid. Electricity is the movement of charged particles we call electrons. Electrons move from the glass to the foil leaves; they do not pour down the wire like water in a tube. Because the leaves are both negatively charged, they repel each other. On the contemporary theory, then, what is observed by the electroscope is the presence of an electrostatic charge.

[3]Image in the public domain. Source: Wikimedia Commons, from the book *Opfindelsernes Bog* 1878 by André Lütken.

[4]William Sturgeon, *Lectures on Electricity*, London: Sherwood, Gilbert, and Piper, 1842, p. 41.

As we noticed in Section 8.2, the lack of neutrality in observation is a potential problem for scientific knowledge. If observation depends on theory, then the capacity of observation to confirm or falsify theory seems to be diminished. This kind of bias seems to emerge in the case of observations with the electroscope. Proponents of one theory observed the “electric fluid” flow into electroscope. This observation makes sense from the perspective of the fluid theory. If electricity is a liquid, we can observe it flowing into a jar. Hence, observations with the electroscope seem to confirm the theory that electricity is a kind of liquid. But proponents of the charged-particle theory looked at the same device and saw the electrons on the foil leaves repel each other. Since, on their view, electricity is not a liquid, there can be no observation of a flow into the jar. Observing the electroscope thus cannot confirm one theory and falsify the other, for each theory has a different interpretation of the observation.

Can scientific observation be sufficiently unbiased to produce knowledge, if theory influences observation? Many philosophers and scientists would answer that yes: theories are still reliably confirmed (or falsified) even though observation is not neutral. There are three considerations that support this position. First, strong support for a theory typically relies on multiple sources. Second, theories confirmed in different domains support each other. And third, the use of science to study itself and refine its observations is an epistemic strength of science. Let us consider these ideas in turn.

In our discussion of how Newton’s mechanics was confirmed (Section 4.4), we saw that the theory was tested in a variety of domains. The theory entailed predictions about how planets, projectiles, and pendulums move, and the theory was confirmed in each of these domains (as well as others). This provided powerful confirmation for the theory because it was supported by a wide variety of evidence. Newton’s theory illustrates a general point about confirmation: well supported theories have broad support from different kinds of evidence. While the observations are informed by the theory, the observations are informed in different ways. So, even if the observations are not entirely independent of theory (that is, they are not neutral), the variety of observations makes the support for the theory less biased.

Second, theories are also supported by other theories. We saw this in Section 2.2, where Copernicus’s theory of the solar system (or, more precisely, Kepler’s laws) was supported by Newton’s physics. Newton could explain why the planets moved around the sun, but his theory was inconsistent with Ptolemy’s earth-centered theory. There was independent reason to accept Newton’s theory, since his physics had support from observations of projectile motion, pendulums, and other phenomena that had nothing to do with astronomy. This second consideration is thus an extension of the first. Because the two theories support each other, they further increase the breadth of observations that provide evidence. And again, even if the observations are informed by the theory, they are informed in very different ways. It would be extremely surprising if all of the biases fit together to support an inaccurate theory.

We can apply the two considerations discussed so far to the puzzle about the electroscope: scientists with different theories were looking at the same bottle, but seeing different things. Theory influenced the observation, to be sure, but the electroscope was not the only source of evidence for either theory. The idea that electricity is a liquid was supported by a number of experiments showing electricity to have some liquid-like properties. The theory that electricity was a flow of charged particles came later, and it was supported by observations and measurements taken from a wide variety of instruments. Many of the experiments and phenomena could not be explained by the liquid theory, but were explainable by the electron theory. The charged-particle theory was also supported by the new theory of the atom. On this theory, the atom is composed of positively charged particles (protons) surrounded by negatively charged particles (electrons). Just as Newton's theory had support independent of astronomy, the charged-particle theory had support independent of the experiments with the electroscope. Because of its superior support, the charged-particle theory of electricity ultimately succeeded the theory that electricity was a fluid.

The example of the electroscope is a bit unusual insofar as the two theories had different explanations about how the instrument worked. When an instrument is used for scientific observation, it is more typical for the instrument to rely on theories that are very different than the theories it is is being used to test. For example, astronomical observation, such as the observation of Uranus or of the phases of Venus, requires a telescope. We believe the telescope to be a reliable instrument because we have a theory of optics that explains how it works. The observations made by the telescope are thus influenced by theory, but the theory of optics is independent of any theories about the solar system.

The fact that we can explain how the telescope works highlights the third consideration in favor of rejecting neutrality: science studies itself and thereby increases the reliability of observation. An explanation of why an instrument works helps us understand the conditions under which the telescope might be reliable and what errors it might be prone to. We can thus use our understanding of the instrument to gain confidence in the observations made with it. This is an important instance of a point made in Chapter 2: science is self-correcting. In the case of observation by instrument, we direct our scientific inquiry onto the instruments and learn about them. This knowledge, in turn, strengthens the reliability of observations made with our instruments.

We conclude that neutrality, understood as the idea that observation should be completely independent of theory, is not desirable—and perhaps not even possible—in scientific observation. To be clear, a problematic circularity does arise when a theory supports just that evidence which favors the theory. The three considerations just discussed—that support (and falsification) arises from multiple sources, that different theories support (and falsify) each other, and that science searches for its own sources of error—show that such problematic circularities need not arise in science. Indeed, the foregoing discussion has shown now the reliability of observation is substantially enhanced when scientists appeal to theoretical background knowledge to seek out sources of error or to find new ways to observe something.

The considerations discussed in this section show again how the confirmation of theory is a *holistic* matter. To say that confirmation is **holistic** means that theories are confirmed as a body, as a whole, rather than in parts. Holism is a direct consequence of the structure of confirmation and falsification that we discussed in Chapters 3 through 6. Falsification requires that a hypothesis be validly deduced from the the theory being tested. If the hypothesis is observed to be false, then some part of the theory must be false. But a false hypothesis does not impugn any particular statement of the theory. *Some* statement must be incorrect, but we need to figure out which one is most likely to be false. The theory as a whole has been called into question, not just one specific part of it. The three considerations discussed in this section extend the holism of confirmation to the relationship among theories, and to the instruments that make many scientific observations possible. The holism of confirmation, then, further supports our conclusion that observation can be influenced by theory, but under the right conditions, this lack of neutrality enhances, rather than threatens, the reliability of evidence.

8.5 Conclusion: Certainty, Falliblism, and Scientific Knowledge

This chapter has argued for an objective conception of scientific observation. When we speak of the observations that support scientific inquiry, we are not speaking of the experiences of any scientist. Rather, the observations that support and test scientific theory are about objects, properties and events. Because the content of observation is objective, observations are the sort of thing that can be predicted by theory, and thus contribute to confirmation or falsification. Objects, events, and properties can be observed by multiple agents, perhaps using instruments. Thus, observation is intersubjective. The intersubjectivity of observation, along with the appeal to background theoretical knowledge, enhances its reliability. The objective conception does not require that scientific observations be strictly neutral, but as we have seen, this too enhances reliability.

While the objective view has significant advantages as an account of scientific observation, it requires that we give up on the idea that observation provides a certain, unassailable, basis for science. The other source of knowledge, inference, does not generate certainty either. Scientific inferences requires induction, and it is always possible that the conclusion of a strong inductive argument be false. Scientific knowledge is not certain.

The claim that knowledge is not certain may seem paradoxical. Knowledge, on many accounts, requires certainty, so if scientific inquiry does not generate certainty, then science is not a form of knowledge. But science *is* a form of knowledge. Indeed, scientific inquiry is one of the most powerful and effective ways to generate knowledge. How do we reconcile these ideas?

The first step is to recall the distinction between arguments and inferences made in Section 3.1. An inference is a relationship between thoughts or beliefs. When a person draws an inference, they form a belief (the conclusion) because they have other beliefs (the premises). Arguments, on the other hand, relate statements. The premises support the conclusion whether anyone believes them or not.

Certainty, like inference, is a psychological phenomenon. It is a conviction that something is true, a conviction that admits of no doubt. The observations and arguments that constitute scientific inquiry, on the other hand, are not psychological. Inductive relationships of strong support show that a conclusion is very likely to be true, if the premises are true. And the observations on which these inductions are based can be shown to be reliable by appeal to a instruments, experiments, or sampling proceses by which they were produced. The question, then, is whether we should align our beliefs with the conclusions of reliable observations and strong inductive inferences. When the question is put this way, the answer seems to be an obvious "yes." While doubt remains possible—and hence there is no certainty—aligning our beliefs with the best results of science is the rational thing to do.

For these reasons, most scientists and many Western philosophers have rejected the idea that knowledge requires certainty. They have adopted a position known as **falliblism**. According to falliblism, it is consistent to say that I know something, and yet is possible that I am wrong. The statement has the air of paradox, but the paradox is practical, not logical. If I am really in possession of some evidence that undermines my inferences or shows my observations to have been unreliable, I should not claim to have knowledge. But to say that scientific knowledge is fallible is not to say that I have such evidence. Rather, it is simply the recognition that there *could be* such evidence. As far as we can tell, our observations are reliable and our inferences are strong. At the same time, we recognize that neither observation nor inductive inference guarantees truth. It is possible, albeit unlikely from our current perspective, that we are wrong.

While many Western scientists and philosophers have accepted falliblism, it raises a profound questions for Buddhist scholars. In Buddhism, knowledge of the highest sort has been understood to require certainty. If Buddhist scholars hold onto the idea that knowledge requires certainty, then they have to conclude that science does not produce knowledge (though perhaps it produces a form of cognition that falls short of proper, full knowledge). This line of thought is troubling because science has seemed to confirm certain Buddhist commitments, such as emptiness and dependent arising. Rejecting the idea that scientific inquiry produces (proper, full) knowledge means diminishing the positive impact it can have on Buddhist thought. On the other hand, changing the Buddhist conception of knowledge would be a profound change. Accepting that knowledge does not require certainty will require making other changes in Buddhist epistemology. Understanding and debating the choice between these alternatives is one of the most important projects initiated by the encounter between Buddhism and science.

Chapter 9

Realism, Anti-Realism, and Scientific Progress

9.1 Does Science Progress? Two Views

Chapter 2 presented a brief history of physics and astronomy, describing how Ptolemy's theory of the solar system was succeeded by Copernicus's theory, and Copernicus's theory was succeeded by Kepler. Implicit in that description was the presupposition that science progresses. That is, over time, scientists gather more observations, and through falsification and confirmation, earlier theories are rejected and replaced by better theories. Successor theories are better because they represent reality more accurately. We will call this the *realist* view of scientific progress, and those who agree with it *realists*.

Until very recently, almost all scientists have been realists. The practice of looking for ways to make current theories better, and the corresponding falliblist attitude about one's own theories, naturally lend themselves to a realist view. Many philosophers have also been realists about science. In the twentieth century, however, an alternative view of science emerged. The *anti-realist* view of scientific progress holds that, contrary to appearances, successor theories do not represent reality more accurately than their predecessors. The changes in theories over time represent changes in perspective, just as different views of a mountain make some features apparent, but obscure others.

In this chapter, we will explore the debate between realism and anti-realism. This debate has been an important part of contemporary philosophy of science. There are many more possible arguments than we can describe in a short text. (Indeed, there are many more possible versions of realism and anti-realism than can be described in a short text!)[1] Nonetheless, we hope to sketch the main outlines of the debate.

[1] This chapter will speak about "the" realist and "the" anti-realist, but the reader should be aware that there are *many* versions of realism and anti-realism at large in contemporary philosophy of science. References to realism and anti-realism in the debate below refer only to the versions defined in the following sections.

9.2 The Realist View of Scientific Inquiry

Let us begin with a precise characterization realism's commitments.

Realism: A realist about scientific inquiry holds that

1. scientific inquiry aims at true theories,
2. the objects postulated by a true theory exist, and
3. successor theories represent reality more accurately than the theories they replace.

The first commitment describes goal of scientific inquiry. It does not say that scientific theories *are* true. This would be too strong. Indeed, since all previous scientific theories have been shown to be false in some respect, we have good reason to suspect that our current theories have defects too. This inductive inference about science gives us strong reason to be a falliblist. While we cannot identify the defects of our current scientific theories (though, as we will see below, sometimes we have suspicions), we can be confident that there are falsehoods somewhere. By holding only that science *aims* at truth, Commitment 1 is consistent with fallibilism. A falliblist can accept that the goal of science is to produce true theories, even if we always fall short of that goal.

Historically, one of the ways that theories have progressed is by postulating the existence of new kinds of entities, or by reconceptualizing familiar things. Ptolemy's theory proposed that the planets move in peculiar paths—the epicycles. Newton proposed that objects fell toward the earth because of the force of gravity. The atomic theory of matter proposed that atoms are made of protons, neutrons, and electrons, while quark theory proposed that protons and neutrons were composed of quarks. Commitment 2 further specifies the goal of truth by committing to the existence of such novel entities, when the theories proposing them are true. Indeed, a realist takes the existence of the entities to be at least a necessary condition for the truth of scientific descriptions; should the purported entity be shown to not exist, then the theoretical statements invoking it must be false.

A *successor theory* is simply one that replaces an earlier theory. To say that a successor theory is more accurate than its predecessor is to say that the successor entails either more truths or fewer falsehoods than its predecessor. The successor theory corrects the earlier theory, replacing false statements with true ones. It follows from Commitment 3, then, that science progresses over time.

