Understanding Scientific Reasoning

Third Edition

Ronald N. Giere
Minnesota Center for Philosophy of Science
University of Minnesota—Twin Cities

Holt, Rinehart and Winston, Inc.

Fort Worth Chicago San Francisco Philadelphia Montreal Toronto London Sydney Tokyo

Publisher Ted Buchholz
Acquisitions Editor Jo-Anne Weaver
Project Editor Steve Welch
Production Manager Kathy Ferguson
Art & Design Supervisor Vicki McAlindon Horton
Text Designer Bill Maize, Duo Design Group
Cover Designer Vicki McAlindon Horton/Bill Maize, Duo Design Group

Library of Congress Cataloging-in-Publication Data

Giere, Ronald N.
Understanding scientific reasoning / Ronald N. Giere. — 3rd ed.
p. cm.
Includes bibliographical references and index.
1. Science—Philosophy. I. Title.
Q175.G49 1991
501—dc20 90-49877
CIP

ISBN: 0-03-026419-7

Address for Editorial Correspondence: Holt, Rinehart and Winston, Inc., 301 Commerce Street, Suite 3700, Fort Worth, TX 76102.

Address for Orders: Holt, Rinehart and Winston, Inc., 6277 Sea Harbor Drive, Orlando, FL 32887. 1-800-782-4479, or 1-800-433-0001 (in Florida).

Printed in the United States of America

1 2 3 4 090 9 8 7 6 5 4 3 2 1

Holt, Rinehart and Winston, Inc.
The Dryden Press
Saunders College Publishing

For permission to use copyrighted materials the author is grateful to the following:

Chapter 2: Section 2.10, "Gene Analysis Upsets Turtle Theory," *New York Times*, March 14, 1989. © 1989 by The New York Times Company. Reprinted by permission; "New View of the Mind Gives Unconscious An Expanded Role," *New York Times*, February 7, 1984. © 1984 by The New York Times Company. Reprinted by permission; "Was That a Greenhouse Effect? It Depends on Your Theory," *New York Times*, September 4, 1988. © 1988 by The New York Times Company. Reprinted by permission; Exercise 2.1, "Einstein's Impossible Ring: Found," by M. Mitchell Waldrop, *Science*, June 24, 1988, Vol. 240, p. 1733. © AAAS. Reprinted by permission; Exercise 2.2, "Why is the World Full of Large Females," by Roger Lewin, *Science*, May 13, 1988, Vol. 240, p. 884. © AAAS. Reprinted by permission; Exercise 2.4, "Green Antarctica," by Bill Lawren, *Omni Magazine*, August 1987, p. 30; Exercise 2.5, "Hot Extinction Theories," *Sky and Telescope*, July 1988. Reprinted by permission of Sky and Telescope magazine. Exercise 2.6, "A Heresy in Evolutionary Biology," by Roger Lewin,

(continued on page 314)

PART ONE

Theoretical Hypotheses

CHAPTER 2

Understanding and Evaluating Theoretical Hypotheses

In this chapter, we will develop a framework for understanding and evaluating a wide range of scientific cases. The chapter begins with a case study, the discovery in 1953 of the structure of DNA. The case study will provide everyone with enough background on this particular episode so that it can be used as an example for most of the chapter. After the framework is in place, it will be applied to a variety of other examples.

2.1 THE DOUBLE HELIX: A CASE STUDY

In the fall of 1951, a 23-year-old American named Jim Watson arrived at The Cavendish Laboratory of Cambridge University in Cambridge, England. He had come in pursuit of his personal scientific quest to discover the physical structure of DNA—deoxyribonucleic acid—a discovery he was sure would bring him fame. Why DNA?

The idea that human inheritance is transmitted from parents to offspring by identifiable bits of matter in germ cells (sperm and eggs) has a long history. Since around 1900, there had been slow, but steady, progress in determining the chemical structure of these particles, now called genes. By 1950, it was well known that germ cells contained both DNA and proteins, which are large chains made up of units called amino acids. In spite of Oswald Avery's 1944 experiments, which strongly suggested that genes are made of DNA, in 1950 most biologists and chemists still thought that genes are made of proteins rather than DNA. One of the relatively few people who took Avery's work seriously was Salvador Luria, an Italian-born geneticist teaching at Indiana University. Watson, after completing his B.A. in zoology at the University of Chicago in 1947, went on to study for his Ph.D. with Luria. When Watson finished his Ph.D. in 1950, he and Luria decided that the best route to further progress in genetics would be through detailed knowledge of the structure of DNA. Watson got a grant for further study in biochemistry with an expert in Copenhagen.

The state of knowledge in 1951 concerning the makeup of DNA is summarized in Figures 2.1 and 2.2. A DNA molecule was thought to consist of one or more chains of nucleotides. Each nucleotide consists of a sugar molecule (deoxyribose), a phosphate molecule, and a base. There are four different possible bases, two each of two kinds: purines (adenine and guanine) and pyrimidines (cytosine and thymine). Such chains of nucleotides,

Figure 2.1

A short section of DNA as represented by organic chemists in 1951. Note that this representation omits any reference to the three-dimensional arrangement of the atoms.

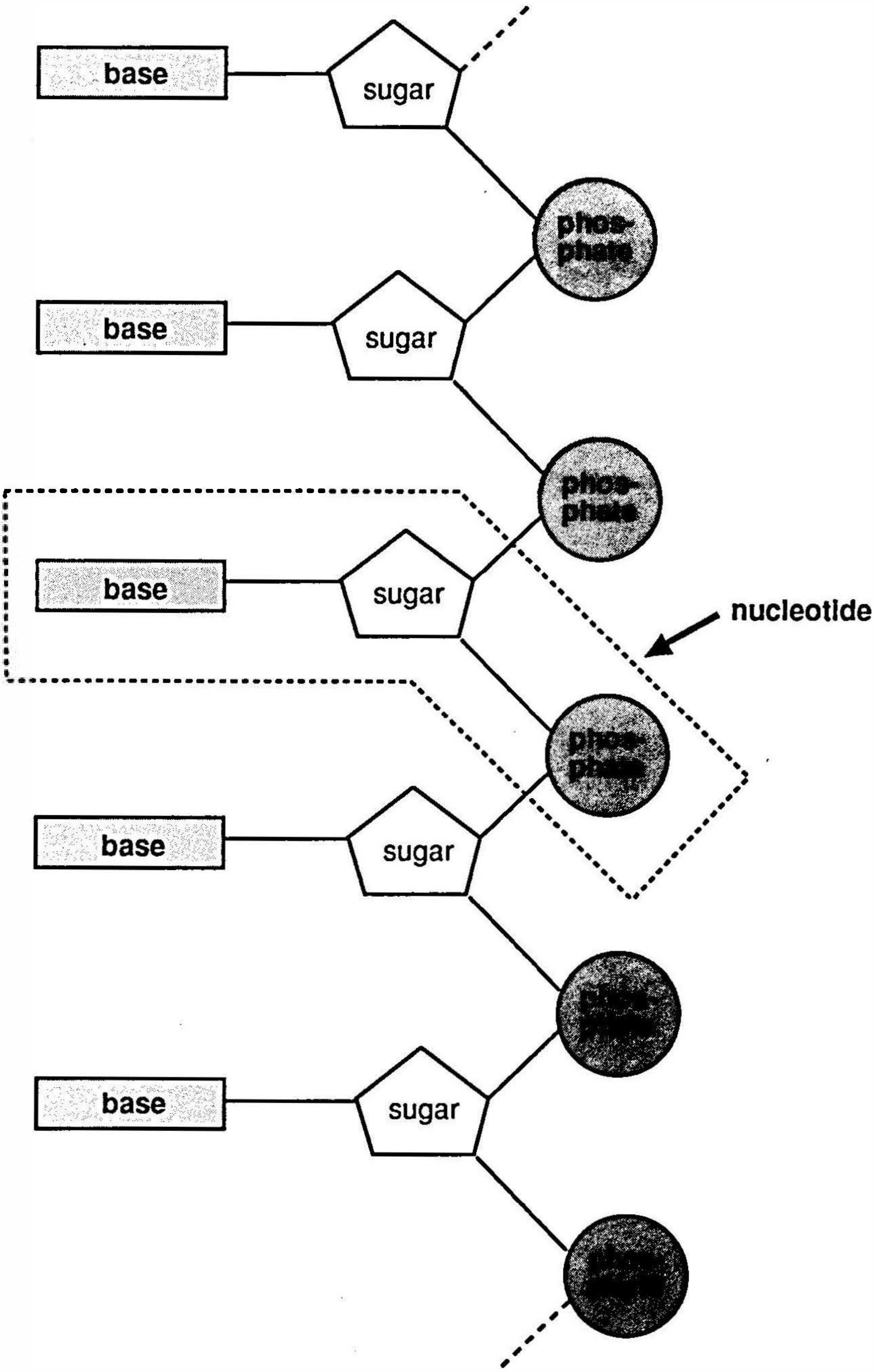
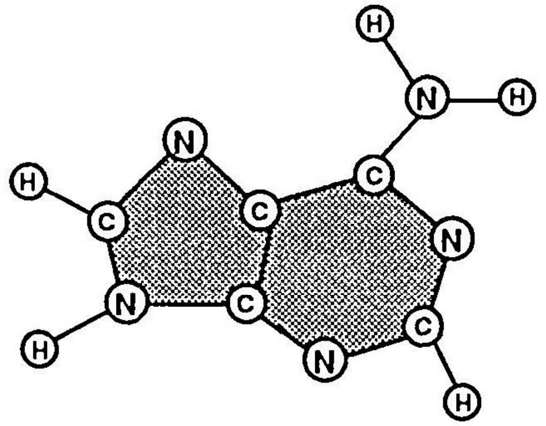


Figure 2.2

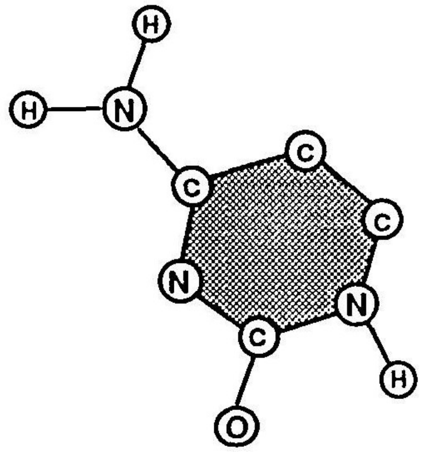
Representations of the four bases known in 1951 to be present chemically in DNA.