Arguments for each of the commitments of realism can be drawn from the characterization of scientific inquiry found in the earlier chapters of this book. One argument for Commitment 1 draws on the notion of falsification (Section 3.4). In falsification, a hypothesis is validly deduced from a theory. A hypothesis not only follows from the theory, it is a proposition that can be tested by observation. That is, by making observations, we can determine whether the hypothesis is true or false. Because the deduction is valid, we can be certain that if the hypothesis is shown to be false, then one of the premises of the deduction must be false. Since the premises of the deduction were propositions of the theory, when a hypothesis is shown to be false, at least one proposition of the theory must be false. Of course, while a falsified hypothesis requires one make changes in the theory, but it does not tell one exactly what changes must be made. But changes must be made, and the whole point of falsification is to make the theory better, as Commitment 1 holds.

Now, scientists never rest content with rejecting a theory. The variety of responses to a falsified hypothesis support the idea that successor theories are more accurate than their predecessors (Commitment 3). Scientists often need to explore alternatives to find an improved theory—or set of theories—to then subject to further testing. In the simplest case, the successor theory will entail all of the old, true hypotheses, but not the false hypothesis. In more complicated cases, the new theory will give us new materials with which to understand the phenomenon. All of these responses are attempts to modify the theory so as to block the entailment of the false hypothesis. In revising or replacing a theory, then, there is thus a plain sense in which the successor theory is presumed to be more accurate than the theory it replaces (as Commitment 3 holds).

A further argument for Commitments 1 draws on the discussion of inductive support. A strong inductive argument makes its conclusion likely to be *true*. Therefore, treating theories as supported by inductive arguments presupposes a commitment to the idea that a theory is the sort of thing that can be true or false. Any by subjecting our inductive arguments to critical analysis—seeking out sources of error—we presuppose that the aim of scientific inquiry is theories that are true (as Commitment 1 holds)

The discussion of inductive support for theories also supports Committment 2. Contemporary scientific theories are subject to multiple kinds of tests. Theories that have survived such testing come to be well supported, with broad support from different kinds of evidence, using differing instruments informed by different background information, and multiple connections with related theories. For example, we saw in Section 4.4 how Newton's physics was supported by its ability to explain projectile motion, pendulums, and planetary motion. The theory deductively entailed regularities in each of these areas, and it did so by invoking "forces" postulated by the theory. Its empirical success across this range of applications provides powerful support for the forces postulated by Newton's theory. It would seem almost miraculous for it to have this range of successes were it very far off the mark.

Another argument for Commitment 2 draws on our discussion of causal modeling in Chapter 7. There we saw how scientists understand complex phenomena by constructing and testing causal models. In a causal model, a phenomenon is analyzed into parts, and the larger phenomenon is shown to be a product of the causal interaction of the parts. The parts of a causal model are often novel and difficult (if not impossible) to observe: quarks, bacteria, planetary orbits, and of course Newtonian forces. In this way, scientific inquiry is mechanistic. It postulates hidden mechanisms to account for the phenomena we observe.

Causal models are subject to testing, and as we saw in Chapter 6, experimentation is one of the best ways to identify causal relationships. Causal models are composed of simple causal relationships, and they therefore provide many possibilities for experimental intervention. Testing causal models, then, gives us good evidence that the entities postulated by the models exist. And since the entities correspond to the entities described by scientific theories, the practice of building and testing causal models presupposes a commitment to the existence of entities corresponding to (true) theories (as Commitment 2 holds).

Realism thus appears to be supported by the picture of science presented in this text. However, we have also seen considerations that mitigate against it. Those considerations motivate anti-realism.

9.3 The Antirealist Challenge

In response to the realist view, an anti-realist will argue that the realist has not fully appreciated some of the lessons learned in foregoing chapters. In particular, there is a tension between the realist's commitments and both the holism of confirmation (see p. 105) and the way that theory informs observation (see Section 8.4). The anti-realist argues that this tension undermines both the arguments for realism and the realist position itself.

According to confirmation holism, a theory is confirmed (or falsified) as a whole, not proposition by proposition. The characterization of falsification, in the first argument for realism, above, admits as much. A false hypothesis shows that *some* proposition of the theory must be false, but it does not tell us which one. If this is true, then it is too quick to conclude that falsification leads to theories that are more accurate. There is no guarantee that the adjustments made in the theory in response to falsification are true. For all we know, by making the adjustments that block the false hypothesis, we have increased the number of false statements in the theory, not decreased it.

There is, of course, a sense in which going through process of showing a hypothesis to be false, and then adjusting the theory so that the hypothesis no longer follows, increases the accuracy of a theory: the theory makes one less false prediction. The theory has become more *empirically adequate* in the sense that the hypotheses it entails are true. This suggests a more modest goal for science than truth. Scientific theories are useful because they facilitate observable predictions. An empirically adequate theory will support accurate predictions about observation, and with accurate predictions we can apply scientific results to the practical problems of the day. The aim of science, according to the anti-realist, is to describe the observations as well as possible, that is, to be empirically adequate. But, the anti-realist insists, we need not take the further step and suppose our theories to be true.

One might wonder about the point of confirmation and falsification on the anti-realist view. When a theory is confirmed or falsified, do not we conclude that it is true (or false)? In response, the anti-realist can appeal to the distinction between arguments and inferences, as distinguished in Section 3.1. The inductive and deductive *arguments* that figure in confirmation are important for establishing empirical adequacy. However, the anti-realist declines to make the *inference*. That is, he or she does not form the belief that the theory is true. Such a belief is not only unnecessary: since all prior theories have been shown to be false, belief in our current theory seems bound to be mistaken.

The distinction between argument and inference also supports an anti-realist response to the realist's arguments, above, that draw on the use of falsification and confirmation in scientific practice. The anti-realist will agree that scientific theories can be supported by the inductive and deductive *arguments* involved in confirmation and falsification. However, there is no need to *infer* that the theories are true (contra the realist Commitment 1).

Nor does inductive confirmation or falsification provide reason to believe the objects and mechanisms of scientific theories actually exist (contra the realist Commitment 2). The anti-realist holds that since the goal of science is to form empirically adequate descriptions, it is sufficient to treat the mechanisms of causal models and other theoretical entities as useful fictions. This is just to say that things behave *as if* there were electrons or gravitational forces standing behind them. Words like "electron" do not correspond to anything. We add such terms to our theory because they facilitate more accurate predictions. Theoretical terms thus improve the empirical adequacy of our theories. We need not make the further commitment that such entities exist.

As an example of the anti-realist attitude toward theoretical terms, consider the use of psychological language when talking about computers or other sophisticated machines. It is useful to treat the computer as "wanting" a password, "needing" a certain kind of input, or even "thinking" that it has opened the correct file. We do not believe that computers have such psychological states. Rather, by treating them as having such simple psychologies, we make it easier to work with computers. Similarly, by adding terms for unobservable entities, we make it easier to use our theories to predict observations.

The anti-realist thus rejects the first two commitments of realism. Treating the goal of science as empirical adequacy, rather than truth, also motivates a rejection of the realist's view of scientific progress. Science progresses, according to the realist, because successor theories eliminate the falsehoods of earlier theories. If we do not accept scientific theories as true descriptions of hidden mechanisms, then the realist picture of scientific progress is unmotivated. The anti-realist has an alternative picture, and we will explore it in the next section. In the meantime, we are in a position to characterize anti-realism in terms of three commitments that correspond to the commitments of realism.

Anti-realism: An anti-realist about scientific inquiry holds that

1. scientific inquiry aims only at empirically adequate theories,
2. the unobservable objects postulated by a empirically adequate theory are useful fictions, and
3. successor theories offer different perspectives than the theories they replace.

9.4 Scientific Change: The Anti-Realist Argument

From a realist perspective, scientific progress looks incremental. False statements are sought out and replaced with truths, like rotten timbers in a house. To make theories better, scientists postulate hidden entities and processes, and they propose new ways of conceptualizing familiar things. Over time, our theories become more and more accurate, providing deeper understanding of the world around us. As realists will admit, however, this picture of scientific progress is too simple. Science changes in fits and starts, not smoothly. Theories often have **anomalies**, that is, they have known, persistent problems that resist solution. Indeed, scientists may know that a theory they use is false, but continue to use it anyway. Nonetheless, the realist will insist, such roadblocks are temporary and ultimately surmountable.

To get a feel for the roadblocks, let us turn once again to the astronomical theories of both Ptolemy and Copernicus. Both theories primarily aimed at explaining the observed positions of the planets over time, and this includes their retrograde motions. Remember that "observed position" means the relative position of a planet against the background stars (and perhaps landmarks on the horizon). Now, it is relatively easy to predict where a planet will be tomorrow night, based on where it is tonight. Such predictions do not require a theory. Predictions about where a planet will be in a month or a year is much more difficult, and this is where a theory of the solar system is necessary.

While Ptolemy's theory did not match the observations perfectly, it did match them with an accuracy that, in some ways, rivaled contemporary theory. Using Ptolemy's calculations, it was possible to predict eclipses of the moon about one year in advance,[2] as well as times when two planets would appear in the same region of the sky. To achieve this accuracy, Ptolemy had to postulate some unexpected characteristics of the solar system. For example, while planets and the sun were thought to orbit the earth in circles, the earth was not at the exact center of their orbits. In Figure 2.3 on p. 9, notice the small × at the center of the circular orbit, and that the earth is just below. The × is the *eccentric*. Adding the eccentric to the theory was necessary to improve the observational accuracy. However, many scientists, including Copernicus, found the eccentric troublesome, as it seemed in

[2]To be precise, Ptolemy's theory could only predict when the eclipse was likely. An eclipse of the moon is not visible everywhere on earth, and calculating whether an eclipse would be visible at a particular place required information not available to the ancient world.

tension with their understanding of the physics of movement. According to Greek physics, the motions of the heavenly bodies were uniform circles. Natural motion was toward the center, that is, toward the earth. However, the earth is not at the center of the solar system. Each planet is slightly offset (and each planet's eccentric is different). As a result, Ptolemy's system was inconsistent with ancient physics. This was an *anomaly*: a way in which the theory was known to be incorrect.

Copernicus's research in astronomy was partly motivated by the unsatisfactory use of the eccentric, but also because of a widespread dissatisfaction with the accuracy of predictions obtained from Ptolemy's theory. By putting the sun at the center, his theory gave an elegant account of retrograde motion, and it eliminated the need for the eccentric. However, if the planets simply moved in circles around the sun, then the theory was much less accurate than Ptolemy. In order to increase the accuracy, Copernicus had to add epicycles as well. As we noted in Section 2.2, one of the attactions of Copernicus's theory to scientists like Galileo was that the heliocentric theory was simpler. Because Galileo rejected the Greek idea that the physics of heavens was different from physics on earth, he did not think that the uniform motion of the heavens had to be circular. Natural motion was in a straight line, and curved motion required an explanation. Hence, the addition of epicycles required explanation. What forces pushed the planets to move in epicycles? No answer was forthcoming from Copernicus or his contemporary scientists. For Galileo, Copernicus' epicycles were an anomaly.

Copernicus's view faced other challenges as well. According to his theory, the earth moved. Indeed, it had to be moving around the sun at an extremely high speed. This postulate seemed to defy common sense. After all, the earth does not *feel* like it is moving. Copernicus's contemporaries challenged his view with the following observation. When a stone is dropped from a tower, it falls straight down and lands at the base. But a stone dropped from a moving vehicle does not land at the base of the vehicle, but well behind it. If the earth and tower are moving, then the stone, it should fall away from the base of the tower. Hence there is an apparent consequence of Copernicus theory—a hypothesis—is not confirmed by observation.

Of course, Copernicus's theory *was* superior to Ptolemy's in other ways. We have already discussed the importance of the phases of Venus, and how Galileo's observation of the phases of Venus was an important piece of evidence for Copernicus and against Ptolemy (Section 2.2). However, even this was subject to challenge. The differences among phases of Venus are too small to see with the naked eye. Galileo could see them only because he had a telescope, which was a novel instrument at the time. Critics of Galileo rejected the observations he made. Why, they asked, are observatons made with this instrument reliable? Galileo could point to observations made on earth that could be checked. For instance, one could look at a faraway mountain and see a house that was not observable with the naked eye. By traveling to the mountain, one could confirm that the image seen through the telescope was accurate. But, the critics would continue, the heavens are different from the earth (on this point, Galileo and his critics disagreed), so why should the telescope make reliable observations about heavenly bodies? Since Galileo had no account for why his telescope worked, he was unable to definitively answer this criticism.

As we have told the story so far, it is clear that the change from one theory to another is more complicated than the simple realist picture shows. The simple realist picture of progress presents it as inevitable: observations force changes in the theory, forcing scientists to somehow jump to corrected theories. But, the historical episode portrayed here shows that observations do not force theory change. The lesson of confirmation holism should be taken into account to strengthen the point. Failed predictions, like the failure of the stone to fall away from the base of the tower, can be accommodated by adding further propositions to the theory.

The realist can accept the conclusion that observations do not force theory change. The realist of course, can respond by making their view of theory change more sophisticated, but we will hold that development until Section 9.5. The point for now is that the realist regards anomalies as mere bumps in the road of scientific progress.

The anti-realist draws a stronger conclusion from the historical situation summarized in Figure 9.1. Both theories have anomalies, persistent problems for

Ptolemy		**Copernicus**	
Confirmation	Anomaly	Confirmation	Anomaly
Accourate planet positions	eccentric	Accurate planet positions	Required epicycles
	Phases of Venus	Phases of Venus	Required telescope
Earth does not move, stone falls down			Stone should fall away

Figure 9.1: Comparison of Ptolemy and Copernicus

which there is no clear fix. While both are observationally adequate in some ways, both have inaccuracies as well. To choose one theory over the other requires making a trade-off of one anomaly or observational inaccuracy for another. It is a mistake, the anti-realists concludes, to suppose that one of these theories represents reality more accurately than the other.

The anti-realist's Commitment 3 expresses an alternative to the realist. For the anti-realist, the goal of science is to create empirically adequate theories. Both Ptolemy's and Copernicus's theories are empirically adequate, though both have defects. The anti-realist suggests that *any* alternative would have anomalies or observational inaccuracies as well. Therefore, there will always be more than one way to make sense of the observations. We should see the difference between the theories as differences in perspective. The observations can be explained and systematized in multiple ways. Ptolemy and Copernicus are simply different interpretations of what we observe.

9.5 The Realist Response

In the foregoing sections, the anti-realist has mobilized two lines of attack on the realist position. First, the process of theory change through falsification and confirmation is much more complicated than it has seemed. Falsification does not require any particular change to a theory, so there is no guarantee that theories get more accurate over time. And observations do not force theory change. It would have been just as rational to keep modifying Ptolemy's theory as it was to adopt Copernicus's theory. Second, while the arguments used to falsify and confirm theories are an important part of science, we need not make the inferences such argument suggest. In particular, we need not believe that the theories are true descriptions of hidden mechanisms. It is enough that we treat theories as empirically adequate and hidden mechanisms as useful fictions.

The realist, like any good philosopher, has some responses ready...

9.5.1 Realism and Theory Change

The realist can agree that the simple and optimistic picture of falsification leading to more accurate theories is too simple. It is true that kind of conundrum presented in Figure 9.1 has occurred in the history of science. But, the realist will argue, the kind of balance between successful predictions and anomalies does not last. In the case of Copernicus's theory, it took about 200 years for the heliocentric theory to be widely accepted by scientists. But it was accepted, and the reasons why are illuminating.

Let us first consider the anomaly of the tower. Again, the puzzle is why a stone dropped from a tower would fall straight down, rather than fly away, if the earth is moving at a great speed. Galileo had the answer, and it was later built into Newton's laws of motion. As Newton expressed it, moving objects have *intertia*: objects in motion tend to stay in motion. This idea was an important departure from ancient Greek physics, for which the natural motion of earthly objects was down. The idea of inertia had a wide variety of consequences for physics. With respect to the tower problem, the idea of inertia gave Galileo (and Newton) an answer. One can conduct the following experiment. While in a moving vehicle—a car or a train will do—drop a small object, such as a coin. You will find that it falls straight down; it does not fly to the back of the vehicle. This shows already that there is a problem with Greek physics. Galileo's explanation is that you, the car, and the coin are all moving in the same direction (let's call it "forward") together. When you release the coin, it doesn't stop moving forward, it continues with you and the car. So, from where you are sitting in the car, it seems to fall straight down. Similarly, the tower and the stone are traveling together before the stone is dropped. Once dropped, the inertia of the stone keeps it traveling in the same direction as the tower, and it falls to the tower's base. The anomaly of the tower was thus resolved by the new physics of Galileo and Newton.

The second problem to be resolved was Copernicus's need for epicycles. Without them, his theory was less accurate than Ptolemy's. This anomaly was resolved about 70 years after the publication of Copernicus's work by Kepler. Kepler's innovation was the propose that the planets moved in elliptical orbits, rather than circular ones. With this modification of Copernicus's theory, the need for epicycles was eliminated. Kepler's proposal was further supported when Newton showed that elliptical orbits were a consequence of his three laws of motion. With this demonstration, the heliocentric theory of the solar system was made consistent with the physics of the 17$^{\text{th}}$ Century.

Finally, the questions that might be raised about the reliability of the telescope, and thus about the significance of Galileo's observations of the phases of Venus, were answered. By the time of Newton, the way lenses worked to magnify images was well understood. Light is bent as it passes through the surfaces of clear materials of differing densities, such as air, glass, or water (note how a straight stick looks bent when partially submerged in water). Scientists studied how glass lenses would bend and focus light. This gave them a general account of how combinations of lenses could magnify an image without distorting it. Understanding how the telescope worked then allows one to be as confident of its effects when focused on the heavens as when focused on earthly things. There was therefore no reason to be suspicious of the observations made with the telescope. Indeed, as discussed in Section 8.4, understanding how a telescope works permits scientists to determine the conditions under which it will be reliable and when it is prone to error.

While the scientific developments took almost two hundred years, the anomalies of Copernicus's theory and the challenges to some of the supporting observations were resolved by the further development of physics and optics. Appealing to physics and optics to resolve anomalous consequences of astronomy is clearly not a matter of *ad hoc* tinkering to block falsification. The theories of physics and optics are supported their own bodies of observation. The anti-realist is correct to say that it is *possible* that modifications to a theory will introduce more falsehoods than it eliminates, this is extremely unlikely when theories, confirmed in different domains, are brought together. The success of science in combining theories from different domains into unified explanatory schemes, and thereby resolving the anomalies of previous theories, gives the realist confidence that science progresses, even if it sometimes progresses slowly.

9.5.2 Belief, Truth, and Scientific Knowledge

The realist defense of scientific progress in Section 9.5.1 did not touch the central contention of the anti-realist: that empirical adequacy, not truth, is the goal of science. There are at least two reasons why one might be skeptical of such a position.

First, contemporary science is highly integrated. The previous section showed how astronomy, physics, and optics fit together on several points. The phenomenon is pervasive. Neuroscience draws on results of chemistry, biology, and physics in its understanding of the properties of neural connectivity. Biology requires chemistry, but also physics to understand how animals can walk or fly. The sciences are not broken into separate fields, each supported by its own domain of observation.

According to the anti-realist, theories postulate unobserved entities and processes to facilitate prediction. But why, the realists asks, should all of these different domains of observation support theories that fit together in such productive and useful ways? It seems like it is an enormous coincidence that the properties of Calcium ions, an entity postulated by chemistry, should be just thing needed to help explain neural synapses, a process postulated by neuroscience. By contrast, the realist has a simple explanation: calcium ions exist with the properties postulated by current chemical theory.

The anti-realist has a possible response to this criticism. Chemists developed calcium ions to help explain their observations. Once a tool has been developed, it becomes useful for a whole range of purposes. Neuroscientists borrowed the idea from chemists, so it is no surprise that chemistry and neuroscience share ideas.

A second consideration that the realist might bring to the argument is that when we treat science as merely empirically adequate, we lose some of the intellectual benefits that scientific understanding can provide. Scientific inquiry challenges the way we think about ourselves and our place in the world. If we suppose that scientific theories are aiming at truth, then current science is our best attempt so far to speak the truth. We then need to figure out what the consequences of such discoveries might be, and what consequences they have for us. When Copernicus proposed that the earth was not at the center of the universe, it caused quite a stir. It deeply challenged the way that philosophers had understood our place in nature. We were no longer at the center of the universe; we were riding along on one planet among many, circling one star among billions. Or again, neuroscience proposes to understand human thought and consciousness in terms of neural function. In Chapter 11 we will begin to wrestle with the consequences of this scientific proposal.

If science aims only at empirical adequacy, and if different scientific theories are nothing more than alternative perspectives, then we can ignore the troubling consequences of a theory. If Ptolemy and Copernicus are not really disagreeing about how the world works—only about how best to make predictions—then it is hard to see why philosophers should get excited. If Ptolemy's theory best fit the philosophy of the age, then the philosophers should have used the Ptolemaic theory rather than the Copernican. To the realist, this seems like the easy way out.

More importantly, scientific discoveries speak to moral issues, not just practical problems. For example, many human societies have postulated that humans come in different types or races, and that these races have fundamentally different capacities. The argument that a group of people with specific skin color, face shape, language, or behaviors are too unintelligent to be educated has justified a range of morally repugnant practices, including slavery. Contemporary biology tells us that human characteristics are a consequence of our DNA. Insofar as we are different, it is because we have different DNA. Studies of DNA have found variation (which is to be expected, since humans do differ), but they have not found differences that correspond to race. According to biology, race does not exist. This scientific discovery has the consequence of undermining race-based arguments for enslaving or otherwise mistreating people on the grounds that they are of a different kind than "us."

If science aims only at empirical adequacy, it is difficult to see why the biological discovery that there are no races should have moral consequences. On an anti-realist view, the racist can can ignore scientific results by adopting an alternative interpretation of the observations. For the realist, there is a fact of the matter about which interpretation is correct, and we should align our beliefs with the facts. Of course, the anti-realist can reply that slavery and other forms of mistreatment can be shown to be wrong independently of scientific research. After all, the anti-realist might ask, if different races did exist, wouldn't slavery still be wrong? The anti-realist here is relying on a clear difference between facts and values, and in Chapter 10, we will explore this idea further.

9.6 Conclusion: An Enduring Debate

The arguments of this chapter have made progress on the issue, but they have not fully resolved the matter in favor or either the realist or the anti-realist. The realist has to admit that scientific change is a messy process, and at times we may not be in a position to choose between alternative theories. The anti-realist has to admit that there are often overwhelming reasons to reject one theory and adopt its successor. Even with these concessions to each other, the core commitments of each position remain intact.

Realism and anti-realism are two broad ways to understand the enterprise of science. The issue speaks to deep questions about what science is and why we engage in scientific inquiry. Buddhist scholars, too, will need to wrestle with this issue. Buddhism has a different way to understand what it means for something to be real or unreal, and these conceptions will influence the ways in which the realism debate will unfold in the Buddhist context. But those debates are for you, dear reader, not for this text.

Chapter 10

The Social Context of Science

10.1 Questions of Fact and Value

Scientific inquiry is conducted by humans, and it takes place in a social context. It therefore exists in an environment thick with desires, interests, and values. This chapter will ask how scientific knowledge relates to these human dimensions of scientific practice. In the Western tradition, especially over the last 200 to 300 years, science has been regarded as a distinct form of inquiry from the study of value. The common view has been that science studies facts, while religion and philosophy study religious, moral, political, and aesthetic value. In spite of this purported separation, scientific inquiry is often intertwined with questions of value. To inquire into the place of science in society, this chapter will explore how values are related to scientific inquiry. What sort of influence, if any, should values have on scientific inquiry? How does science relate to our projects of moral development? Can it show us how to live better lives? Can it contribute to our efforts to make the world around us better?

The foregoing questions presuppose a distinction between *facts* and *values*. Before proceeding further, it will be useful to understand these ideas as they have figured in Western thinking about science and society.

Facts are features of the world; states of affairs that exist independently of human opinion. That Drepung Loseling monastery is 4.3 kilometers north of Gaden Jangtse monastery is a fact. When we seek knowledge of facts, we use methods that are like those of scientific inquiry. We observe and we make inferences from these observations. We develop and test theories that explain the observations. By seeking out possible sources of error, our scientific methods are designed to reach true conclusions about facts.

Values are matters of what is good, right, or beautiful. To say that a person ought to be compassionate is to express a value. Laws and rules fall within the domain of value as well: that using a phone during an exam is forbidden also expresses a value. We also speak of personal or cultural values, which are things held to be good by one or more individuals. Values may be expressed directly, in talk of what is good or bad. Values may also be reflected in talk of obligation, of duty, and of what is prohibited or permitted. So, to say that using a phone during an exam is forbidden is to say that one ought not use a phone during an exam, and it is presumably forbidden because the use of a phone commonly makes for what is bad (distracting others, perhaps, or cheating by getting answers using the device). To say that the end of suffering for all sentient beings is the central good leads one to say that one ought to strive to end suffering for all sentient beings.

Judgments of fact and value work together in complex ways. Consider Argument 10.1. The conclusion, that this is a bad watch, is clearly an evaluation. The

> This watch is grossly inaccurate in its time-keeping and it is too heavy to carry about comfortably.
>
> ---
>
> Therefore, this is a bad watch.
>
> Argument 10.1

premises of the argument appear, at first glance to be factual. We can use scientific tests to determine the accuracy of its time-keeping and measure its weight. But notice that there is also an evaluative dimension to the premises. To say that the watch is too heavy to carry about comfortably might be thought to be merely factual in character. That is, it reports only that lugging it around causes discomfort in normal users. But given the purposes for which watches are used, it certainly implies that this watch had a bad feature. Moreover, accuracy (within some tolerance) is something that makes a watch useful as a watch. The argument thus appeals not only to facts, but the values of accuracy and portability that we associate with a *good* watch.

While scientific methods are designed to discover facts,[1] they are a poor way to discover values. If we want to know whether killing is wrong, it would not do to use some kind of sampling procedure and then make an inductive generalization. Finding that there are about 200 murders for every 100,000 people in Karnatika does not tell us whether murder is right or wrong. The judgment that the watch is a bad one, above, therefore seems to have two sources, one a judgment of fact and the other a judgment of value.

The central questions of this chapter, then, are about how judgments of fact and judgments of value fit together. As the previous two chapters have done, we will proceed by contrasting two positions. The first we will call the thesis of *value freedom*:

Scientific Value Freedom: Good scientific inquiry ought not be influenced by re-

[1] Clearly, this characterization of scientific inquiry presupposes realism. If one were an anti-realist, the language of "fact" would be have to be interpreted in terms of empirical adequacy. This chapter will adopt the realist language because it makes the exposition simpler. The anti-realist may also recognize a distinction between facts and values, but the characterization would need to be different.

ligious, moral, political, or personal values.

The second position denies that science is free from influence by values. That is, science is *value laden*:

Scientific Value Ladenness: Under some conditions, good scientific inquiry is influenced by religious, moral, political, or personal values.

The qualification "under some conditions" is important. The proponent of value ladenness does not hold that values should always influence science. She admits that some kinds of influence produce unreliable research. However, there are some cases where the inquiry is improved by the values that influence it.