PURINES

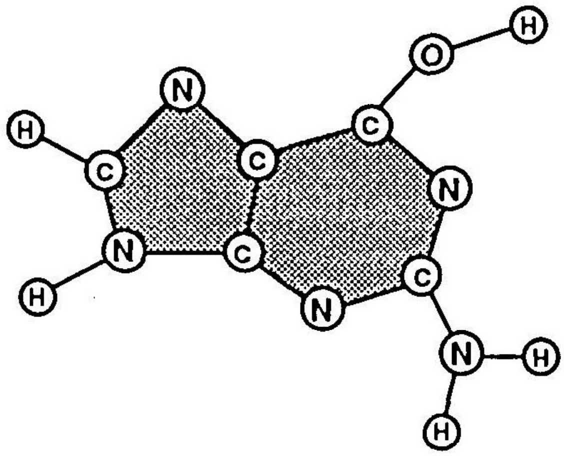
PYRIMIDINES



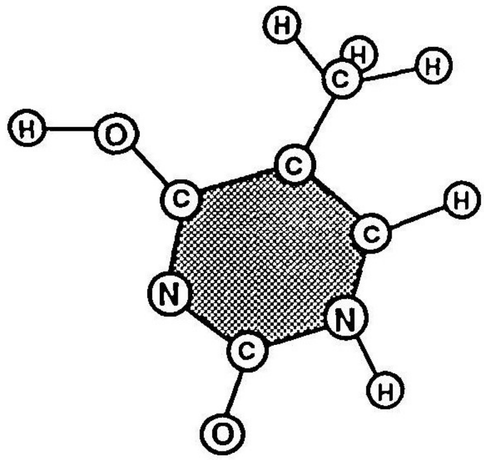
adenine



cytosine



guanine



thymine

called "polynucleotides," were thought of as consisting of a "backbone" composed of sugar and phosphate supporting a sequence of bases. What Luria and Watson wanted to know was how all these pieces physically fit together. They believed that such structural knowledge would make clear how genes function in the process of inheritance.

Watson found the work in Copenhagen a waste of time. The following spring he went off to Naples, Italy, for two months where he spent his time reading articles from the early days of genetics and daydreaming about discovering the secret of the gene. While in Naples he attended a small, scientific meeting concerning the structure of large molecules in living organisms. One talk excited him. Maurice Wilkins, an Englishman working at Kings College at the University of London, showed a photograph taken by focusing x-rays on a small amount of chemical DNA. The picture indicated that DNA had a fairly regular crystal-like structure. That meant that there was a reasonable chance of actually figuring out some details of the structure. Watson was delighted, and began to think that x-ray methods were a better route to the structure of DNA than the biochemistry he had vainly been trying to master. He tried to make friends with Wilkins but without success.

While visiting a friend in Geneva on the way back to Copenhagen, Watson learned that the world's greatest living physical chemist, Linus Pauling, had just discovered the structure of a significant protein molecule, α-kerotin. The structure was a helix, and Pauling had discovered it by building a physical model of the molecule, using information obtained from x-ray photographs. Now Watson was sure this was the way to go. But where could he go to pursue his quest? He had been put off by Wilkins and he was sure Pauling would pay little attention to someone with so poor a knowledge of physical chemistry. The only place he could think of was The Cavendish, where he knew that some people were using x-ray techniques to study large molecules. He wrote to Luria for help. By good fortune, shortly after receiving Watson's letter, Luria met one of The Cavendish scientists at a meeting in the United States and arranged for Watson to begin work there the following fall. In science, as in other endeavors, it helps to have influential friends.

Three-Chain Model

It was at The Cavendish that Watson first met Francis Crick. Although a dozen years older than Watson, Crick was still working on his Ph.D. He kept getting distracted by other interesting theoretical problems. Crick, however, shared Watson's belief in the importance of DNA and his hunch that the best strategy for discovering its structure was to build models as Pauling had done in discovering the α-helix. They speculated that DNA also possessed a helical structure.

One possible embarrassment was that Wilkins was already working on the problem, and, unlike Americans, English scientists tended to respect such territories. However, Wilkins's own work was going very slowly because he did not get along with the person in his laboratory engaged in x-ray studies of DNA, Rosalind Franklin. Partly because of his conflicts with Franklin, Wilkins voiced no objections to Watson and Crick's fiddling with models of DNA.

Crick soon provided a major contribution to the project by developing a theoretical account of how x-rays are diffracted by helically-shaped molecules. If one is going to use x-ray pictures in building models of helically-shaped molecules, one needs to know what an x-ray picture of such molecules should look like.

They still needed more information about existing x-ray photographs of DNA. Luckily, Franklin was scheduled to give a talk on her recent work in London the middle of November. Watson was dispatched to the talk to learn what he could. The day following the talk found Watson and Crick on the train to Oxford for a weekend visit with a friend. Crick was excited. His theory of x-ray scattering, together with the data Watson related from Franklin's talk, indicated that there could be only a small number of possible helical structures for DNA molecules. It should consist of at least two, but not more than four, polynucleotide chains. They decided to try a model with three chains.

The next big question was the position of the sugar phosphate backbone, as it was called, relative to the bases. There were only two major alternatives. Either put the intertwined backbone in the center and let the bases hang out on the outside, or put the backbone on the outside and try to fit the bases into the inside. Fitting the bases into the inside seemed too complicated, so they decided to try building a model with the bases on the outside.

Upon returning to Cambridge, they set about building a model using pieces of wire and specially fabricated metal plates to represent the various components of the polynucleotide chains. In this task, their major reference work was Pauling's book *The Nature of the Chemical Bond.* This book provided the best available information about the distances and angles between the various groupings of atoms held together by chemical bonds. A good model had to reflect these basic features of atoms.

In less than a month they had completed what they regarded as a quite satisfactory model. They invited Wilkins and Franklin up from London to inspect their handiwork. It took Franklin only a few minutes to discover a major flaw in the model. Natural DNA is surrounded by water, which is loosely bound to the molecule. Watson and Crick's three-chain model left far too few places for water molecules to hook on to the DNA molecule. In fact, real DNA accommodates ten times the amount of water permitted by the model. Indeed, Franklin had given the correct information in her talk the previous month. Watson had misremembered what she had said!

In the aftermath of their humiliation by the group from London, the director of The Cavendish, Sir Lawrence Bragg, forbade Watson and Crick to engage in any more DNA model building. Watson went off to spend Christmas with the family of a friend at their manor house in Scotland. His dreams of fame and glory seemed far from being realized.

Two-Chain Model

Returning from vacation, Watson took up learning how to take x-ray pictures of the tobacco mosaic virus (TMV). He was not wasting his time because TMV should have a helical structure. And, indeed, several months and a new x-ray tube later, he obtained good pictures clearly indicating a helical structure. But neither Watson nor Crick stopped thinking about DNA, even though they were officially forbidden to work on it.

Meanwhile, two scientists at Cold Spring Harbor reported an experiment that strongly supported the idea that the primary genetic material is DNA, not proteins. In eight years, the scientific climate had changed. Unlike Avery's work, these new experiments were being taken very seriously by many other geneticists. For Watson, news of these new results was both good and bad.

It confirmed that he was right to focus on DNA. But now many other people would start working on DNA. His advantage was slipping away.

Another new result aroused Watson and Crick's interest. An Austrian-born biochemist at Columbia University, Erwin Chargaff, had carefully measured the base contents of DNA from several different biological species. The relative amounts of the pyrimidines, adenine and cytosine, varied from species to species. But, remarkably, the amounts of adenine and thymine were the same in all samples, as were the amounts of cytosine and guanine. Like Chargaff, Watson was sure these results were highly significant, but no one seemed to have any idea just what the significance might be. Crick, too, became increasingly preoccupied with the Chargaff results.

September of 1952 found Watson turning his attention to the idea that bacteria come in male and female pairs. If true, this meant that the genetics of bacteria are much more like that of higher organisms than had earlier been thought. Crick was once more back at work on his still unfinished Ph.D. dissertation. One new aspect of their lives was that Linus Pauling's son, Peter, had joined their group at The Cavendish. Through Peter they were able to keep abreast of the news from Pasadena.

The first ominous word was that Pauling was working on α-coils. A little later came word that he was working on DNA, but no details. Then, in the middle of January, came a draft of a paper in which Pauling outlined a model of DNA. To their great relief, Watson and Crick found that Pauling had come up with a model superficially resembling their own ill-fated, three-chain model. In addition, it had several other features that they felt sure had to be mistaken. They figured they had at most six weeks before Pauling discovered his mistake and turned the full power of his genius to rectifying the blunder. Nevertheless, they were determined to turn all their energies to the problem once again. The serious prospect that Pauling, an American, might beat his British group to the solution was enough to convince Bragg to let them try again.

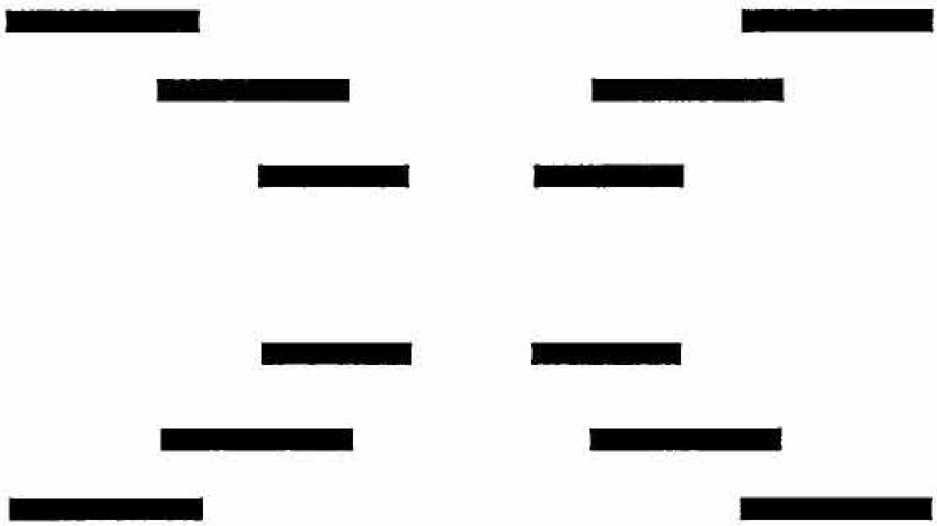
Watson journeyed down to London to show Pauling's paper to Wilkins and to enlist his support for their new effort. Not finding Wilkins immediately, he went around to Franklin's lab. Citing Pauling's paper, he tried to convince her of the urgency of the situation and to enlist her, and her carefully acquired x-ray data, in the effort. She chased Watson out of her lab, reportedly insisting that she would have no part of their "little boys' games."

Wilkins consoled Watson by showing him a picture Franklin had taken of what she called the "B form" of DNA, which contained much more water than the then standard "A form." The pattern, as shown in Figure 2.3, was incredibly simpler than any Watson had seen before. The strong black crosses could only come from a helical structure. That followed immediately from Crick's theory about how x-rays are diffracted by helixes. Watson returned to Cambridge more excited than ever.

They still did not have enough information to answer the big questions that had faced them the year before. How many chains are there? Are the bases outside or inside? If the bases are inside, how are they arranged? This time Watson decided to try two-chain models, appealing to the general idea that important biological entities come in pairs.

Figure 2.3

A schematic rendering of the pattern revealed by Rosalind Franklin's 1952 x-ray photograph of the B form of DNA.



The decision whether to put the bases outside or inside was more difficult. They had always worried that there would be too many possible ways of arranging the bases on the inside, thus making it difficult to decide which is correct. But seeing that they were getting nowhere with base-outside models, they decided to have a go at models with the bases inside.

The idea was that a base attached to one sugar phosphate backbone should bond with a base on the opposing sugar phosphate backbone, thus forming a kind of miniature spiral staircase. The immediate problem was that while the distance between the two strands, the diameter of the spiral, should be constant, the bases are all of different sizes. Hooking up any two bases across the inside could either deform the bases or make bulges in the spiral backbone. Nevertheless, in spite of deformations and bulges, Watson proceeded to build a model with bases bonded like-with-like, that is, adenine with adenine, and so on.

He was shortly set straight by an American crystallographer, Jerry Donohue, who had worked with Pauling. Watson had taken information about the hydrogen bonds on guanine and thymine out of a standard textbook. Donohue informed Watson that the standard texts were wrong. The bases could not possibly bond the way Watson's like-with-like model required. Crick voiced still other objections. Watson reluctantly gave up on the like-with-like scheme.

The machine shop at The Cavendish was late in producing the little tin plates they needed to represent the bases. Being impatient to get on, Watson cut his own set of bases out of stiff cardboard. Playing around with his cardboard models Watson discovered that the combination adenine-thymine has a very similar shape to the combination guanine-cytosine. With these "steps" one could build a spiral staircase with a uniform diameter. Donohue confirmed that the required hydrogen bonding would work. Moreover, this scheme provided an immediate explanation for the Chargaff results because each pair consists of one purine and one pyrimidine, and only these particular combinations would bond in the required manner. Crick, who later claimed independently to have come to the same conclusion without benefit of Watson's cardboard models, was quick to proclaim that they had found "the secret of life."

But there was still work to do. When the shop delivered the metal cutouts, they took a plumb line and measuring stick to the model, carefully aligning all the pieces to make sure that they did fit together in a configuration consistent with knowledge of the relevant chemical bonds. It seemed to be all in order. Even Sir Lawrence Bragg was pleased. More to the point, Wilkins, and even Franklin, agreed that the proposed structure was confirmed by a detailed examination of their own x-ray data.

"We wish to suggest a structure for the salt of deoxyribose nucleic acid (D. N. A.)." So began the 900-word paper by Watson and Crick published in *Nature,* March 25, 1953. Figure 2.4 shows the schematic rendering of the structure of DNA as it appeared in this first paper. By prior arrangement, Watson and Crick's paper was followed by two papers by Wilkins and Franklin, respectively. These papers set the direction for work in molecular biology that continues as we traverse the final decade of the twentieth century.

2.2 UNDERSTANDING EPISODES IN SCIENCE

Now that we know how the double helix was discovered, let us go back over the case and begin developing a few general analytical tools for understanding this and similar episodes in science. We will first survey some general features of such episodes, and then go on to examine a few of these features in more detail.

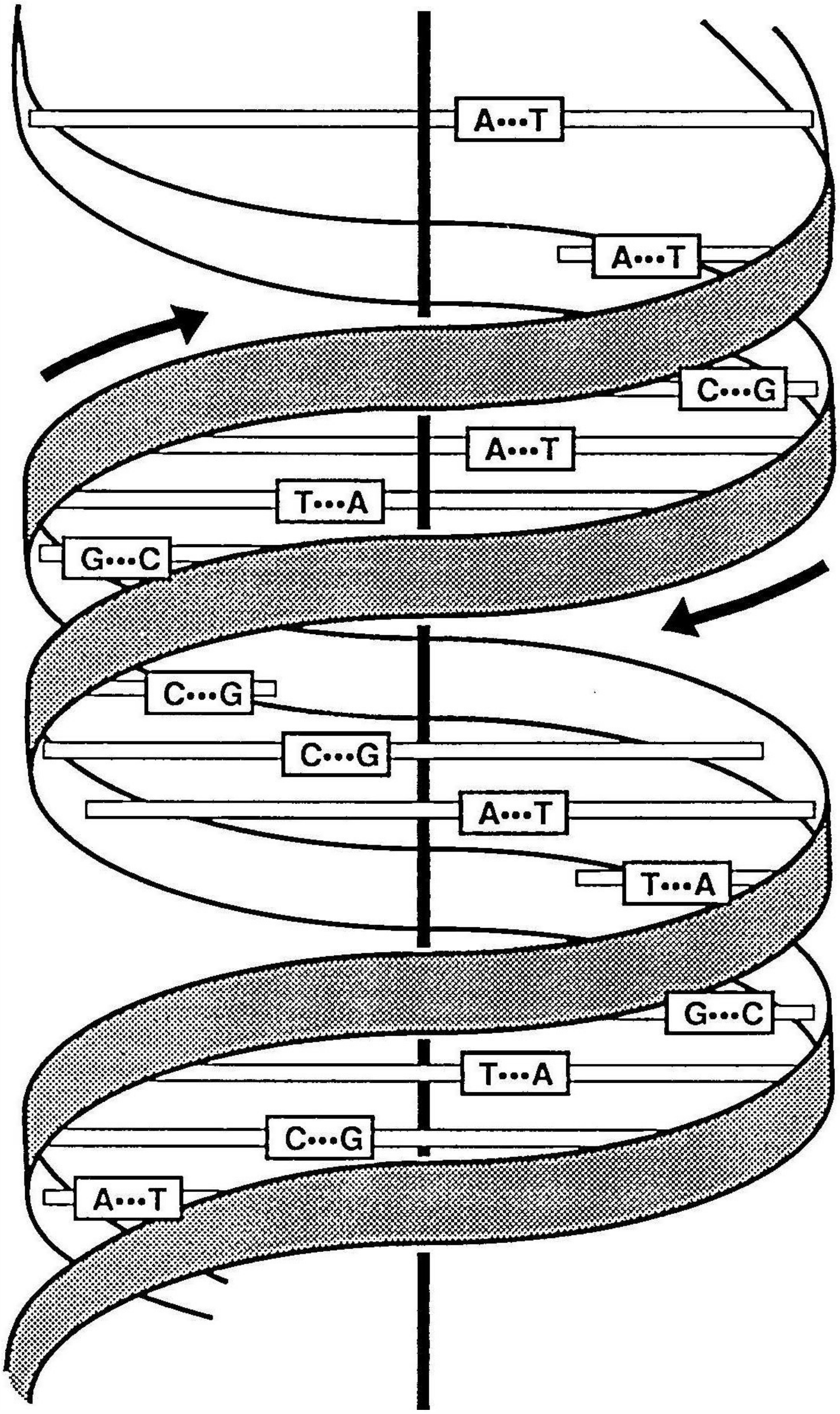
Human Context of Science

The Double Helix, Watson's own personal account of the discovery, was first published in 1968, fifteen years after the event. Many people objected that the book was too personal. In particular, it exposed the all-too-human foibles and weaknesses of the participants. Prominent among these were the conflicts between Rosalind Franklin and Maurice Wilkins. There were also the sometimes unflattering characterizations of Francis Crick, as talking too much, Sir Lawrence Bragg, as being too stiff, and Linus Pauling, as being a showman. And of course there is also the picture of Watson himself as obsessively pursuing a Nobel Prize in active competition with Wilkins, Franklin, and Pauling.

Why did people object to the personal nature of *The Double Helix?* Some, like Crick and Wilkins, objected that Watson had distorted the story. Others, however, objected mainly because they wished to preserve an image of science as a "rational," "objective" activity. Besides being flattering, this image often serves scientists well in dealings with people outside science. Watson showed that the official image of science is simply not correct.

Figure 2.4

A schematic representation of the double-helical structure of DNA as shown in Watson and Crick's 1953 paper in *Nature.*



Of course, no one should ever have been taken in to start with. Science is a human activity, and scientists are ordinary human beings. Thus, among scientists, as among members of any other profession, one should expect to find a full range of human strengths and weaknesses. So one general lesson from this episode is that the successes of science are not due to scientists being unusually blessed with virtues like objectivity, modesty, or honesty. We must look for another explanation why science works as well as it does.

A more practical lesson is that behind every episode you read, or hear about, there are many typically human stories. People's reputations and careers are on the line. The typical report of new scientific findings one sees in newspapers or magazines, however, reveals only snatches of these stories, if anything at all. One can only imagine what the real flesh-and-blood scientists were doing. Nevertheless, in trying to understand an episode, it always helps to remember that real people were there behind the scenes all along.

Exploring How the World Works

Let us grant, then, that science is a human institution run by people with a wide range of interests and motivations. Still, there must be some things that distinguish science from other institutions, such as commerce, the military, the arts, politics, or religion. The most general claim one can make is this: Insofar as they participate in science as an institution, scientists are engaged in exploring how the world works. Whatever their differences in personal motivation and style, Pauling, Watson, Crick, Wilkins, and Franklin were all engaged in trying to figure out the structure of DNA. This was part of the larger project of discovering how genes replicate, and, thus, how parents pass on characteristics to their offspring.

One can also say something in general about how scientists explore the workings of the world. They engage in careful and deliberate interactions with the world. They do experiments and make observations, some of which are designed to help them decide which of several possible ways the world might work is most like the way it really does work. This activity distinguishes a scientific tradition, for example, from a religious tradition that seeks understanding of the world through the interpretation of sacred texts, or a literary tradition based on particular literary forms, such as the novel.

It is often said that science is distinguished by the use of something called "the scientific method." This is doubtful, at best, because the methods scientists actually employ are as varied as the subjects scientists study, from astronomy to zoology. What one can say in general about the methods of science, about experimentation, for example, cannot be specific enough to be of direct use to any scientist engaged in actual research. Some things, however, can be of use to an onlooker seeking to understand and, to some extent, even to evaluate the goings-on in particular scientific inquiries.

Finding a Problem

No scientist is simply engaged in the general pursuit of discovering how the world works. They all focus on some particular aspect of the world.

They have special problems they seek to solve. Watson, for example, set himself the problem of determining the three-dimensional structure of DNA.

How an individual scientist comes to focus on a particular problem is largely a matter of the accidents of personal history. Watson was, apparently, interested in biology from an early age. He had been an avid bird watcher and claimed that he first developed an interest in the nature of genes as a senior in college. His interest in DNA in particular seems to have been sparked by his graduate school adviser, Salvador Luria.

The lesson for us is that knowing how a particular scientist came to work on a problem is not likely to be of much use in understanding or in evaluating specific scientific results. It may be interesting as biography and useful for understanding why particular avenues of inquiry were pursued when they were, but that is not our concern here.

Constructing Models

Watson and Crick spent a great deal of time thinking about, talking about, and actually building, a model of DNA. Not all scientists literally build scale models like the wire and metal-plate model Watson and Crick constructed. In a more abstract sense, however, most scientists can be said to be engaged in constructing models of some aspect of the world, if only in their minds. Understanding episodes in science requires some understanding of the particular models being developed. Also, learning to analyze reports of scientific episodes requires an appreciation of the nature and role of models in science. Acquiring such an appreciation will be one of our first tasks.

Deciding Whether a Model Fits

Scientists do not construct models just for the fun of it. Constructing models is part of the process of figuring out how the world works. While working on their models, scientists are all the time trying to decide whether these models actually exhibit a reasonably good fit to the real world. In making such decisions, scientists obviously consider all sorts of facts, particularly the results of careful experiments. But their decision-making process is influenced by many other factors as well. Among these are the desire to be the first to make an important discovery and the fear of being shown wrong. Watson's desire for success clearly outweighed his fear of making mistakes, as demonstrated by his initial enthusiasm for the disastrously ill-fitting, three-helix model. To understand episodes in science, we will have to learn more about how scientists decide whether a model really does fit the world.