10.2 Value Free Science

The idea that science should be value-free arises from a concern that views about how the world *ought to be* will influence our inquiry into how the world *is* in such a way as to make scientific inquiry less reliable that it would be without such influences. Historically, some examples of bad science hve been promoted and good science have been suppressed because scientists and others have let values influence their observations and reasoning.

Perhaps the most well known example concerns the prosecution of Galileo. When Copernicus's theory was published, the Catholic Church took an anti-realist stance. They admitted that Copernicus's theory gave a mathematically elegant way of calculating and predicting the observed positions of the planets, and that it provided the basis for a more accurate calendar. However, the Catholic Church rejected the idea that the earth moved around the sun. They insisted that Copernicus's theory was merely a clever device for predicting the motions of the planets, and it did not describe physical reality. A motionless earth at the center of the solar system was important to the Catholic Church, at least in part, because it symbolized the importance of humans and their central place in creation. The Christian Bible also used the idea that the earth did not move as an expression of the solidity of the Christian religion.

Galileo defended a realist interpretation of Copernicus's theory in his writings. This led the Catholic Church to put him on trial for heresy twice, once in 1616 and again in 1633. After the latter trial, he was found guilty. His book was banned, and he was forbidden to publish again. He was put under house arrest for the remainder of his life.

The story of Galileo illustrates how a commitment to particular values can reduce scientific reliability. The Copernican theory may have been surprising and even troubling, but the conflict between a well-founded scientific theory and a value should not result in the suppression of the theory. In order to maintain scientific reliability, then, many scientists and philosophers have proposed scientific value freedom: values should not influence scientific practice.

An immediate challenge to the idea of value-freedom is that without values, scientists would not know where to direct their inquiries. We saw in Chapter 2 that scientific inquiry begins with a problem of some kind. Just as saying that a watch is "too heavy" involves a judgment of value, to say that something is a "problem" invokes values. Ptolemy, Copernicus, and Kepler were all responding to the need for an accurate calendar for religious rituals and practical purposes. Historically, without these values, predicting the motions of the sun, moon, and planets would have not seemed such a pressing problem to be investigated. (In later stages, this concern may have become secondary to a kind of curiosity—the problem of getting to the truth. But, does this curiosity not itself involve a value?) The central point is that values, including values that are not merely a matter of valuing knowledge, weigh on scientists as they focus their scientific projects. Thus, in the modern world, we invest public money in scientific research because we believe that research will help us in some way—cure diseases, or provide a technology for feeding the world, or a technology for not polluting the world. Values are necessary to identify the topics of scientific research.

The proponent of value-freedom can respond to the foregoing argument by insisting on a division of labor. The scientists job, once the scientist is on the job, is to provide accurate and reliable theories that describe the facts. It is not the business of scientists to tell us how things ought to be or what we ought to do. It is then everyone's job—the role of politicians, religious leaders, philosophers, and all the consumers of scientific results—to determine what to do in light of the facts scientists discover. Of course, in doing so, everyone will need to be informed by values. (This community would include scientists, although not in their role as scientist, since for practical purposes scientists are consumers of scientific results.) Using the example of the watch, we might say that it is the job of the scientist to tell us how the watch works, and measure its accuracy and weight. It is up to the watch user to decide what weight and accuracy is valuable. Or again, using the example of drug research: it is the role of the scientist to determine the effect of a substance at various dosages (both the rates of recovery and any side-effects), and what processes produce the substance. It is the job of physicians and patients to determine what to do in light of those facts.

The thesis of value freedom has, until recently, been widely accepted by Western philosophers. Many scientists today would probably still espouse it. However, there are two kinds of argument that show that, at the very least things are not so clear. Section 10.3 will argue that the scientific study of facts cannot be so neatly pulled apart from the philosophical study of values. Section 10.4 will argue that the ethical demands on scientists has important consequences for scientific methods.

10.3 Value Laden Science

The thesis of value freedom is a tidy picture, but it is subject to an important objection. Consider Argument 10.2. Just as in the watch argument, we see an interplay

> Cigarette smoking causes lung cancer and heart disease.
>
> ---
>
> Therefore, cigarette smoking is bad for your health.
>
> Argument 10.2

of factual and evaluative elements. Whether cigarettes cause cancer and heart disease is something that can be determined by scientific inquiry. In the case of the watch, the values that make something a good watch seem rather separate from the science of how watches work. This example is different. Cigarette smoking is said to be *bad* for your health because it causes disease. But notice that the concepts of health and disease are partly evaluative. Insofar as science is to deliver information about health and disease, it is impossible to pull the values away from the scientific inquiry. Of course, science could deliver facts merely about the reaction of the human body to the various substances in cigarette smoke. But are these changes healthy, or diseased? Any way you slice it, to call some state of the body diseased or healthy is to evaluate it.

Scientific inquiry into health requires that some phenomena are recognized as healthy and others as unhealthy. Something is unhealthy because it causes pain or other kinds of suffering, or because it shortens one's life. Pain and suffering are disvaluable. The fundamental categories with which the science of health operates are therefore partly evaluative. They are only "partly" evaluative because they are also descriptive. Whether or not one has cancer is a fact, and it can be identified by scientific methods. The concepts of health and disease have two sides, we might say, an evaluative side and a descriptive side.

While not all sciences include concepts that have evaluative and descriptive sides, many do. We study poverty, unemployment, crime, depression, and wellbeing. The problems with which these scientific inquiries begin are already described in partly evaluative terms. Contrary to the perspective of value freedom, we cannot excise the values from such problem descriptions without destroying the very concepts with which we do science. Science is value laden. (Indeed, so is watchmaking, the proponent of value ladenness might say: it is a poor watchmaker who does not understand the need for accuracy.)

Value Laden Good scientific inquiry may include commitment to moral, political, or religious values.

If science is value laden, then can it contribute to our knowledge of ethics? Yes, but not alone. As we have already seen, scientific inquiry is designed to give us true descriptions of facts. The mixing of fact and value shows that when science studies value laden phenomena like health, it needs assistance from a discipline like philosophy. Reflection on the values implicit in the concepts of health and disease will make the character of the scientific problems clearer, and therefore improve the scientific inquiry. Conversely, science can give us a clear understanding of the causes and conditions of good health.

Something similar has already taken place in the Buddhist studies of compassion, or so the proponent of value ladenness might argue. Compassion is also a concept with both descriptive and evaluative dimensions. The character of compassion is investigated by Buddhist philosophy. Buddhist scholars have also investigated the conditions under which compassion can be cultivated. They have done so through observation and reasoning. As Buddhist science improves its methods by attending to possible sources of error, developing new kinds of observational test, and seeking to falsify (and thereby improve) previous work, we can hope that the study of compassion will continue to develop.

The discussion of value ladenness so far has been very optimistic. However, as the example of Galileo shows, the influence of values in scientific inquiry can be negative. Values can also make us blind to observations or inferences that might otherwise make our inquiry more reliable. For example, individuals within a species show substantial variation. Looking at a grove of mango trees, we can see that they have somewhat different shapes, and we describe these shapes in evaluative language: this mango tree is *stunted* from lack of water, that one was *misshapen* by storm damage. When we make such judgments, we are invoking the idea of a *normal* tree. The judgment that something is normal can be value laden insofar as, like the mango trees, it is saying something about how the trees ought to be. Like mango trees, humans show variation too and we judge some to be normal and others to be misshapen, so to speak. Until recently, there was a tendency in Western thought to treat men as the normal form for humans. This had unfortunate effects on medical research. Symptoms shown by men were taken to be the defining symptoms of many diseases. In many cases, however, men and women exhibit different symptoms for the same disease. This led to women being under-diagnosed and under-treated.

The example of disease symptoms shows that values can work in subtle ways to make scientific inquiry unreliable. Part of the problem is that, all too often, our sciences are unreflective about implicit values. Once we accept that good science might be value laden, we take it upon ourselves to be vigilant about those values that might distort scientific inquiry. The first step is to simply recognize that some scientific concepts have evaluative dimensions, and to make these evaluations explicit. We need to critically reflect on our evaluations—to be to challenge, defend, and sometimes revise our evaluations philosophically. But equally importantly, we need look carefully at they way in which they influence our observations and reasoning. The idea that humans have a special place in creation might seem innocuous. But it's consequences and logical connections to other commitments need examination. What, for example would God's concern for humans have to do with the arrangement of the solar system? Why would God's plan for creation, even were it to have involved a special concern for the humans created, need to have put the earth and human beings on it at the unmoving center of the physical universe? Further, should such tenuous connections have led historically to Catholic scholars ignoring some of the evidence that Galileo brought forward to support the Copernican theory.

Because the evaluative dimensions of our concepts are implicit, they are often difficult to see. Values are part of the conceptual materials with which we understand the world. Like eyeglasses, one does not notice that one is wearing them. Just as it is much easier to notice that someone else is wearing eyeglasses, it is easier to notice implicit values when they are values to which one is *not* committed. This means that we can improve the reliability of our scientific inquiries by subjecting them to criticism from a broad range of sources. In contemporary science, the process of peer review aims to provide such criticism. Before a scientific paper can be published, it is[2] reviewed by scientists who hold different theoretical views. The point about implicit values shows that in an ideal peer review process, different philosophical or religious stances might be relevant as well. The matter is complicated, because not all scientific domains give rise to biases in the same way. For example, it seems unlikely that biases arising from religious commitments would appear when studying the strength and related properties of various composite materials. Christians, Muslims, and Buddhists are likely to agree that strength and durability are good making features of the materials used in airplane wings. The evaluative dimensions of the concepts involved are more like a "good watch" than they are like "disease." At the same time, one might not have expected the Christian conception of humans' place in creation to influence theories about the shape of the solar system. A peer review process that encompasses broad diversity, then, is more likely to capture unexpected biases.

It follows that the reliability of scientific inquiry depends on the social institutions that support it. Peer review and similar processes require a community of scientists. These scientists have to be trained to seek out appropriate criticism of their own views and the views of others, even if they hold different theories and different values. The community then has to be organized in the right way to make such critical voices effective. There have to be channels of communication by which philosophical and scientific criticism can be communicated, and there must be a practice of thoughtful response to such criticism. The history of science in the West shows that such social conditions are very fragile. There are many examples of scientific communities being co-opted by political ideologies, charismatic leaders, or the lure of wealth and fame. If we are to conduct reliable scientific inquiries, we must be sure that the scientific community within which we work is critically reflective, open, and intellectually vibrant.

10.4 Research Ethics and Social Responsibility

The defender of value freedom proposed to give science and the study of values different, but complementary roles. In one's role as a scientist, according to the

[2] Or should be. Peer review does not always succeed in providing the appropriate level of outsider criticism.

thesis of value freedom, one should avoid value judgments. Making and defending value judgments is not the job of the scientist, but the job of the philosopher, or perhaps a question for democratic deliberation and decision. Section 10.3 gave one reason to be suspicious of value freedom. In this section, we will explore another: the ethics of scientific practice.

Like any human endeavor, scientific inquiry can be done in morally respectable and morally reprehensible ways. Separating the role of the scientist from the role of the person making value judgments seems to excuse the scientist from the moral consequences of scientific practice. But, like everyone else, scientists need to take responsibility for their actions, including scientific practice. This has two sorts of consequences for science. First, scientific methods need to be subject to moral scrutiny, especially when we are studying humans or animals. We will discuss the principles of research ethics in Section 10.4.1. Second, there is a larger set of questions about a scientist's responsibility for the consequences of their discoveries, and the place of science in society. We will explore these in Section 10.4.2.

10.4.1 Research Ethics

In the history of science, there have been a number of notorious episodes where scientific research let to egregious abuses of humans and animals. During the Second World War, the Nazis wanted to study the effects of very high altitudes on humans. So, they exposed religious minorities (Jews in particular) and other political prisoners to extreme cold and low air pressure. They also studied how the human body responded to damage by inflicting deliberate injury. Fortunately, horrifying experiments like these have been rare in the history of science. Nonetheless, the experience of the Second World War, as well as some other ethical lapses in twentieth century science, led to substantial reflection on the ethics of research. Research ethics is a rich and active area of scholarship. In this section, we will present some of the central ideas on which there is wide agreement.

A useful way to formulate the main ideas of research ethics is in terms of principles. Moral principles, in the Western tradition, are rules for correct behavior. When we are wondering how to design a study, we consult the principles and try to satisfy all of them. As with any set of rules, there may be conflicts. Ethical dilemmas arise when two principles indicate different answers to a question of what to do. Using principles to guide research ethics, then, means thinking through and resolving such dilemmas.

Principles of Research Ethics: Ethically correct scientific research should satisfy the following principles.

Informed Consent. Subjects of research must consent to participate on the basis of an understanding of the procedures, risks, and benefits of the study.

Harm. Study design must minimize the risk of harm to the participants.

Justice. The burdens imposed by the research must not fall unfairly on some part of the population.

The first principle, *Informed Consent*, requires that participation in a scientific study be voluntary. Most scientific studies ask the subjects to undertake some risk. An experimental drug, perhaps, will have side effects. There is nothing wrong with asking someone to take a risk for the sake of a scientific inquiry. It is noble for a person to volunteer to help others by risking injury to him- or herself. In so doing it is important for the person who is recruited into the experiment to be given the best present information about the likely benefits and harms of taking part in the experiment. The subject needs to know the personal risks of injury, such as the side effects of a drug. The subject also needs to know the possible benefits, whether those are benefits to themselves (for instance, if the experimental drug is a successful cure), to other patients, or to society at large. In short, the subjects need to know what they are signing up for. This information makes it possible for the subjects to make a truly voluntary choice.