Convincing Others

It is not enough for an individual scientist to decide that a model fits. Other scientists must be persuaded to make the same decision. This requires data and arguments that will appeal to other scientists who are approaching the subject with a wide variety of interests, backgrounds, and skills. Franklin, for example, came to the conclusion that DNA has a helical structure long after Watson, Crick, and Wilkins were quite convinced it must. Pauling fairly quickly agreed that the double-helix model was correct after seeing both the model and the x-ray data at Kings. Other scientists in the field quickly agreed after hearing about it or reading the papers by Watson, Crick, Wilkins, and Franklin.

Spreading the Word

Once most of the scientists directly involved in a scientific area decide that a model fits, there begins a much slower process by which that conclusion spreads to the general, nonscientific public. Scientists may be involved in this process, as Watson has been, but so are many others, particularly teachers, journalists, and even filmmakers. It is at this stage of the game that the rest of us learn most of what we know about science. To understand and evaluate this information requires learning how to use what is presented to reconstruct some features of the models and the decision-making processes that went into producing the information in the first place. We turn now to this task.

2.3 MODELS AND THEORIES

Scientists often describe what they do as constructing and using models. Understanding science requires knowing something about models, and how they are used. In fact, there are at least three different types, or uses, of models to keep in mind.

Scale Models

Watson and Crick were helped greatly by actually trying to construct a physical model of DNA. This was a model in the ordinary sense in which model airplanes and doll houses are models. They are all scale models. The big difference between Watson and Crick's model and more familiar scale models was the extreme scale, which in the case of the DNA model was roughly a billion to one. That is, an inch in the model represented roughly one one-billionth of an inch in an actual DNA molecule.

Scale models are widely used in science and even more widely used in engineering. For example, one can learn a lot about the wind resistance of various automotive designs by testing scale models of automobiles in small wind tunnels. This is a lot easier, and cheaper, than building wind tunnels large enough to hold full-sized cars. Nevertheless, when scientists talk about models, they are often not talking about scale models.

Analog Models

In *The Double Helix,* Watson talks about noticing spiral staircases, and of thinking that the structure of DNA might be like a spiral staircase. He also had the example of Pauling's α-helix. Here we would say that Watson was using a spiral staircase and the α-helix as analog models for the DNA molecule. He was suggesting that the DNA molecule is analogous to the α-helix or to a spiral staircase. One might also say that he was modeling the structure of DNA on that of the α-helix or a spiral staircase.

Probably the most famous analog model in modern science is that of the solar system as an analog model for an atom. The nucleus of an atom, containing protons and neutrons, is said to be analogous to the sun. The electrons are said to be analogous to planets circling the sun. There is no doubt that this analogy between the solar system and atoms was extraordinarily fruitful during the first half of the twentieth century. It suggested all sorts of questions that formed the basis of much research. For example, "How fast are the electrons moving around in their orbits?" "Are the orbits circular or elliptical?" In investigating such questions, scientists learned much about atoms. In particular, they learned about many respects in which atoms are not like the solar system. In the end, a good analogy often leads to its own demise.

Analog models are typically most useful in the early stages of research when scientists are first trying to get a handle on the subject. At this point, almost any suggestion as to how they might construct a new model may be helpful. At later stages, when the question turns to evaluating how well the new model fits the real world, the original analog model is less useful. In trying to convince Wilkins and Franklin that they had the right structure for DNA, Watson and Crick did not appeal to features of spiral staircases. Nor did they simply appeal to Pauling's success with the α-helix. Similarly, facts about the orbits of the planets were not used as evidence that the solar system model of the atom is correct. For these evaluations, other evidence was needed.

In sum, thinking about analog models may be very useful in attempting to understand a proposed new model. Such models are much less useful in attempting to evaluate a proposed new model.

Models and Maps

The models most commonly referred to in scientific contexts are theoretical models. In attempting to understand what theoretical models are, it is helpful to invoke an analogy between theoretical models and maps. That is, we are going to use maps as analog models for theoretical models. Maps are more abstract than scale models, but still less abstract than a typical, theoretical model.

It is instructive if at this point the student can produce his or her own map. This could be a map showing a trip between home and school, between a dormitory and a classroom building, or between home and work. Figure 2.5, for example, is a map that depicts part of the University of Minnesota campus, including the main library and the building housing a particular department. Draw your own map before proceeding.

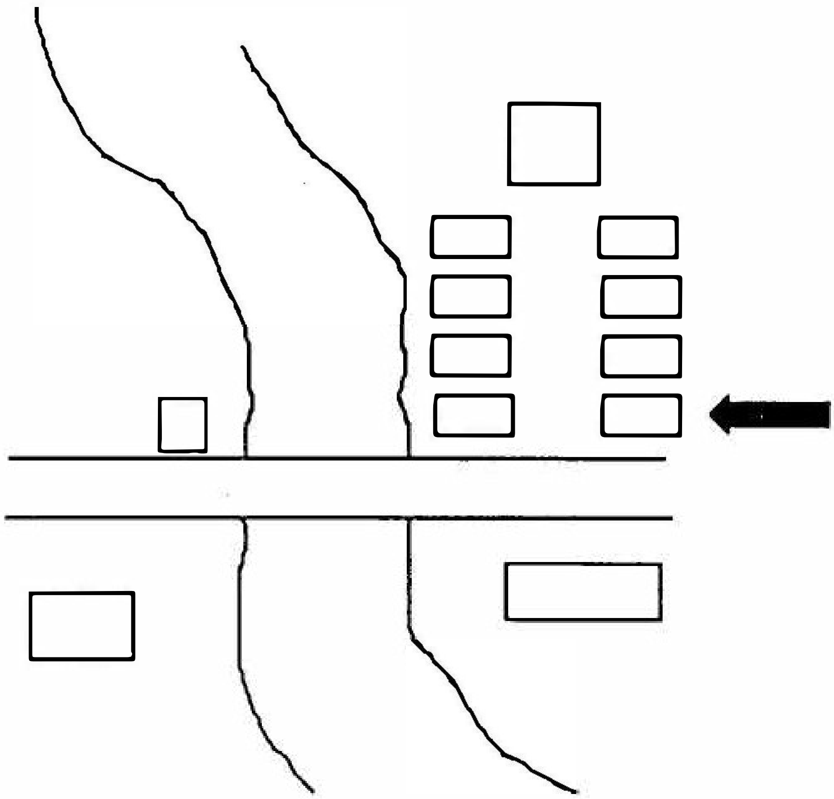
In the map, as shown in Figure 2.5, there is a solid arrow. What is the object to which the arrow is pointing? Stop and answer this question before reading any further.

The standard answer to the question is that the arrow is pointing to a building, presumably the university building in which the department office is located. What would you think if someone told you that your answer is mistaken? Not only is it mistaken, it is not even close to being right. It is totally off base. No doubt, you would begin to suspect that there is some trick being played on you. You would be right. But the trick has an important point.

The answer is that the arrow is pointing to a rectangle drawn on the page. That is, quite literally, what the arrow is pointing to. The reason one is inclined to say that the arrow points to a building is that it is pretty clear from the

Figure 2.5

A partial sketch of the Twin Cities campus of the University of Minnesota.



map and accompanying text that the rectangle represents a building. One, therefore, interprets the arrow as pointing to what the rectangle represents, rather than to the rectangle itself.

The point of this exercise is that a map is not the same thing as what it represents. In the case of maps, no one is likely to make this mistake. After all, one can fold up a street map and put it in one's pocket. One cannot fold up the city and put it in one's pocket. Nor is one likely to mistake a scale model for the thing modeled. Surely no one was in danger of confusing Watson and Crick's wire and tin scale model for a real molecule of DNA. Theoretical models, as we shall see, are another matter.

Granting that a map is distinct from what it maps, what is the relationship between the two? It is true, but not very informative, to say that the map maps the area mapped. It is somewhat more informative to say that the map represents the area mapped. The next question is, how does a map manage to represent a particular space?

The first part of the answer is that a map exhibits a particular similarity of structure with the space mapped. In the case of maps, the particular similarity of structure is spatial. The spatial relationships among marks on a street map, for example, correspond to the spatial relationships among the streets in the city represented by the map.

The second part of the answer as to how a map manages to represent the area mapped is that we have a whole set of fairly well understood social conventions for constructing and reading maps. Without these conventions, a map would be just a piece of paper with lines drawn on it. The conventions for street maps are so well known that most people are not even aware of them. But this is a special case. Few people know the conventions for interpreting Watson and Crick's scale model of DNA. One has to know a lot of physical chemistry to be able to interpret a particular tin plate as representing a purine base. Franklin could do it easily and quickly recognized that the three-chain model did not have enough places to attach water molecules.

The analogy between maps and models suggests further interesting questions. One is, "Could there be a perfect map, for example, a perfect map of Chicago?" The answer depends on what one means by a "perfect map." Suppose it means a map that contains a perfectly accurate representation of every feature of the city. Is that possible? Hardly. To represent every feature would mean representing every alley, house, garage, tree, bush, broken sidewalk, and abandoned car. It would mean representing not just the locations of buildings, but their height as well. That is an impossible task. So one way in which maps are not perfect is that they are incomplete, that is, they represent only selected features of their subject, such as streets, and ignore others, such as heights of buildings.

Restricting our attention to those features that are represented, there remains the question of how accurately those features are represented. For example, does the map accurately portray the relative distances between the Water Tower and the Chicago River, or between Michigan Avenue and Halsted Street? And is it accurate to the nearest ten yards? yard? foot? inch? Clearly no map is going to be perfectly accurate down to the millimeter.

In summary, a map can be used to represent a place because there exists a set of social conventions that allow one to interpret features of the map as representing features of the place. All maps are incomplete in that they do not represent all features of the place represented. And no map gives a perfectly accurate account of the features that are represented. Nevertheless, there remains a similarity of structure between the map and the place represented. They are similar in some specifiable respects and to some specifiable degree of accuracy. All these things hold, as well, for the relationship between theoretical models and the parts of the world they represent.

Theoretical Models

A theoretical model is part of an imagined world. It does not exist anywhere except in scientists' minds or as the abstract subject of verbal descriptions that scientists write down. When Watson was building the three-helix model, he could have written down a description of what a DNA molecule would be like if it were like this model. His description could have begun: "The model has three sugar phosphate backbones that twist in a helical structure with bases arranged . . . ." This description, obviously, cannot describe a real DNA molecule, because we now know that DNA has only two chains, not three. What it describes, rather, is a possible molecule that turned out not to exist at all.

Watson and Crick did build a scale model with three chains. What is the relationship between that scale model and the corresponding theoretical model? The scale model can be used in place of words to characterize the theoretical model. One simply says: "The theoretical model has three sugar phosphate chains with bases arranged like this," and, then, points at the scale model. This strategy works because there is a similarity of structure between the scale model and the theoretical model. One can understand that similarity if one knows the conventions used in building the scale model (such as that red wires stand for hydrogen bonds).