The need for informed consent has important consequences for the design of research studies. When we study human behavior, for instance, we know that people behave differently when they are being watched than when they are in private. As a result, scientists sometimes wish they could be invisible. There is a temptation to deceive the subjects, either by not letting them know they are being studied or by misleading them about the study. According to the Principle of Informed Consent, such a study would be unethical. People do behave differently when they are being observed, and this means that scientists need to think creatively about the real possibilities of error so as to find ways of gathering information that is consist with informed, voluntary participation in the study.

The *Harm* principle requires that we choose methods with the lowest risk of those available. If methods A and B would produce similar results, and if A is less risky, then it would be wrong to do B. Designing a study of humans requires the scientist to carefuly think through the various kinds of harm that might be done to the subjects by alternative experimental designs. Not all harms are physical. A study might make subjects anxious or depressed, for example. Also, many kinds of research gather information that the subjects might want to keep confidential. Breaching confidentiality is an important risk of much research in the social sciences. Note that the Harm Principle asks scientists to minimize *risk*, not minimize harms. So, there might be the possibility of substantial harm, but in such a case, the likelihood of occurrence should be very small.

The *Justice* principle prohibits researchers from exploiting vulnerable subjects. We do not want a system where one group of people disproportionately bears the burden of the risks of research while another group reaps the benefits. For example, some medical research uses subjects from in developing countries. Since existing health care is poor and incomes are low, subjects are more willing to participate, or are willing to participate for less compensation, than are subjects from wealthy countries. The drugs tested, however, are very expensive and will not be affordable to the subjects of the drug test. It is unjust to use such test subjects, when they are bearing all of the risk and getting none of the benefit of the research.

We have been discussing research on humans, but research on animals also requires ethical deliberation. Animals are not able to give informed consent. However, a version of the Harm Principle applies. In animal research, we want to be sure that the risk of harm is absolutely necessary, and that every attempt has been made to minimize pain and discomfort. While considerations of justice do not have the same force, the Justice Principle does apply. The likely harms and benefits to animal subjects of experiments should be weighed against the likely harms and benefits (for animals and humans) of the possible experimental results.

Thinking through the ethical dimensions of one's research is a difficult task. These principles are only a starting point. Much depends on the details of the proposed methods and the context of the study. The larger lesson of the discussion in this section is that considerations of ethics are a crucial part of scientific methodology. Hence, we have another reason to suppose that science cannot be value free. One of the concerns of the defender of value freedom, again, is that letting values permiate science will undermine the reliability of our methods. Careful consideration of the sources of error is a crucial part of the development of ethically responsible methods. It should be clear, then, that ethical considerations in research design do not reduce reliability.

10.4.2 The Social Consequences of Scientific Inquiry

We live in an era of amazing technological capacities. Scientific inquiry has made it possible for us to travel quickly around the globe, or into space. The telephones in our pockets are high powered computers that not only permit instant communication, they map our position, predict the weather, and (of course!) show us cat videos. These developments have come at a price. High speed travel is among the factors that have damaged the planet and its atmosphere. The industry that produces parts for cellphones also produces dangerous forms of pollution and toxic waste. Neither the benefits nor the harms of technology would be available without the imagination and research of scientists. To what extent, if any, are the scientists who made technology possible responsible for the harms it has caused?

The question is a difficult one, and it probably does not have a general answer. To see some of the complexities involved, consider the story of Leo Szilard, a physicist whose research was crucial for the construction of the atomic bomb. In the 1930s, Szilard developed the idea of a *chain reaction*. Some atoms, like uranium atoms, are unstable and will break apart or *decay*, forming two more stable atoms. When this happens, energy is released, along with particles called neutrons. These neutrons can trigger further atoms to decay. Szilard showed that if sufficient uranium of the right type were put together, the decay of atoms would increase at a rapid rate. Should the chain reaction happen fast enough, an enormous amount of energy would be released in an explosion. This was the root idea of the atomic bomb.

In 1939, Szilard believed that German scientists were working on the possibility of building an atomic weapon. World War II had just begun, with the British, French, and Americans fighing against the Germans, Italians, and Japanese. Szilard wrote a letter, also signed by Albert Einstein, urging that the United States begin investigating the possibility of building an atomic bomb. As a result, the United States began the Manhattan Project, which ultimately produced the bombs that fell on Hiroshima and Nagasaki.

Later in the war, it became clear that the Germans were not successful in producing an atomic weapon. Indeed, Germany surrendered in 1945 without having produced one. At this point Szilard came to believe that the United States should *not* build and use atomic weapons. The huge explosion caused by an atomic weapon would produce an enormous number of civilian deaths, since its damage could not be limited to military targets. He tried to convince United States officials to stop or change the program, but he was ultimately unsuccessful.

Szilard's story is a complicated one. In some sense, since he developed the idea of a chain reaction, he was responsible for the creation of the atomic bomb. And he certainly felt responsible. On the other hand, as he well knew, others were working on the same problems in physics. If he had not developed the idea, someone else would have. Indeed, he was aware of this as well, which is why he initially urged the President of the United States to support scientific research on atomic weapons. In many ways, the development of atomic weapons was out of Szilard's control, even though he was responsible for a crucial piece.

As we have emphasized in a number of places, science is a social enterprise. Scientists work in large communities, they share information, and they build on one another's ideas. Hence, even though we give credit to individuals when important discoveries are made, the discoveries are typically as much a group product as an individual one. This makes it difficult to assign moral responsibility to an individual. As individuals, we are responsible for our actions. Szilard had the choice to stop working on the relevant part of physics, once he realized what the consequences were. He chose a different and somewhat more dramatic path by trying to change the course of the whole project. He can only be primarily responsible for his own actions, and he did not alone develop the atomic bomb. Therefore, it would be inappropriate to blame him for its development, just as it would be in appropriate to give him full credit for the beneficial applications of nuclear technologies.

While the question of the responsibility we have for the benefits and the harms of science is ultimately not an individual one, because scientific inquiry requires communities, responsibility must be addressed at the community level. The Manhattan Project would not have existed were it not for the resources of the United States government. Now, the United States was at war at the time, so the question of whether it should have pursued the Manhattan Project is a complicated one. The point for us to consider is how, as a community, we should influence the course of science. Should democratic processes be used to decide how public resources are allocated to scientific projects? Should democratic processes be used to assess the risks and benefits of a technology? Should there be limits on the commercial development of technology? These are important questions, and they point beyond the limits of this text. Again, the lesson is about the relationship of science to society: this is a case where values important to a society can and should influence scientific inquiry.

10.5 Conclusion

In this chapter we have seen that while the thesis of value freedom is attractive, it is ultmately untenable. The problems with which scientific inquiry begins are important because of the things that are important to us, including our health and welfare. We characterize parts of our inquiry in terms that have both descriptive and evaluative dimensions. And ethical considerations should shape our choice of methods. Science is permeated with values, and this leaves us with a responsibility. Values can undermine the reliability of science, leading us to reason poorly and ignore observations. Our responsibility as scientists, then, is twofold. We must make the evaluations that are implicit in our science explict, and be prepared to defend them philosophically. We must also be vigilant for the negative effects that moral, political, or religious commitments can have on scientific inquiry, and be ready to modify our views.

Chapter 11

Consciousness and The Limits of Science

11.1 Are There Domains that Science Cannot Reach?

In the earlier chapters, we have focused on the kinds of knowledge that science can produce, and the kinds of inquiry by which science can provide such knowledge. This chapter looks at the other side, Are there objects or subject matters concerning which science is ill-equipped to generate knowledge? There are two aspects to this question. We might ask about objects: are there any kinds of objects that cannot be studied with scientific methods? We might also ask about modes of knowledge: are there ways of knowing that could not be part of scientific practice?

When we ask these questions, we must be mindful that science is a quite varied and multi-faceted practice. Not all science looks like astronomy and physics. In this text we have tried to give a flavor of some of these differences by discussing a variety of examples. Even so, there are vast areas of science we have ignored. In particular, we have given no examples from the study of human thought and behavior, or of social processes and institutions. While the objects of these sciences require methods that are in some ways similar to those we have discussed, appropriate methods also have distinctive features. Similarly, mathematics and computer science are surely sciences, and yet they are different from other sciences in both their object and ways of knowing. So, when we ask whether there are kinds of knowledge, or objects of knowledge, for which science is ill-equipped, we must not assume there is a single method or kind of object that all sciences share.

Phenomenal consciousness—the awareness or subjective feel of something—is a prominent and much discussed example of something that might elude scientific understanding. It is sometimes said that it is not the job of science to taste the soup, but to explain the processes by which the taste is produced. Science seems able to explain why soup tastes as it does. We can describe and analyze the different chemical compounds and how they interact with sense receptors on the tongue and in the nose. And neuroscience can characterize the neural mechanisms which give rise to experiences of more or less sweet, sour, bitter or salty. What about the experience itself? There seems to be a qualitative feel to the taste of the soup. Like the visual experience of the red of a monk's robe, or the smell of a rose, the taste of the soup seems to be private or subjective in the sense that I have it and I cannot share it. What, if anything, can science do to explain these subjective, private, or qualitative feels? This is the problem of phenomenal consciousness.

The problem of phenomenal consciousness is an issue of great contemporary interest among scientists and philosophers. In this chapter, we will investigate it in some detail. Before diving in, it is worth noting that science has faced similar questions of limits in the past. From the time of ancient Greek science through the time of Newton, the study of life was particularly challenging. Physics and astronomy were the most well developed sciences in the ancient world, and by the seventeenth century, chemistry was joining them as a mature science. All of these sciences were "mechanical" in the sense that they analyzed phenomena into causal systems of interacting particles and forces (much as we illustrated in Chapter 7). Prior to the nineteenth century, living organisms seemed to have a number of properties that could not be analyzed into such systems. Historically, some philosophers and scientists argued that scientific understanding was limited to non-living systems.

Living things do several things that non-living things do not do, and these seemed difficult to fit within a mechanical system. Living things are born and grow. Some animals, like salimanders, can regenerate a tail or limb if it has been cut off. Some small creatures, such as the flatworm and the Hydra, have the capacity to reproduce themselves even if they are cut into pieces. Such strange behavior seemed to defy analysis into the causes and effects with which ancient Greek and early modern scientists were familiar. Moreover, as we explored in Section 7.3.1, the organs, limbs, coloring, etc. of living things have purposes or functions. Teeth are for chewing, eyes are for seeing. It did not seem possible to reconcile purposes with mechanical causation, since the reason why an animal has such features lies in the future. A newborn cat has claws for catching mice, but all of its mouse chasing is yet to happen. Such puzzling and difficult to explain features of living things made it plausible, up until the nineteenth century, that life might be out of reach of science.However, great advances were made in the nineteenth and early twentieth century. As we saw in Section 7.3.1, Darwin's theory of evolution gave us a way to understand the natural purposes of animals. In addition, there were breakthroughs in the understanding of reproduction, nutrition, respiration, and the other processes associated with living things. These phenomena thus were fit into a causal system and are now the subjects of scientific study.

With respect to phenomenal consciousness, contemporary neuro-scientists are much like their colleagues in biology were in the nineteenth century. We are beginning to gain some important new insights into the brain. But while we can envision a science of experience of a sort—one which analyzes the complex processes involved in sensory perception, and (for example) explains why one substance tastes sweeter than another—it yet seems as though there may be something fundamental about consciousness. While we now agree that we explain life without anything left over, will we ever be able to explain consciousness without something left over?

11.2 The Visual System and Peceptual Consciousness

The question of whether science can explain perceptual consciousness clearly depends on the state of the science. So, let us begin by looking at some of the science relevant to subjective experience. Since perception of color is such a common example of a phenomenal state, and because much is known about the visual system, let us consider the visual system. In its broadest outline (Figure 11.1), light re-

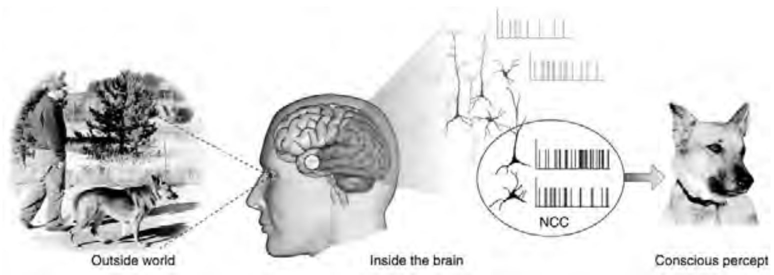


Figure 11.1: Consciousness of Physical Objects[1]

flects off of objects in our environment and into our eyes. Our eyes focus the light onto specialized cells, and these cells send information into the brain. We have a conscious experience of color when particular systems of neurons are activated.

The story begins with the basic physics of color. Color is a property of light, and the contemporary understanding of light treats light as an electromagnetic wave. These waves are like the vibrations of a string that has been pulled tight. A string on a musical instrument will make higher and lower pitches as the string vibrates faster and slower. The faster vibrations are shorter wavelengths, and slower vibrations are higher wavelengths. Similarly, light can have higher and lower wavelengths. These wavelengths correspond to colors. Whenever we see something red, our eyes are detecting wavelengths in the neighborhood of 650 nanometers.

The light that comes into your window in the morning, or that comes from a typical light bulb, is known as "white" light. Issac Newton demonstrated that white light is contains the full range of wavelengths visible to the human eye. When the narrow bands of light, corresponding to a specific colors are combined, the result is white light. Newton showed that when white light shines through a triangular glass bar, known as a prism, it breaks into separate colors, as Figure 11.2 illustrates.

[1]Christof Koch (2004) *The Quest for Consciousness: A Neurobiological Approach*, Englewood, Colorado: Roberts and Company Publishers, p. 16. Available for use thorough Wikimedia Creative Commons license.