Why can we not just stick to scale models and dispense with the notion of theoretical models? Because not all theoretical models have corresponding scale models. Watson, for example, never completed a scale model of the two-chain molecule with like bases bonded together. But this model existed as a theoretical model. Watson even described it in a letter to Max Delbrück. More fundamentally, scientists construct theoretical models of a whole variety of complex processes for which it would be difficult if not impossible to build working scale models.

In any case, the idea of a theoretical model should not be mysterious. We all create something like theoretical models all the time. For example, one can imagine giving a party, including imagining who comes with whom and who says what to whom. Here one is constructing a theoretical model of a complex social event. The party may never occur or, if it does, it may be nothing like originally imagined. In the process of doing science, scientists imagine all sorts of complicated things and processes, including large molecules, such as DNA.

Theoretical Hypotheses

The most important question one can ask about a theoretical (or scale) model is whether it is, indeed, similar to the world in the intended respects and to the intended degree of accuracy. By way of shorthand, we will often simply ask, does the model fit the world as it was intended? Even more simply, does the model fit?

There is another way of talking about the fit between models and the world that is often used by scientists, journalists, and other commentators on science. To accommodate this way of talking, we will have to introduce some additional terminology.

When scientists make the claim that their model is, in fact, similar to the world in the desired respects, we can say that they have formulated a theoretical hypothesis and are claiming that this theoretical hypothesis is true. A theoretical hypothesis is a statement (claim, assertion, or conjecture) about a relationship between a theoretical model and some aspect of the world. It asserts that the model is, indeed, similar to the world in indicated respects and to an implied degree of accuracy. If the model is similar to the world, as claimed, then the theoretical hypothesis is true. If the model is not similar to the world, as claimed, then the theoretical hypothesis is false.

For example, at one point, Watson and Crick formulated the theoretical hypothesis that DNA has a helical structure with three polynucleotide chains. That hypothesis was shown to be false. They later formulated the theoretical hypothesis that DNA has a two-chain structure. That hypothesis turned out to be true. In general, asking whether a specified theoretical hypothesis is true or false is just another way of asking whether the corresponding theoretical model fits the real world. This relationship is pictured in Figure 2.6.

"Truth" is a heavy-duty concept whose meaning has been debated by philosophers and others for two thousand years or more. Theories of the nature of truth are discussed routinely in courses in logic and philosophy. The above discussion of the truth or falsity of theoretical hypotheses falls into the category of what is usually called the correspondence theory of truth. But there is no need to enter these troubled waters here. For the practical purposes of understanding and evaluating reports of scientific findings, it is sufficient to think in terms of the fit between a model and the world. Here the analogy of the fit between a street map and the streets of the corresponding city provides a sufficient guide. If that is not enough, one can always fall back on the more restricted relationship between a scale model and the real thing. If that does not work, an abstract inquiry into the nature of truth is unlikely to be of much help.

Finally, in everyday speech the word "hypothesis" often carries the connotation of a claim that is highly speculative—a conjecture without any real support. Thus, one may reply to a claim one disputes by saying, "Well, that is one hypothesis." Here the implication is that there are other, equally plausible, hypotheses that one might propose.

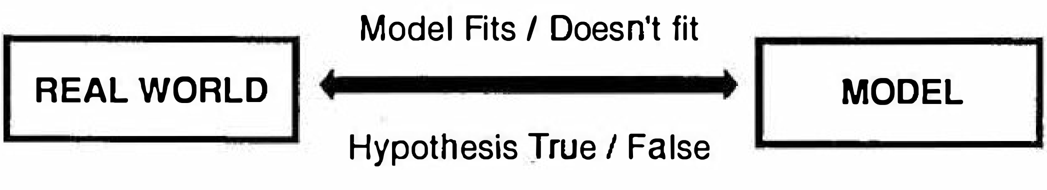
For the purpose of developing a systematic framework for understanding and evaluating scientific findings, it is best to ignore this way of talking. For us, the claims made in behalf of both the three-helix model and the two-helix model are both hypotheses. Indeed, in our preferred terminology, all general scientific claims are hypotheses. The difference is that some hypotheses are well supported by the evidence, and others are not. The important thing is to learn to distinguish between those hypotheses that are well supported and those that are not. That tells you which hypotheses it is reasonable to regard as true.

Theories

Everyone knows that scientists produce theories. Yet up to now I have not explicitly talked about theories, as such. The reason is that "theory" is a quite vague and often ideologically loaded term. The main reasons for calling

Figure 2.6

A picture of the relationship between a model and the real world. The hypothesis that the model fits the real world may be either true or false depending on whether or not the model actually fits.



something a theory may be to give it honorific status or, alternatively, to call it into question. Which of the two functions is served by using the word "theory" depends largely on the intentions of the speaker and the nature of the audience. Here we will attempt to employ a more neutral analysis. We already have in hand all the elements of such an analysis.

For our purposes, a scientific theory has two components. One is a family of models that may include both scale models and theoretical models. The second is a set of theoretical hypotheses that pick out things in the real world that may fit one or another of the models in the family.

In 1953, there was basically just one model, the double helix, and one chemical substance, DNA, to which this model was applied. But several years later, when people began talking more about "the theory of molecular biology," there was a whole family of similar models, and a range of other substances to which they were applied, including RNA (ribonucleic acid). Later, we will encounter many other theories that clearly include many distinct, but similar, models.

Like the word "hypothesis," the word "theory" often carries the common meaning of something speculative. It is common, for example, for those who question the theory of evolution to say that evolution is merely a theory and not a fact. Here again, we shall reject this common usage. For our purposes, the theories of molecular biology and the theory of evolution are all theories. Whether they are also facts depends on whether the corresponding models fit the world or, alternatively, whether the corresponding theoretical hypotheses are true. If so, they are facts. If not, they are not facts.

The important question is how well each theory is supported by the evidence. Here the relevant distinction is not between theory and fact but between theory and data, a distinction that is crucial for evaluating theoretical hypotheses and, thus, theories.

2.4 DATA FROM THE REAL WORLD

Everybody knows that in order to determine whether a proposed model fits the world one needs some information about the part of the world in question. But not all information is relevant. We will use the term "data" (singular, "datum") to refer to all the special information that may be relevant to deciding whether the model in question does fit. There are several, general characteristics that data must have.

The first feature that information must have in order to function as data is that it be obtained through a process of physical interaction with the part of the real world under investigation. Interaction may be active, as when one does experiments on the materials in question. Or the interaction may be passive, as when radio astronomers measure radio frequency signals from distant galaxies. In either case, the data result from a physical interaction with the relevant part of the real world.

A second general feature of data, as opposed to mere information, is that relevant differences in the data can be reliably detected. Detection may be a simple matter of looking, as when one observes a chemical solution change from blue to green. More often, detection requires elaborate instruments that produce outputs among which a scientist can discriminate just by looking. Often these outputs are computer printouts of graphs or tables of numbers. Figure 2.7 provides a schematic picture of the relationship between the real world and some data.

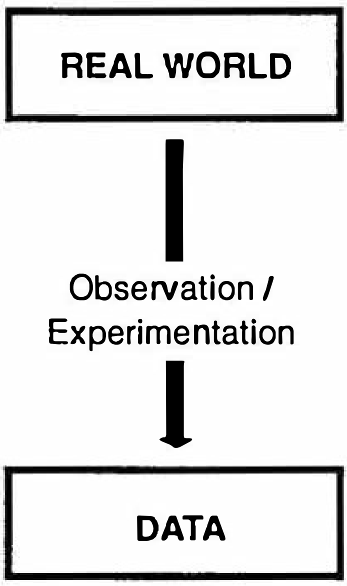
Among important bits of data in the story of the double helix were Chargaff's results on the one-to-one ratio of particular purines to specific pyrimidines in DNA, Franklin's results on the amount of water in DNA samples, and, of course, Franklin's x-ray pictures of DNA. In each case these data were obtained by actually working with samples of DNA.

Among information used by Watson and Crick that did not count as data in favor of their model of DNA was Pauling's discovery of the helical structure of α-kerotin. This information was influential in the decision by Watson and Crick to investigate helical models of DNA. Acquiring this information, however, required no physical interaction with samples of DNA. It could not, therefore, play a role as data for their hypothesis that DNA fits a double-helical model.

Here is one respect in which the analogy between maps and theoretical models breaks down. One can discover that a street map is deficient simply by finding a street that is not on the map. That requires no special skills or instruments. One can just look and see. Most models in modern science are not like that. Modern science typically investigates things that are very small (DNA), very large (our galaxy), very far away (distant stars), or otherwise inaccessible (the center of the earth). In all these cases, one cannot just look to see whether a proposed model fits. This fact has profound implications for how one might evaluate whether a model fits the real world.

Figure 2.7

A picture of the relationship between the real world and data generated through a physical process of observation or experimentation.



2.5 PREDICTIONS FROM MODELS

The fact that one can only interact indirectly with the objects of a scientific investigation means that one can only investigate the fit between parts of the objects and some limited aspects of a proposed model. It is, therefore, important to be able to figure out what kind of data an object that did fit the proposed model would produce in the circumstances of a particular experiment. Scientists often speak of using a model to make predictions about what kind of data would be produced. This requires some explanation.

Sometimes, scientists use a model to make predictions in the literal sense of trying to figure out ahead of time what the data will be. But often, predicting the data simply means using the model to determine what the data should be like, even though the experiment has already been done. For example, the double-helix model was said to "predict" the Chargaff ratios even though Chargaff's experiments were done several years earlier. Similarly, the model allowed Crick to calculate the kind of x-ray pattern a double helix would produce and thus, in this respect, "predict" the kind of picture that Franklin, at that point unbeknownst to Watson and Crick, already had in her possession.

The example of Crick's prediction of the x-ray pattern exhibits another important feature of making predictions from models. It requires more than just the model in question. It also requires that one have a well-attested model of the experimental setup. Thus, Crick had also to have a good model of how x-rays are diffracted by atoms. Otherwise, he could not calculate what the pictures should look like if the x-rays are being diffracted by a helically shaped molecule.

Figure 2.8

A picture of the relationship between a model and a prediction obtained by reasoning about the model in the given experimental context.

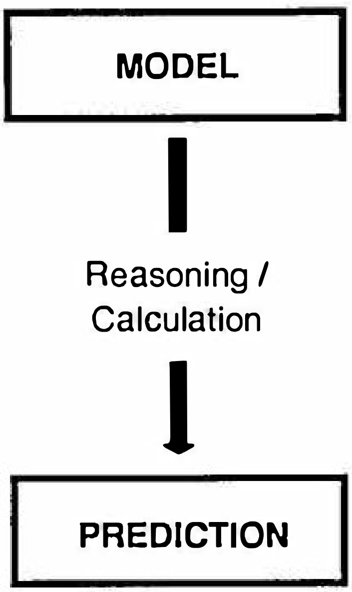


Figure 2.8 provides a schematic picture of the relationship between a model and a prediction derived from the model. Note that here the arrow represents not a physical interaction with the world but merely a process of reasoning or calculation.

2.6 THE ELEMENTS OF SCIENTIFIC EPISODES

We are now in a position to use the results of the previous three sections to construct our own simple model of a scientific episode involving a particular theoretical model (or family of similar models). There are four ingredients: (1) real world; (2) model; (3) data; (4) prediction. It is helpful to arrange these elements as shown in Figure 2.9. There are four relationships among the elements.