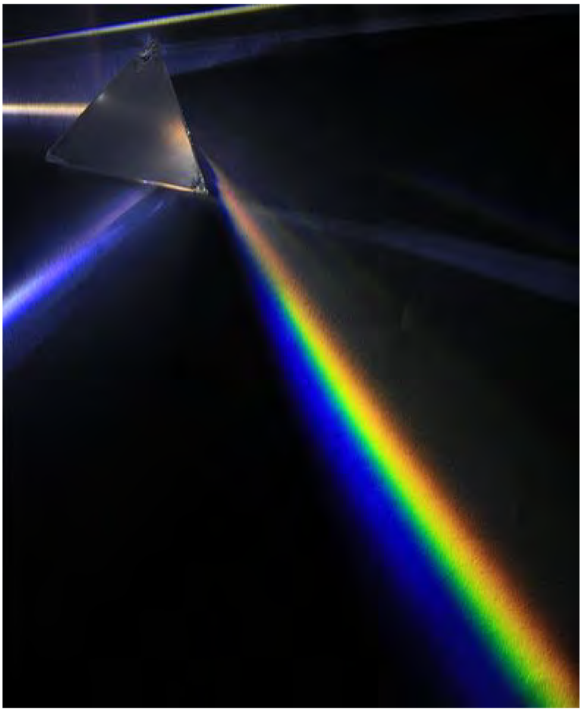


Figure 11.2: The prism breaks white light into bands of colored light[2]

The angled sides of the prism separate the wavelengths so that each is seen in its pure form. Rainbows seen after a rainstorm are a the same phenomenon; the raindrops act as prisms to break the white light of the sun into its components.

Colored light, then, is light that has a specific wavelength. Why then do *objects* look colored? The answer is that our eyes respond to the light that reflects off of objects. Materials reflect light of some wavelengths and absorb other wavelengths. The robes of a monk look red because because, when these robes are struck by white light, they reflect much of the light in wavelengths near to 650 nanometers and absorb much or most of the light of different wavelengths. Thus, the robes look red (in normal conditions) because the light available to our eyes is predominately red light.

Our eyes work very much like the camera in your cell phone. In a digital camera, like the one on your phone, the lens focuses an image onto an array of tiny detectors. Each detects the light and sends this information into the phone to be recorded as one tiny part of the picture. Like the lens in a camera, the lenses of our eyes focus an image onto a specialized array of detectors, in this case, neural cells. There are two kinds of cells scattered across the back interior surface of our eyes, on which light is focused: “rods” and “cones.” Rods respond the intensity of light, that is, they are activated in degrees corresponding to light levels. Rods can be activated by light at lower levels than would produce a response from cones.

[2]Image in the public domain. Source:D-Kuru/Wikimedia Commons

Cones are responsible for color vision. (Thus, in low light conditions, we tend to see things around us in tones of gray, for the cones are not in play.) There are three kinds of cones in the human eye, and each kind is attuned to a different range of light wavelengths. This means that they respond only when light within their range hits the back of the eye. So, when we see red, light has been focused on the back of the retina where it then stimulates those rods that are sensitive to light in the range close to 650 nanometers.

These events in the eye, structured as they are by the patterned activation of the rods and cones, are not enough to account for the perception of an object or of colors. Were the neurons extending from the eye to the rear of the brain to be severed, one would see nothing. So, things have to happen in the brain. Neural pathways connect the retina to areas at the back of the brain called the "visual cortex." This part of the brain exhibits patterns of activation across layers of neurons that are structured like those on the retina. When the retina is stimulated, the brain begins a cascade of processing, as patterns of activation at one layer of brain tissue cause transformed patterns of activation at layer upon layer of brain tissue. Figure 11.3

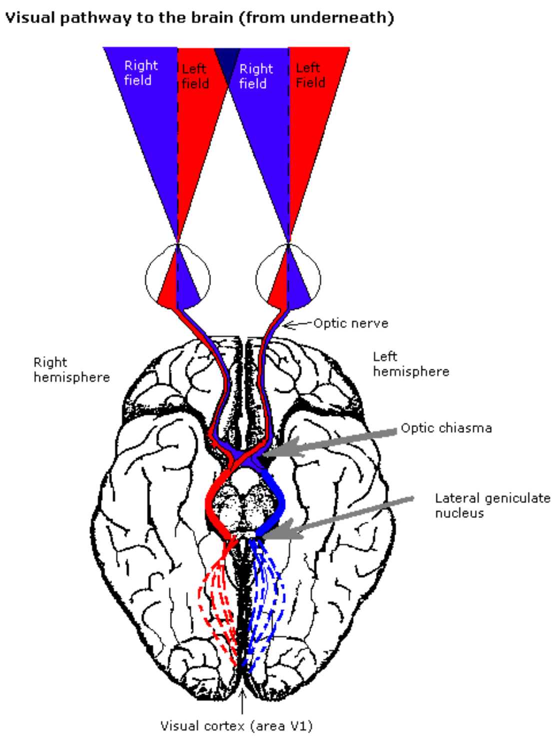


Figure 11.3: Visual Processing in the Brain[3]

provides an image of where visual information is processed in the brain. Some of the processing in the visual cortex involves contrasts. There are areas dedicated to

[3]Image used under GNU Free Documentation License, Version 1.2. Source: from Wikimedia Commons

identifying red–green and blue–yellow color contrasts. Other areas contain clusters of neurons that respond only to specific color ranges. Suffice it to say that there is a lot going on.

One may think of these aspects of visual processing as dimensions of human color experience, so that a given color experience can be thought of as located in this space of subjective quality dimensions. For example, the look of a piece of cloth might be more yellow and less red than another, but about the same as a third. That is, one can locate the experienced color as being more or less along a dimension, such as more red than blue, more green than yellow, and so on. In this way the scientist can study how visual experiences arise out of the system of neural processing that begins with the activation of rods and cones by light near a wavelength of 650 nanometers and ends with patterns of activation in the the visual cortext corresponding to the sensation of red.

So, according to this story, the sensation of color depends on differential responses of the visual cortex to the information from the cones. In turn, the cones are responding to different wavelengths of the light reflected into our eyes. And the light reflected into our eyes comes from surfaces that have absorbed some wavelengths and not others. This is the causal model of color vision. When a person has the subjective experience of seeing a particular color in color space, their brain is exhibiting a particular pattern the space of possible patterns in the visual cortex. It seems, then, that the subjective experience of "red" can be studied by building causal models of the visual system.

11.3 Monocrhomatic Mary

Some Western philosophers are unsatisfied with the account of vision provided in Section 11.2. The scientific story leads up to the experience of red, but, they contend, it stops short of explaining phenomenal consciousness. The pattern of stimulation in the visual cortex is certainly important in coming to have a conscious experience of color, for without it, we would have no experience of color. But, the pattern of activity in these neurons is not yet conscious subjective experiences. The problem is that the knowledge of causality does not seem to give us an account of the experience. Phenomenal consciousness, they argue, is not an object of scientific knowledge.

To support the idea that phenomenoal consciousness escapes scientific study, recent philosophers[4] have discussed a fanciful example that sharply focuses the problem of phenomenal consciousness. This story is called "Monochromatic Mary." Mary is a exceptionally brilliant and accomplished neuroscientist who has always lived in a black and white environment. She has never looked upon a single colored object. Her computer monitor and books are all in shades of gray. (Admittedly, this is a bit far-fetched. Don't ask us how she has never seen the color or her own

[4]The original presentation of the Monochromatic Mary story was by Frank Jackson in "Epiphenomenal Qualia," *Philosophical Quarterly*, 32: 127–136 (1982).

skin, hair, or eyes.) She has access to scientific equipment and all of the research of others. Suppose she knows the kind of story told in Section 11.2 in all of its fine-grained scientific detail, including the micro-processing of the different parts of the visual cortex. It would not be exaggerating to say that she knows volumes and volumes!

Mary has never left her lab, the story continues, but she has cameras that show her the outside world. She knows, in particular, that there are rose bushes outside of her lab. She knows that these flowers reflect light in the range close to 650 nanometers, and so she knows that normal humans will process this light as fitting into the "red" zone of their subjective color space. Now suppose that she is able to leave her lab and enter the normal world. She opens the door and sees a large, red rose in front of her. When she sees the red rose for the first time, does she learn anything new? Will she gain new knowledge when she sees something red? Is there some object of knowledge or some way of knowing that was not available to her in the lab?

The story of monochromatic Mary has been discussed extensively by philosophers and neuroscientists. The goal of this chapter is not to try to explain all of the lines of argument that have been developed, nor to argue for one answer or another. We want only to give a sense of the debate, and hope that you will be eager to debate it yourselves.

11.3.1 Mary Learns Something Nonscientific

One common response to the story of Mary is to say that she learns something new, and she learns something of a nonscientific kind. This intuition is the basis of arguments that phenomenal consciousness will not yield to scientific approaches to knowledge. At the beginning of this chapter, we distinguished between objects and modes of knowledge. The intuition that Mary's story shows phenomenal consciousness to be outside of scientific knowledge involves both objects and modes of knowledge, and each can be developed into an argument for the conclusion that scientific knowledge has interesting and important limits.

Some philosophers think of experienced color as a qualitative feel or experience having—what they call—an intrinsic "what it is like." This intrinsic qualitative feel is an object, one might contend, that can only be apprehended by a being capable of phenomenal consciousness who is undergoing an episode where this qualitative property is instanced. Let us call this kind of thing a "perceptual quality." A perceptual quality is the object of conscious perception, and it has a feel or experiential quality. The perceptual quality can only be known by having the relevant experience.

When Mary leaves her room, it is natural to say that she has a new experience: she sees red for the first time. If redness is understood as a perceptual quality, then Mary becomes acquainted with a new object of knowledge. But, the story stipulates that Mary had complete scientific knowledge when she left her room.

So, one might conclude, there is a kind of object—perceptual quality—that is not an object of scientific knowledge. This argument is presented as Argument 11.4.

Perceptual qualities are objects of knowledge.

Before leaving her room, Mary has complete scientific knowledge of the visual system.

After leaving her room, Mary becomes acquainted with a new perceptual quality.

Therefore, there are some objects of knowledge (perceptual qualities) that are not objects of scientific knowledge.

Argument 11.4: Mary experiences a new object of knowledge

Moreover, Mary seems to acquire this knowledge in a different way than scientific knowledge is acquired. As we have been emphasizing throughout this text, scientific knowledge is acquired through inference and observation. Upon seeing the red rose, Mary's knowledge of the perceptual quality of red does not involve any inferences. There is no need of premises or arguments (deductive or inductive), one might say, in order for her to grasp what it is like to see this color (red). She simply sees it. Does this mean that she observes it? In a sense, obviously yes. But when we look closely, this is not the sense of observation that supports scientific inquiry.

In Chapter 8 we discussed two ways of thinking about observation, one that was more objective, while the other was subjective. We argued that the objective approach to observation was better for scientific inquiry. If Mary's new experience of red is an observation, it is an observation of a subjective sort. The experience is something she *has*, and because she has this experience, she has a direct acquaintance or apprehension of the perceptual quality of redness. Scientific observations of red objects need involve no such qualitative experience of seeing red. For example, the astronomical observation of red planets in another solar system might involve instruments measuring the wavelengths of light reflecting off the planet, and this might be represented in a printout. Similarly, before Mary left her room, she had scientific observational evidence that the roses were red because she knew they reflected light in the range of wavelengths close to 650 nanometers.

One might conclude, then, that when Mary exits her room, she is learning about the redness of the rose in a new way. Her knowledge depends on a mode of knowledge different from scientific observation. We might represent the argument as Argument 11.5

Arguments 11.4 and 11.5 fill out the intuition that Mary learns something new when she leaves her room, and that her knowledge is somehow different from that attained through scientific inquiry. Philosophers who accept these arguments contend that they show an important limitation to scientific knowledge. If, as they

> Scientific observation is not a personal, subjective experience.
> When Mary leaves her room, she learns about the redness of the rose through a personal, subjective experience of red.
>
> ---
>
> Therefore, when Mary leaves her room, she learns about the redness of the rose in a nonscientific way.
>
> Argument 11.5: Mary deploys a different mode of knowledge

suppose, the objects of sensory knowledge are perceptual qualities, and these must be known with a non-scientific kind of observation, then science seems unable to encompass some parts of human experience. Of course, these conclusions are consistent with the idea many aspects of human experience *can* be and are studied by science, as the discussion of the science of vision in Section 11.2 shows. Neuroscience can give us detailed accounts of perceptual similarity and difference. According to proponents of Arguments 11.4 and 11.5, while such studies explain the neural basis of conscious experience, they stop short of a scientific understanding of the experience itself.

11.3.2 Mary Learns Something Scientific

Many philosophers take issue with Arguments 11.4 and 11.5. While they agree that Mary learns something when she leaves the room, they disagree that this new knowledge is somehow outside of the purview of science. Argument 11.4 assumed that when Mary saw the red rose, she was immediately acquainted with a perceptual quality. And again, a perceptual quality is conceived as a kind of object with an intrinsic feel or look. The argument asserts that such object are not the sort of thing studied by science. One might take issue with this assumption. Section 11.2 suggested that one might have a scientific study of subjective experience. The neuroscientific account shows how subjective experiences might have locations or relations in multidimensional quality space, and and this quality space is constituted by the neural system's capacity to make discriminations, such as the red–green distinction. Moreover, the neuroscientific account shows how visual experiences are the causal outcomes of interactions between organisms' sense receptors and their physical environment. If this is correct, that these subjective qualitative experiences are not themselves a new sort of thing unknown to science. Of course, directly grasping their intrinsic (as opposed to relational) "what it is like" might yet require undergoing the various experiences, a point to which we will return below.

Philosophers who hold that subjective experience is studied by the neuroscience of vision (and other senses) denys that perceptual qualities are a kind of object lying outside of the causal realm. These philosophers thus take the neuroscientific account of vision to be like the evolutionary account of natural purposes (Section 7.2). Earlier scientists took natural purposes to be a special kind of property, and they puzzled about how purposes might be explained. The evolutionary account built a causal model. Causal models take a phenomenon and show how it can arise from the causal interaction of various parts. For a skin (or fur, feathers, scales, etc.) color to have the purpose of camouflage is for it to have protected the organism's ancestors from predation (or, in the case of a predator, protecting it from being seen by its prey). The purpose of camouflage is not a new kind of thing over and above the causal process of evolution. Similarly, this position argues, the phenomenon of experiencing red is nothing more than having one's visual system in the right state, and the state of the visual system is a possible object of scientific knowledge.

How, then, does this position deal with the idea that Mary learns something new? One might agree that *something* new happens when Mary leaves the black and white room. The question is *what*? The new experience need not be understood as a new kind of object. According to the view we are exploring here, the perceptual quality, this property of subjective experience, has already been the object of scientific study. It has been studied as a location of an experience in a multidimensional quality space. But there is something new happening in her brain when she leaves the room and sees the red rose. Her brain has never been in the state that your brain is in when it sees a rose. So, in that sense, something new happens to Mary.

A proponent of Argument 11.5 might reply that the sort of observation Mary makes of the rose is not a properly scientific observation. When Mary sees the rose, she is having a subjective experience, and as argued in Chapter 8, scientific observation is objective. The philosopher who rejects Argument 11.5 may agree with the first premise(that scientific observation is not a personal, subjective experience). The difficulty, the philosopher might argue, is found in the second premise (that when Mary leaves her room, she learns about the redness of the rose through a personal, subjective experience of red). As just noted, when Mary leaves her room, her eyes are struck by light in the range close to 650 nanometers, the cones in her eyes transfer a signal to her visual cortex, and her visual cortex processes the signal, placing the signal in a field of contrasts, including the distinction between red and green. When this happens, she has detected that the rose is red. According to the objective conception of scientific observation (p. 96), she has made the sort of observation that would count as scientific.

Now, the proponent of Argument 11.5 might contend that there is more to human perception than detection. The person not only must be sensitive to the light and able to process it with his or her visual cortex, the person has to *recognize* that he or she is seeing something red. To recognize something is an ability. This ability involves many elements, including the ability to distinguish red things from things of other colors, see that ripe tomatoes and monks robes are more similar to each other than either is to the sky, understand that a white piece of paper in red light will look red, and so on. When Mary was in her room, she could do these things, but only by using instruments to detect the wavelength of light and her knowledge about how humans respond. When she herself is detecting the red light, she does not need the instruments. She now has the ability to recognize red things just by looking.

If Mary can now recognize things just by looking, then there is a sense in which Mary acquires a new way of knowing (a new mode of knowledge) when she leaves her room. Is this new mode of knowledge a form of non-scientific knowledge? The answer here is "yes and no." The practice of science requires a broad range of recognitional abilities. You will soon use microscopes in your biology class, and you will then see that it takes some time to learn to recognize different organisms. Looking through a microscope is something that takes training and practice. Just like Mary's recognition of red, scientists need to learn to recognize a broad range of specialized phenomena. Abilities to recognize and observe are a central and ineliminable part of scientific knowledge. At the same time, the kind of recognition Mary acquires is unlike scientific recognitions because it commonly arises well before a person begins to develop distinctively scientific skills. Folk commonly learn to perceptually discriminate various colors quite early. Because of this, it is not generally thought of as a distinctively scientific bit of know-how. The philosopher who wants to reject Arguments 11.4 and 11.5 can thus admit that Mary does acquire a new mode of knowledge: the ability to recognize red things by just looking. However, this does not show that science has a limitation.

11.4 Other Subjectivities

Those who feel that the story of Monochromatic Mary shows that some form of knowledge falls outside of science will be unsatisfied by the arguments in Section 11.3.2. Many of these philosophers think that the subjective and intrinsic "feel" of a phenomenal quality is left out of the scientific story. The fact that Mary can leave her room and learn to see red makes things seem too easy. What if it were impossible for Mary or any other scientist to learn to identify red by just looking? One might argue that in such a case, there is an object that remains outside of scientific inquiry.

Thomas Nagel (1937–) used the example of bat echolocation to argue that phenomenal qualities are outside the scope of science. Over half of the known species of bats use echolocation in the night as they navigate their environment and look for food (commonly insects). These bats emit a sound at a higher pitch than that detectable by human ears. Their own ears are very sensitive. They are responsive to small differences in the frequency of returning echoes, and to the direction from which the echoes return. This gives them the ability to detect the size, location, and direction of travel of objects in their environment, even very small ones. Obviously, we humans can do little of this sort. In the essay "What is it Like to be a Bat?"[5] Nagel argued that while there is much that scientists might learn about bat

[5]Thomas Nagel. What is it like to be a bat?. *The Philosophical Review*, 83(4), 435-450 (1974).

neurophysiology, "what it is like" to sense and object's size, position, and direction of travel is nothing that any human will experience. The intrinsic phenomenal quality must remain unknown, he argued, and therefore scientific inquiry cannot encompass phenomenal qualities.

What are the similarities and differences between Monchromatic Mary and (the more realistic example) of bat echolocation? To address this, let us reflect on another imaginary case; call it the case of Tasteless Tom. Imagine that Tom is a scientist who never tasted anything, and who lacks taste receptors through an accident of birth. Tom will never taste a thing. Perhaps Tom feels the pressure of his food in his mouth as he chews, but Tom cannot experience his food as more or less sweet, sour, bitter, or salty. We may suppose that Tom becomes interested in all the reactions others apparently have as they eat various foods. Such a scientist could study how people can discriminate substances by taste. Tom could do this by studying what differences in substances produce different reactions in people, and by asking people about what difference they detect (or experience) when tasting various items. Using such methods, Tom could come to identify the several dimensions of common human taste experience. He could then study how the activation of multiple chemically sensitive receptors on the tongue give rise to multiple dimensions of subjective experience. He could study this by correlating the activation of such receptors with people's reports of their experience (as being a little or a lot sweet, a little or a lot salty, little or a lot bitter, and so on). These reports allow Tom to distinguish dimensions of taste-experience. Just as in the neuroscientific account of vision, the resulting experiences are located in a multi-dimensional quality space—so much sweet, so much bitter, and so on. Now, let us suppose that Tom has arrived at a completed science of taste perception.

The case of Tasteless Tom can be compared that of Monochromatic Mary. Both manage to produce a powerful and complete scientific account of a sense pathway, and this includes an account of the experiences had by agents using the sense pathway. Tom and Mary (before she left her room) had none of the experiences that were the subject of their respective sciences. But, there is this difference: Mary can walk out of her monochromatic room, experience the sight of red, green, and other-colored objects, and learn how to discriminate red from other objects by sight. Tom, however, cannot ever taste a thing. He cannot learn how to discriminate the taste of chocolate from the taste of lemon by eating them.

So our imagined exemplary scientist, Tasteless Tom, studying common human taste experiences and the system producing them, can know a lot about humans' qualitative experiences of taste, including (a) the ways in which these may be similar or dissimilar, and in what degrees these are similar in these respects, and (b) the biological processes giving rise to these qualitative experiences. Tom can know all this, despite not being capable of having these qualitative experiences. In parallel fashion, the scientist studying bat echolocation can know a lot about the qualitative experiences had by bats. The scientist can (a) locate these within an abstract space of experiences that are more or less similar in various ways, and (b) can understand the biological processes giving rise to these qualitative experiences and their similarities. In both cases, then, the scientific results tell us a great deal about the qualitative experiences of the relevant creatures (humans with their common taste sensibility and bats which can echo-locate). They manage this without ever being able to themselves undergo the states with the relevant intrinsic what it is like of those states.

Now, we noted that it is plausible that when Mary walks out of her room, she learns or begins to learn something new, something beyond the (a)-type and (b)-type scientific knowledge she possesses. According to the proponent of Arguments 11.4 and 11.5, Mary directly grasped an intrinsic quality, and that to know this is to know more than its location in an abstract quality space. According to this argument, science is limited insofar as such knowledge "escapes its net." The reply distinguished between knowledge *that* something is the case, and knowledge *how* to do something. The critic of Arguments 11.4 and 11.5 contended that Mary is best understood as acquiring a new ability, a knowledge-how, not the knowledge that there is a new kind of object (a phenomenal property) in the world. What about Tasteless Tom? He will never be able to acquire such know-how. Similarly, no scientist will ever be able to acquire the ability to track small objects by echolocation.[6] Does this not show that Tom's (a)-type and (b)-type knowledge is limited? Scientists will never know what it is like to be a bat.

In reply, the critic of Arguments 11.4 and 11.5 might distinguish between "knowledge how" as general category, or mode, of knowledge, and the specific abilities involved in knowing how to do this or that. We have already noted that as a mode of knowledge, know-how is a part of science. Scientific instruments, such as microscopes or telescopes, require know-how for their use. Conducting an experiment requires substantial practical expertise. Insofar as experiments and instruments are necessary for scientific knowledge, know-how has to be included within the category of "scientific knowledge." However, there are innumerable, specific practical abilities that are not commonly thought of as part of scientific knowledge. The critic of Arguments 11.4 and 11.5 has acknowledged that learning to recognize red is something that most humans do early in their cognitive and linguistic development. Science does not tell us the taste of the soup: distinguishing cream soup from tomato soup is not a bit of know-how that is commonly[7] part of science. Tom's inability to taste does not show a limitation to science, it is simply a limitation of his sensory capacities. The inability to echo-locate is a limitation of human sensory capacities. The critic of Arguments 11.4 and 11.5 contends that such a limitation of human sensory capacities is no more interesting as a limitation of science than Tasteless Tom's inability to taste.

[6] Unless, of course, we develop a technological prosthesis, a kind of super-hearing-aid, for doing so. We set aside such a possibility for the sake of this argument.

[7] It is impossible to draw a principled line between those recognitional abilities that are part of science and those that are not. Chemists, for example, used to use taste discrimination in chemical tests. (For safety reasons, this is not done any more!). This point supports the critic of Arguments 11.4 and 11.5.

11.5 Conclusion

This chapter has explored the question of whether scientific knowledge has an interesting and important limitation. We began our discussion with the caution that scientific knowledge is complex and multi-faceted, and that we should be cautious when generalizing about it. As a result, we focused on a specific sort of scientific knowledge: the study of the human sensory system, and the visual system in particular. In the course of exploring these arguments, we have discovered two modes of knowledge that are properly part of science: knowledge-that and knowledge-how. When we think of scientific theories and discoveries, we are thinking of scientific knowledge-that. Most of the questions in this book have been about the kinds of evidence and reasoning that may be used in support of scientific claims. All of these issues concern the scientific knowledge *that* a theory is true or a inference reliable. however, scientific inquiry also requires the ability to use instruments and the ability to recognize things by looking, hearing, smelling, touching, or tasting. Hence knowledge *how* is also a mode of knowledge used in scientific inquiry.

The fanciful example of Monochromatic Mary, and the further example of Tasteless Tom, helped us raise the question of whether there is an interesting and important limitation to science. The dispute ended up centering on two arguments, Arguments 11.4 and 11.5, that concluded that there is a kind of object and a mode of knowledge that is outside of the reach of science. We have not tried resolve these arguments. There is too much to say for a text such as this. We do hope that our readers are intrigued.

Appendix A

Logical Form and Validity Testing

In Section 3.2, we introduced a method of diagramming arguments to demonstrate their validity or invalidity. While this method does not apply to all deductively valid arguments, it does apply to an interesting class of them: syllogisms. Syllogisms were identified by the ancient Greek philosopher Aristotle (384–322 BCE), and have been a crucial part of the Western study of logic.

The important lesson of Section 3.2 was that validity is a matter of *form*. That is, the validity of an argument is a product of how the terms are related, not what the terms mean. The example we used earlier was had the form of Argument A.1.

All A are B
All B are C

All A are C

Argument A.1

Notice that all three sentences of this argument have the same form: "All A are B." The example of the invalid argument, 3.4, used another sentence form: "Some A are B." These are just two of the sentence forms that appear in syllogisms. Syllogisms are arguments created from four possible sentence forms in Figure A.2.

All A are B
No A are B
Some A are B
Some A are not B

Figure A.2

The following examples are all *valid* syllogisms. Each argument form is paired with an example where the variables A, B, and C have been replaced with meaningful terms. These are just some of the valid syllogisms.

Form	Example
All A are B No B are C ――――― No A are C Argument A.3	All monks are humans No humans are immortal ――――― No monks are immortal Argument A.4
Some A are B All A are C ――――― Some C are B Argument A.5	Some monks are compassionate All monks are Buddhist ――――― Some Buddhists are compassionate Argument A.6
All A are B Some C are A ――――― Some C are B Argument A.7	All monks are compassionate Some humans are monks ――――― Some humans are compassionate Argument A.8
All A are B No C are B ――――― No C are A Argument A.9	All monks are compassionate No mean people are compassionate ――――― No mean people are monks Argument A.10

To demonstrate that a syllogism is valid, we must represent each premise and conclusion on a diagram. As we did in Section 3.2, each term is represented by a circle. Since each of the premises and the conclusion of a syllogism relate two terms, each sentence is represented by a pair of circles. The four sentence forms are represented in Figures A.11, A.12, A.13, and A.14.