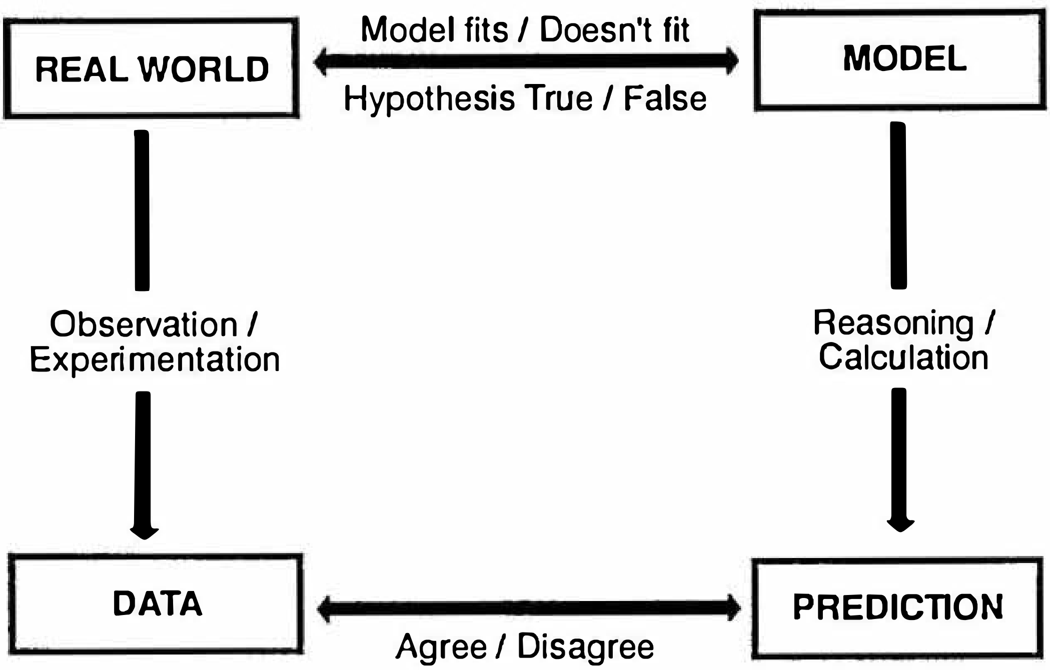
First, the relationship between the real world and the model is expressed by a theoretical hypothesis that asserts that the model fits the real world. It is understood that the model fits only in some respects and then only to some degree of accuracy. If the model does not fit accurately in the intended respects, then the theoretical hypothesis is false.

The real world and the data are related by a physical interaction which involves observation and, perhaps, experimentation. The model and the prediction are related by reasoning or calculation.

Finally, Figure 2.9 contains a relationship not previously noted, namely one between the data and the prediction. If what is going on in the real world, including the experimental setup, is similar in structure to the model of the world, including a model of the experimental setup, then the data and the

Figure 2.9

The four elements of an ideally complete report of a scientific episode involving a theoretical hypothesis.



prediction should agree. On the other hand, if the real world and the model of it are not similar in the relevant respects, then the data and the prediction may disagree.

Whether the data and the prediction agree is crucial to evaluating the truth of the hypothesis that the model fits the real world. But, as we shall soon see, there is more to evaluating the fit of a model than just agreement between data and predictions.

It should be noted that Figure 2.9 represents a picture of fully developed, scientific episodes in that it contains all four elements. Many episodes, and thus many reports of scientific findings, do not include all four elements. It is common, for example, to find reports that describe only the part of the real world under investigation together with some new data. There may be no mention of models or predictions. Similarly, one often finds discussions of new models of real world entities or processes with no mention of data or predictions. Occasionally one finds accounts of models of real world things that include predictions but no discussion of data. It is possible, but unlikely, that one would find a report including descriptions of real world things, models, and data, but no mention of predictions from the model. In such a case, the presumption would be that there is a prediction that, in fact, agrees with the data. If it did not, that would surely be noted.

One may learn a lot from incomplete episodes or reports. Unless all four elements are present, however, it is difficult to evaluate the relevance of the data to the crucial question of whether the model adequately represents the real world.

2.7 EVALUATING THEORETICAL HYPOTHESES

We are now ready to begin developing a general scheme that can be used by nonspecialists to evaluate scientific hypotheses as reported in various popular and semitechnical sources. We will continue with examples from the story of the double helix. Later in this chapter, we will work through several completely different examples.

The basic idea behind the evaluation of hypotheses is to use the agreement or disagreement between data and predictions, information that is relatively accessible, to evaluate the fit between a model and the real world, something that is not directly accessible. In the ideal case, there are only two possible cases. Either the prediction and the data agree or they do not agree. We will treat these cases separately, beginning with the case in which the prediction and data disagree. That turns out to be the simplest of the two cases.

In less than ideal cases, it may not be clear whether the prediction and data agree. Agreement may be a matter of degree. In such cases, it is difficult for a nonspecialist to make any independent evaluation of whether the model in question fits the world or not. Here the best one can do is rely on the informed judgment of specialists. Unfortunately, when agreement between data and predictions is unclear, specialists often disagree among themselves about how well the model might fit the real world. In such cases, the only safe course for the nonspecialist is to regard the data as inconclusive and to suspend judgment about the model until more decisive data become available. If use of the model in question is relevant to some practical decision that needs to be made, the problem, then, is to make that decision in a manner that takes proper account of the uncertainty as to whether the corresponding theoretical hypothesis is true. This latter sort of situation will be treated in Part Three of this text.

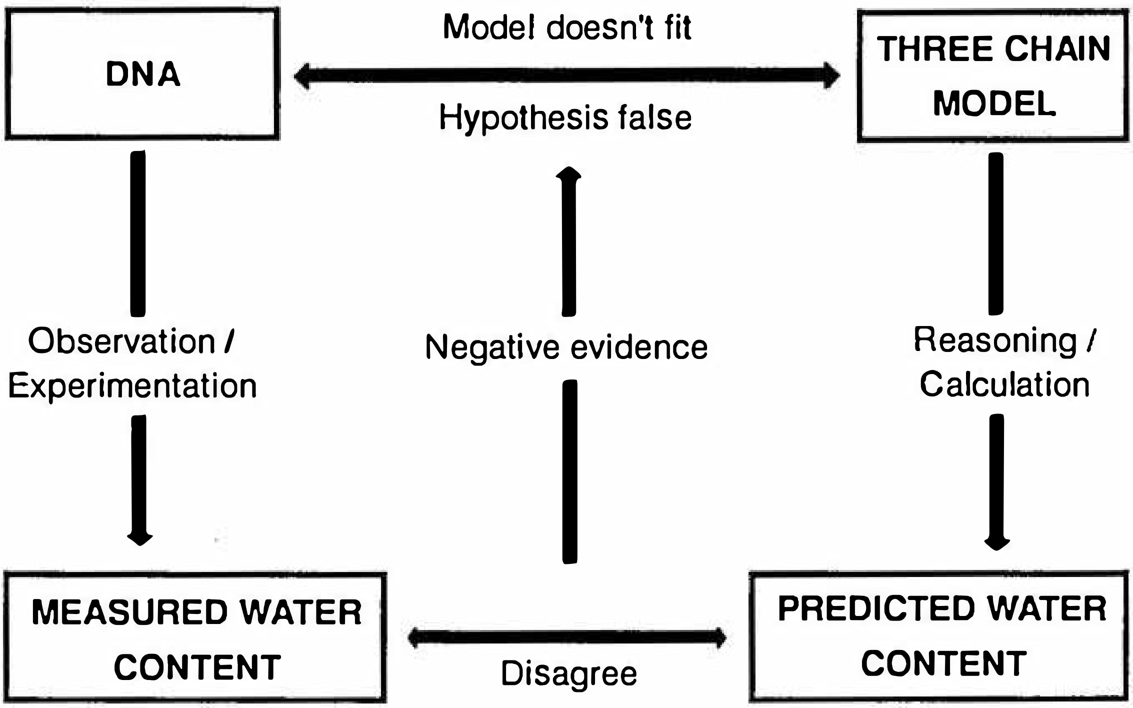
Evidence That a Model Does Not Fit the Real World

The story of the double helix provides a clear example of a model that was judged not to fit the real world—the three-chain model of DNA. Here the decisive data were Franklin's experimental measurements of the amount of water contained in samples of DNA. The three-chain model yielded a prediction as to how much water such a DNA molecule would accommodate. This could be determined by examining the scale model as long as one could interpret the model and knew enough physical chemistry to judge where water molecules might fit into the structure. The trouble was that the prediction from the model gave a value for the amount of water that was only one-tenth the amount Franklin had measured. So there was a clear disagreement between the experimental data from real samples of DNA and the prediction based on the three-chain model of DNA. This situation is pictured in Figure 2.10.

In this case, one is tempted to conclude without further ado that the hypothesis is false. That is, real DNA molecules do not closely resemble the proposed three-chain model. Franklin immediately drew that conclusion, although Watson and Crick took a little longer to come around. As

Figure 2.10

The elements of the episode involving Watson and Crick's three-chain model of DNA.



nonspecialists reading about this episode, we could just follow their lead. If we are attempting to reach an independent evaluation, however, we cannot be quite so decisive. There are two possibilities that militate against so hasty a conclusion, neither of which tend to be accessible to a nonspecialist.

One possibility is that the data were mistaken. That is, the experiment yielded a mistaken value for the amount of water. There are all kinds of things that, unbeknownst to anyone, might have gone wrong with the experiments so as to yield a value for the amount of water ten times greater than the actual amount. Only people with much experience with the actual apparatus and experimental techniques can reliably judge how likely it is that something was seriously wrong with the experiment.

A second possibility is that through misunderstanding of the model itself, or because of a mistaken model of the experimental apparatus, the prediction was mistaken. A proper understanding of the model, or the experiment, might have yielded a predicted value in agreement with the actual data. Again, this is something for which a nonspecialist must rely on the judgments of the experts.

For these reasons, we will take reports of clear disagreement between data and predictions as a basis for concluding only that there is good evidence that the hypothesis is false. That is, there is a good reason, although not necessarily a conclusive reason, for believing that the model does not adequately represent the real world.

Evidence That a Model Does Fit the Real World

One of the nice features of the double-helical model for DNA is that having the sugar-phosphate backbones on the outside left a lot of places for water molecules to attach themselves. So the double-helix model yielded a prediction for the amount of water in agreement with Franklin's data. Should one take that as a ground for thinking that the double-helix model fits?

As a matter of fact, Watson and Crick did not treat the agreement between the amount of water predicted by the double-helix model and the measured amount of water as a basis for arguing in favor of the double-helix hypothesis. Why not? Because they knew many possible ways to build models with the required places for water. It could be done with a variety of three-helix models, for example, as long as one put the backbones on the outside. Thus, predicting the measured amount of water provided no basis for distinguishing the two-helix model from a variety of three-helix models. There was, therefore, no basis for regarding this agreement between prediction and data as evidence that the two-helix hypothesis, rather than some three-helix hypothesis, was true.