To say that "All monks are compassionate" tells us that every individual with the property of being a monk will also have the attribute of compassion. Our convention is to darken the area of the circle that is empty. Since the sentence says that there are no monks that are not also compassionate, the left part of the monk circle must be darkened. In Figure A.11, "monk" and "compassion" have been replaced by the variables A and B. Any sentence with the form "All A are B" is represented with the diagram of Figure A.11.

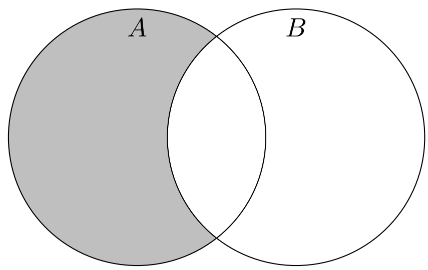


Figure A.11: All A are B

The sentence form "No A are B" tells us that there is no overlap between the two terms. If no monks are drink alcohol, then there are two groups of people—the monks and the alcohol drinkers—and no individual is in both groups. This means that the area in the diagram where the two circles overlap must be darkened, as in Figure A.12.

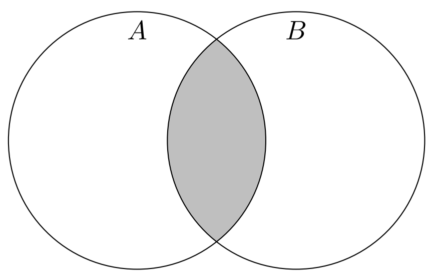


Figure A.12: No A are B

When a sentence has the form "Some A are B," as in "Some monks are tall," no part of either circle is darkened. The left part of A circle represents things that have the A property, but not the B property—monks that are not tall. The sentence does not tell us whether or not there are monks that are not tall, so we cannot assume this part of the circle is empty. Similarly, the right part of the B circle represents things that are tall, but not monks. Again, the sentence does not tell us whether this group is occupied or empty. What we do know from the sentence is that there is at least one individual with both properties. To represent this, we put the symbol

X in the area of overlap, as illustrated in Figure A.13.

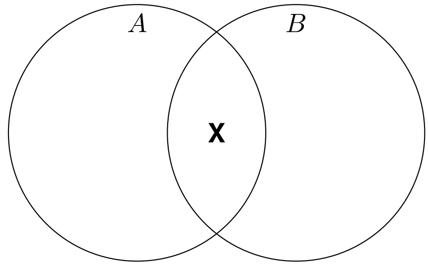


Figure A.13: Some A are B

Finally, the sentence form "Some A are not B" tells us that there is at least one thing in the group of As that does not have the B property. Just as in the previous form we use the symbol **X** to indicate that there is at least one individual in that part of the circle.

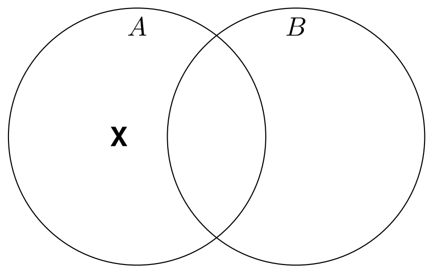


Figure A.14: Some A are not B

When we test an argument for validity, we must represent all three sentences—the two premises and the conclusion—on one diagram. Recall the definition of validity on page 20: an argument is valid if there is no possible situation where the premises are true and the conclusion false. If the argument is not valid, and there is such a situation, the diagram will show it to us. If the argument is valid, and such a situation is impossible, then the diagram will show that too.

Let us begin with the Argument A.4 All monks are humans; no humans are immortal; therefore no monks are immortal. Rather than show all three sentences on the diagram at once, we will build it sentence by sentence.

As a rule of thumb, it is best to start with premises, then add the conclusion. Also, it is best to start with the sentences of the form "All A are B" or "No A

are B," since these restrict the placement of the **X** when we diagram sentences beginning with "Some. . . ".

The first premise of the argument is that all monks are humans. On the diagram in Figure A.15, this is indicated by darkening that part of the monks circle that is outside of the circle of humans.

The second premise of the argument is that no humans are immortal. This means that we must darken the area of overlap between the human circle and the immortal circle. This is demonstrated by the diagram of Figure A.16.

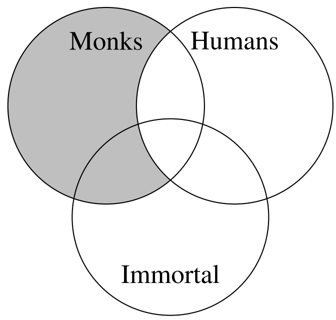


Figure A.15

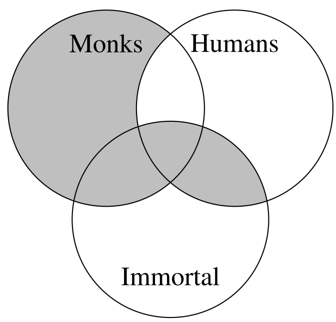


Figure A.16

The conclusion of the argument is that no monks are immortal. To represent this sentence on the diagram, we would darken the area of overlap between the monk circle and the immortal circle. We can see from the diagram in Figure A.16 that this area is already darkened. By representing the premises, we have already represented the conclusion. This means that the truth of the premises guarantees the truth of the conclusion. The argument is clearly valid, since any situation that made the premises true, would also make the conclusion true.

As a second example, let us consider Argument A.5. The example of Argument A.6 was: Some monks are compassionate; all monks are Buddhist; therefore some Buddhists are compassionate. Once again, we will begin with the premises, and use the premises with the "All A are B" or "No A are B" forms first. Figure A.17 shows "All monks are Buddhist" on the diagram for the argument.

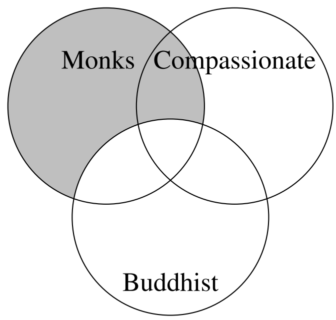


Figure A.17

Now we want to add "Some monks are compassionate" to the diagram. To do so, we will put a **X** in the area where the monk circle and the compassion circle overlap. Notice that part of this overlapping area has been shaded. The premise that all monks are Buddhist has darkened that part of the overlap between the monk circle and the compassion circle. That is, the first premise precludes the existence of non-Buddhist monks who are compassionate. The only remaining place to put the **X**, then, is in the part of the diagram shared by all three terms, as in Figure A.18.

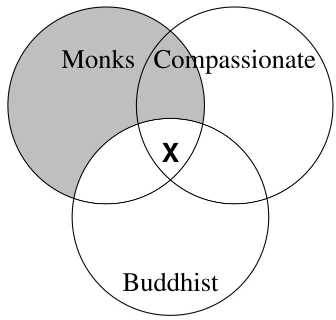


Figure A.18

The conclusion of the argument is that some Buddhists are compassionate. To add this sentence to the diagram, we would put an **X** in the overlap between the Buddhist circle and the compassionate circle. However, in Figure A.18, there is already an **X** in that area. By making the premises true, we have already made the conclusion true. So, there is no possible situation where the premises are true and the conclusion false; hence the argument is valid.

An remember, *any* terms that had this relationship would constitute a valid argument. So, any argument that exhibits the form of Argument A.5 is also valid.

There are two remaining examples from our earlier list of four, Arguments A.7 and A.9. Rather than taking you through these step-by-step, we are simply going to show you the diagrams that demonstrate their validity. See if you can identify each premise on the diagram, and convince yourself that there is no situation where the premises are true and the conclusion false.

All A are B
Some C are A

Some C are B

Argument A.19

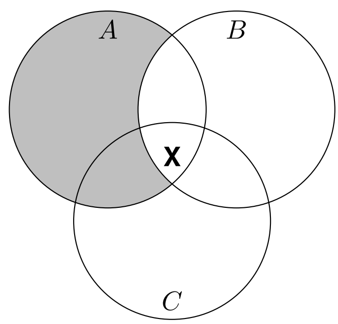


Figure A.20

All A are B
No C are B

No C are A

Argument A.21

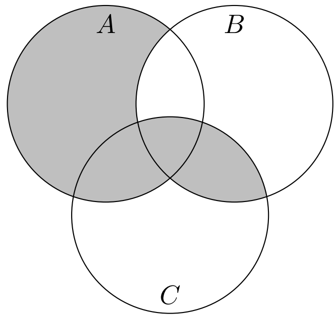


Figure A.22

So far, all of the examples we have diagrammed so far have been valid. Let us consider how the diagrams show that an argument is invalid. An argument is invalid when there is a possible situation where the premises are true and the conclusion false. This means that when we diagram the premises, the diagram will not already make the conclusion true. Rather, there will be some possibility left open where the conclusion could be false.

No monks are rabbits
No rabbits are human

No monks are human

Argument A.23

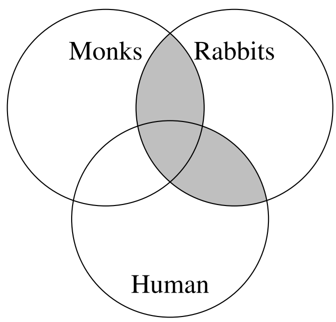


Figure A.24

Consider Argument A.23. Figure A.24 has represented both premises. Notice how the diagram does not yet represent the conclusion. For the conclusion to be represented, the overlap between the monk circle and the human circle would have to be darkened. The premises do not guarantee that this area is empty; there could be monks who are human. In other words, there is a possible situation where the premises are true and the conclusion false, and the argument is invalid.

All monks are tea drinkers
Some tea drinkers are tall

Some monks are not tall

Argument A.25

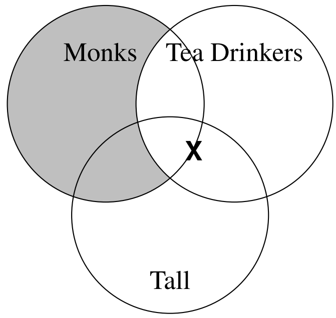


Figure A.26

Argument A.25 introduces a new situation. The first premise is represented in Figure A.26 by the darkening of the left part of the monk circle, as we have done before. The second premise introduces a problem. There are two possible areas where the **X** could go. This is because the second premise does not specify whether the tall tea drinkers are monks. It could be that the tall tea drinkers are not monks, or perhaps they are. We represent this uncertainty by putting the **X** symbol on the boundary. The second premise requires that whether or not the tea drinkers are monks, there must be some tall ones.

Now, let us turn to the conclusion of Argument A.25. This requires us to put an **X** outside of the tall circle, but within the monk circle. This area on the diagram is not darkened. But again, there are two areas where the monks could be. It is possible, given what the premises say, that all the monks are tall. This is a situation where the conclusion would be false, even thought the premises are true. Hence, the argument is not valid.

In this Appendix, we have seen a small sampling of the way that Western logic tests for validity. The field of logic is extensive, and there are many other deductively valid forms than can be represented with the simple diagramming method demonstrated here. But to discuss them would turn take us too far away from our main goals.

While only a glimpse, these methods demonstrate two of the ideas that are central to the Western approach to logic. First, validity is a matter of *form*, not the specific content of the sentences. This is illustrated by the fact that the diagrams which demonstrate validity do not change when the terms are changed. Swapping the word "chocolate thing" for "monk" in any of these arguments would yield the very same diagrams. This is why we can dispense with the terms and use variables like A, B, and C in their stead. Once we do so, we can clearly see the valid logical forms.

Second, validity does not depend on the actual truth or falsity of the premises.

This is the distinction between validity and soundness. Validity is a relationship between the premises and conclusion: if the premises *were* true, the conclusion *must* be. As we saw in the discussion of falsification (Section 3.4), this characteristic of validity is central to scientific testing. When we test a theory we do not know whether it is true. We test by evaluating the logical consequences of the theory, and this requires that we are able to make valid arguments based on premises drawn from the theory.

Index

The Dharma of Science: Philosophy of Science for Buddhist Scholars provides a broad overview of the metaphysics and epistemology of contemporary science. Written as an accompaniment to science education in Buddhist monasteries and nunneries, it explores the grounds of scientific knowledge and some of the philosophical implications of scientific practice.

About the Authors

Mark Risjord is a professor of philosophy at Emory University in Atlanta, Georgia, USA. He writes about the philosophy of science, with a special interest in the social sciences and in medicine. His books include *Philosophy of Social Science, A Contemporary Introduction* (Routledge, 2014), *Nursing Knowledge: Science, Practice, and Philosophy* (Wiley-Blackwell, 2010), and *Woodcutters and Witchcraft: Rationality and Interpretive Change in the Social Sciences* (SUNY Press, 2000). He has coedited (with Stephen Turner) *Philosophy of Anthropology and Sociology* (Elsevier, 2007) and edited *Normativity and Naturalism in the Philosophy of the Social Sciences* (Routledge, 2016).

David Henderson is the Robert R. Chambers Distinguished Professor of Philosophy at the University of Nebraska, Lincoln, USA. He writes on the philosophy of science, with a special interest in the social sciences and in epistemology. His books include *Interpretation and Explanation in the Human Sciences* (SUNY Press, 1993), and *The Epistemological Spectrum: At the Interface of Cognitive Science and Conceptual Analysis* (co-authored with Terry Horgan) (Oxford, 2011). He has coedited (with Miranda Fricker, Peter Graham, Nikolaj Pedersen) *The Routledge Handbook in Social Epistemology* (Routledge, 2019) and (with John Greco) *Epistemic Evaluation: Point and Purpose in Epistemology* (Oxford, 2015).

Mark Risjord David Henderson
Tashi delek!