This explains why the x-ray data were regarded as being so important. According to Crick's calculations, a double helix should produce a quite distinctive pattern, that is, a pattern unlikely to result from molecules with a significantly different structure. Thus, agreement between the predicted x-ray pattern and actual x-ray pictures provided a reliable basis for distinguishing between a double-helical structure and a variety of other structures. In this case, therefore, agreement between the prediction and the data did provide evidence in favor of the double-helix hypothesis. The elements of this case are pictured in Figure 2.11.

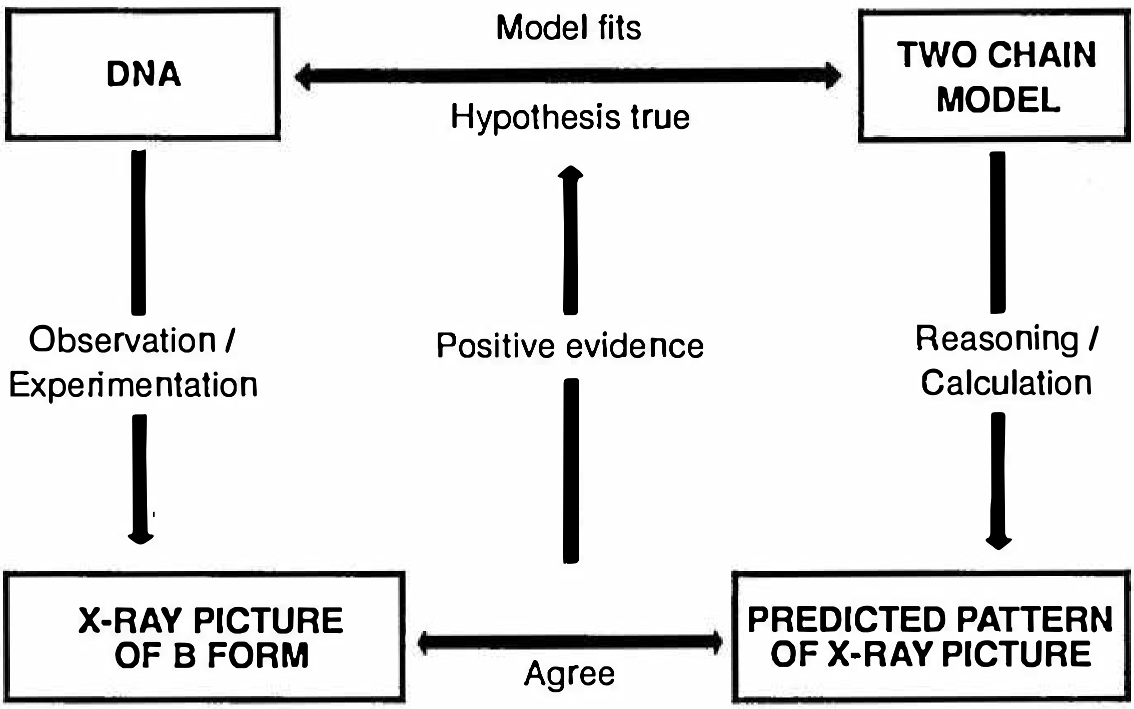
The moral of this story is that mere agreement between a prediction and some data is not enough to provide a basis for thinking that a theoretical hypothesis is true. Agreement counts only when it would not have been very likely if the hypothesis had been false, which is to say, if some significantly different model had provided a better fit to the real world. Ignoring this moral puts one in great danger of thinking that one has evidence in favor of what is, in fact, a false hypothesis.

At this point, one might have the following worry. No matter what the data happen to be, and no matter what model is being considered, is it not always possible at least to imagine some completely different model that, nevertheless, just happens to yield the same prediction as the model under consideration? Does this not mean, therefore, that agreement between a prediction from a model and observed data can never provide any basis for thinking that the model fits the world?

This sort of difficulty has been voiced by scientific skeptics since the heyday of Greek astronomy nearly two thousand years ago. The reply is that in the process of doing science it is never enough for an alternative model to be imaginable in the abstract. It must be plausible against the general background of the models being used at the time by scientists working in the same general area. This means that in principle there is almost always some possible model that would provide a better fit than any model currently

Figure 2.11

The elements of the episode involving Watson and Crick's two-chain model of DNA.



under active consideration. Nevertheless, the model currently regarded as the best fitting model will generally be the best fitting of those models that have actually been considered. It is unreasonable to expect more from any scientific investigation.

The above theoretical worry has a more practical consequence. It is often difficult for a nonspecialist to judge independently whether there are very many other plausible models that would also yield predictions in agreement with existing data. In many cases, therefore, all the nonspecialist can do is rely on the reported judgments of specialists as to whether there are any such alternative models. Worse yet, these judgments are often more implicit than explicit. One must read very carefully to determine whether there is any consensus on the availability of other plausible models yielding the same predictions. With experience, however, one can learn to recognize hints from which one can infer the existence of the relevant consensus.

Finally, as in the case of disagreement between predictions and data, we cannot take agreement between a prediction and data as a definitive basis for deciding whether the model fits. The most we can conclude is that the agreement between prediction and data provides good evidence for thinking that the model fits. The possibility of serious experimental errors, or of mistakes in determining what prediction the model yields, precludes a definitive conclusion on the basis of any single experiment. Even for cases in which it seems to the nonspecialist that scientists themselves reach a definitive conclusion on the basis of a single experiment, they may be relying on knowledge of other experiments, as well as considerable experience with the experimental and theoretical techniques involved.

2.8 A PROGRAM FOR EVALUATING THEORETICAL HYPOTHESES

In this section we will reduce the process of evaluating reports involving theoretical models to six easy steps. This does not require learning anything new. It is just a matter of organizing what we have done into a kind of program for doing an analysis. The advantage of developing such a program is that one can have in one's head a simple, uniform scheme for the evaluation of all sorts of scientific reports, a scheme that is easy to remember and to apply.

The program has two parts. The first four steps instruct one to identify the four basic elements in a complete episode. These steps provide a basis for *understanding* the case. If all four elements are reported, one can go on to the final two steps, which constitute an *evaluation* of the model. If not all of the four elements are identifiable, it may be difficult to perform any evaluation at all.

Some reports mention more than one model and prediction, and report different sets of data. The program assumes that you have identified a single model and prediction, and a single set of data, which is to be the focus of your analysis. Often, it is clear from the context which model or which set of data forms the primary focus of the report. If this is not clear, you might have to run through the program more than once, evaluating different models or considering different data sets. Later, we will consider a special case in which more than one model is evaluated in a single analysis.

The Program

Step 1. Identify the aspect of the real world that is the focus of study in the case at hand. These are things or processes in the world that you should be able to describe in your own words with, perhaps, just a bit of existing scientific terminology.

Step 2. Identify a theoretical model used to represent the real world. Describe the model, using scientific terminology as needed. A diagram may be helpful in presenting a model. Indeed, a diagram may be a version of a model.

Step 3. Identify data that have been obtained by observation or experiment involving the real world objects of study.

Step 4. Identify a prediction, based on the model, that says what data should be obtained if the model actually provides a good fit to the real world.

Step 5. Do the data agree with the prediction? If not, conclude that the data provide good evidence that the model, in its present form, does not fit the real world. If the data do agree with the prediction, go on to Step 6.

Step 6. Was the prediction likely to agree with the data even if the model under consideration does not provide a good fit to the real world? This requires considering whether there are other clearly different, but plausible, models that would yield the same prediction about the data. If there are no such alternative models, the answer to the question is "no." In this case, conclude that the data do provide good evidence that the model does fit the real world. If the answer to the above question is "yes," conclude that the data are inconclusive regarding the fit of the model to the real world.

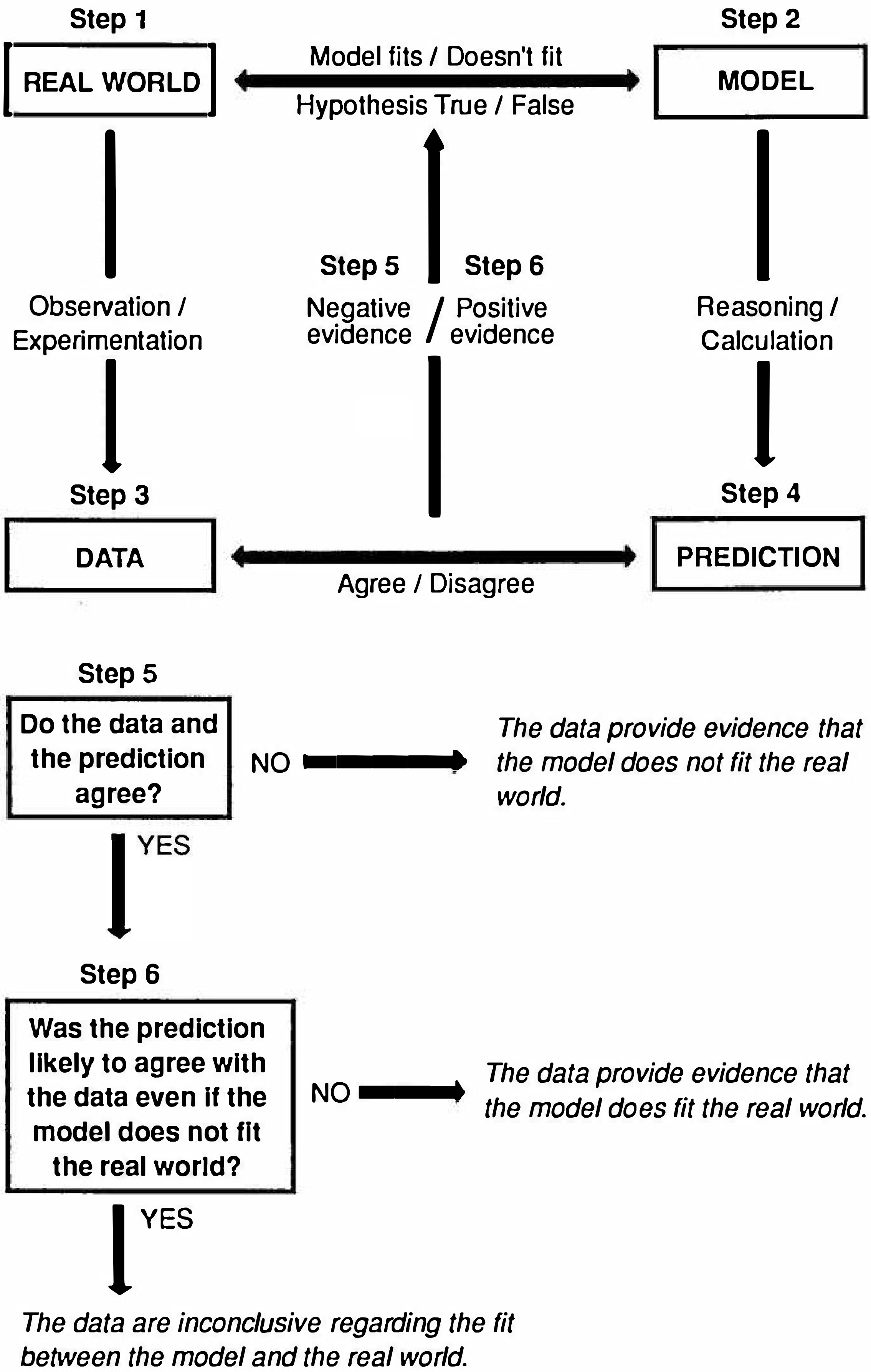
Figure 2.12 exhibits the above program as a "flow chart." This is easier to remember than the verbal instructions, and may help you to recall the details of the written program.

Although the first four steps in the flow chart are numbered in the order given in the program, the chart itself suggests that there is nothing sacred about this particular ordering. What is important is that one identify all four elements before proceeding to the evaluation in Steps 5 and 6, where the order is important. As a general rule, it is advisable to begin with the real world objects. This tells you what the whole episode is about. But the order in which you identify the remaining three elements may vary. The structure of a particular report, particularly the order in which information is presented, may make one order seem to you more natural than another. It is fine for you to follow that order, as long as you clearly identify the elements of the case.

Finally, the program is only a guide, an outline to help you evaluate reports of scientific findings. Learning to use it effectively requires experience. The only way of acquiring that experience is to do many evaluations.

Figure 2.12

A flow chart corresponding to the program for analyzing reports of scientific episodes involving theoretical hypotheses.



2.9 WHY THE PROGRAM WORKS

Why should you use the suggested program for evaluating theoretical hypotheses? Because, by following the program, you will have a good chance of reaching the right conclusion, whatever the right conclusion happens to be. There are three possibilities: (1) the data may provide evidence that the model does not fit the world, (2) the data may provide evidence that the model does fit the world, (3) the data may be inconclusive. In each case, the program will lead you to the appropriate conclusion.

> ***Case 1.*** When the data and a prediction disagree, typically, that is because there are some respects in which the model does not adequately represent the real world. Steps 4 and 5 of the program are particularly relevant to this case. Step 4 is relevant because it requires you to consider whether there is a prediction that is clearly derived from the model. That is important because, if the prediction has little to do with the model, the relationship between the prediction and the data would not be much of an indicator of any similarity, or lack of similarity, between the model and the real world. Step 5, then, asks you to consider whether the data and the prediction agree. If they do not, that is an indication that something is wrong with the model, and the program directs you to the appropriate conclusion.
>
> ***Case 2.*** If the data and prediction agree, the program directs you to Step 6. It asks you to consider the question: Was the prediction something that would be likely to agree with the data, even if the model under consideration does not adequately fit the real world? This question forces you to seek information in the report regarding other plausible models that could yield the same prediction. If there seem to be no other such models, then agreement between the prediction and the data is a pretty reliable indicator of adequate fit between the model under consideration and the world. The program directs you to conclude that the data provide evidence of such a fit.
>
> ***Case 3.*** What if the agreement between the data and the prediction was quite likely even if the model in question does not fit the world? That is, what if there are other models that yield the same prediction, and whose fit to the world is, initially, as plausible as the model under investigation? In this case, the agreement between the data and the model under study provides little basis for thinking that the model under investigation fits the world better than any of the other models. On the other hand, the existence of other possible models provides no basis for thinking that the model under investigation does not adequately fit the world. So the agreement between prediction and data turns out to be inconclusive regarding any fit between the model in question and the world. The program directs you to that conclusion.

In sum, whatever the appropriate evaluation happens to be, the program will lead you to draw the right conclusion.

2.10 HOW THE PROGRAM WORKS: THREE EXAMPLES

We will now work through three examples that together illustrate the three possible results of applying the analysis program for reports involving theoretical hypotheses. Each example begins with an actual report that has been only slightly edited and shortened for the purpose of these illustrations. Readers are strongly urged to attempt their own analyses by applying the program to the reports before going on to read the analyses in the text. One always learns more from seeing how something should be done if one has already tried it for oneself.

A Case of Negative Evidence

Gene Analysis Upsets Turtle Theory

A widely accepted theory explaining why green turtles migrate 1250 miles to an island in the middle of the Atlantic Ocean to lay their eggs and then swim back to Brazil is invalid, researchers have concluded from genetic analyses.

According to the popular theory, the turtles started coming to the island more than 40 million years ago, when it was close to the shore of South America, and just kept coming as the island moved farther and farther away. But the new research concluded that the turtles had been using the island for only a few tens of thousands of years.

The now invalid explanation was advanced fifteen years ago by Archie Carr and Patrick J. Coleman. It was a bold idea, based on knowledge that the Atlantic Ocean was born 70 million years ago and began spreading. This gradually increased the isolation of Ascension Island, formed by volcanic activity on the ocean's centerline.

As long as the Atlantic was narrow, according to this hypothesis, remote ancestors of today's green turtles had no trouble reaching the island and laying their eggs in its beaches. They then returned to the shallow waters along the Brazil coast to feed on its marine grasses. After 40 million years, the ocean had grown to substantial width, but, according to the hypothesis, the turtles continued to reach the island by some mysterious form of navigation that still enables them to find it.

Now three scientists have examined the extent to which genetic material in the Ascension Island turtles has changed from that of the same species elsewhere. The difference, they believe, is far too small to have evolved over 40 million years. But it is sufficient to show that the Ascension Island turtles are a distinctive group from those nesting at other Atlantic sites, having probably used the island for not more than a few tens of thousands of years.

A comparison was made of a specific locus (the mitochondrial DNA) in the genetic material of turtles from four widely separated regions of the earth. Earlier work had indicated how fast, in a particular population, subtle changes in this material occur over centuries or millions of years. In the 1987 nesting season, eggs, or turtle hatchlings, were taken from twelve nests on French

Frigate Shoal in the Hawaiian Islands, ten on Hutchinson Island off Florida, eight on Aves Island off Venezuela, and sixteen on Ascension Island in the mid-Atlantic.

The turtles in Hawaii had presumably been isolated from the other locations since the Isthmus of Panama formed about 3 million years ago. Their DNA, as expected, was most distinctive and provided an index of how fast turtle DNA becomes modified.

Analysis

Step 1. The real world subject matter is the population of turtles that has been found to migrate between the coast of Brazil and Ascension Island 1250 miles out into the middle of the Atlantic Ocean. The question scientists have asked about these turtles is how they ever came to make such a long and difficult trip to lay their eggs.

Step 2. According to the previously standard model, the members of this species of turtles began visiting the island roughly 40 million years ago when it was quite near the coastline of Brazil. But over the next 10 to 20 million years the island moved farther out to sea as the floor of the Atlantic Ocean spread out, moving South America farther and farther from the middle of the ocean. The turtles, in this model, kept going to the island even though the trip got longer and longer.

Step 3. Examination of parts of the DNA of turtle eggs taken in 1987 from four different groups of turtles around the world provides a measure of how long a particular population of turtles has existed as a distinct population. By this method, it was determined that the Ascension Island turtles have existed as a distinct population for only a few tens of thousands of years.

Step 4. The model predicts that any measurements of the age of this population of turtles should yield an age of roughly 40 million years. According to the model, that is how long ago the turtles would have had to begin migrating to Ascension Island.

Step 5. The data and the prediction clearly disagree. The prediction is that the age of the population should be measured at roughly 40 million years. The data puts the age at more like 40 thousand years. That makes the prediction roughly a thousand times greater than the data indicate! The data, therefore, provide clear evidence that the model fails to fit the real world. As the headline says, the gene data upsets the turtle theory.

Note that the analysis has only five steps. The flow chart makes it clear that whenever the data and prediction disagree, your analysis will terminate at Step 5. Note also that, strictly speaking, the prediction is not that the turtle population is 40 million years old. Rather, the prediction must be about the data that would be expected if the model fit the world. So the prediction is that the data obtained by examining DNA samples from turtle eggs would be of the type that is known to indicate an age of 40 million years.

A Case of Positive Evidence

New View of the Mind Gives Unconscious an Expanded Role

For decades, mainstream research psychologists suppressed the notion that crucial mental activity could take place unconsciously. But now, in the wake of exciting new studies, experimental psychologists are taking the unconscious more seriously. Among the most influential of the new studies are the investigations into the role of the unconscious in the visual perception of objects and words.

One of the main researchers in this new area is Dr. Anthony Marcel of Cambridge University. He has developed a model of unconscious perception in which the unconscious mind perceives and remembers things of which the conscious mind is unaware. One of the most impressive tests of this model involves what is called "unconscious reading."

In these experiments, Dr. Marcel flashes a word on a screen for a very short time. In addition, the word of interest is masked by being surrounded with other nonsense words, such as "esnesnon." When asked directly, the subjects were unable to say what real word appeared on the screen. Dr. Marcel then asked his subjects to guess which of two words looks like the masked word. For example, the masked word might be "blood" and the two choices for look-alikes might be "flood" and "week." The subjects were correct in their guesses an astonishing 90 percent of the time.

Analysis

Step 1. The aspect of the real world being investigated is the human mind, particularly human perceptual abilities.

Step 2. The model is not described in much detail. The important thing is that, in Dr. Marcel's model, the human mind has an unconscious component that "perceives and remembers things of which the conscious mind is unaware."

Step 3. The data come from the unconscious reading experiments. In these experiments, the subjects guessed the correct word "an astonishing 90 percent of the time" even though they could not consciously report what word they had seen flashed on the screen.

Step 4. The model, as reported, does not yield any very precise predictions. It suggests, however, that the subjects might be able unconsciously to perceive and remember the masked words that they could not consciously identify. If so, then their guesses about the word should be right more often than not.

Step 5. The prediction and the data clearly agree. The subjects guessed the correct word far more often than not.

Step 6. That the data are described as "astonishing" is a good indication that these results are thought not to be very likely in the absence of something like unconscious perception. The only obvious alternative model is that the subjects were just guessing randomly. But it is very unlikely that random guessing between two alternatives would produce 90 percent correct answers. So the data do support the hypothesis that a model that includes unconscious perception does fit the human mind.

In this example, the model is only vaguely described and it is difficult to see a clear connection between the model and the prediction. Such cases often result in an inconclusive evaluation. What saves this case is that the data are quite dramatic and it is difficult to imagine another plausible model that could explain the data.

A Case of Inconclusive Data

Was That a Greenhouse Effect? It Depends on Your Theory.

The memorably uncomfortable summer of 1988 has left many Americans with a suspicion that nature is at last getting even for mankind's wanton pollution of the atmosphere. From California to the Carolinas, the summer's heat wave and drought took a sobering toll. Electric power faltered, vast forests went up in flames, river navigation was throttled, crops failed.

The greenhouse effect—the trapping of solar heat by pollutant gases in the atmosphere—became a household phrase. Some climatologists warned that unless we quickly mend our ways, the world's grain belts will turn to dust bowls, coastal regions will be flooded, forests will die, and countless species will become permanently extinct. On June 23, Dr. James E. Hansen of the National Aeronautics and Space Administration caught the nation's attention when he told a Senate committee that the warming trend almost certainly stems from the greenhouse effect. A crisis, he warned, may not be long in coming.

But forecasting climate has never been as straightforward as scientists could wish. Many are not even sure that this summer's weather was really symptomatic of any trend at all. According to the Climate Analysis Center of the National Weather Service, July of this year, when the heat wave was at its peak, was only the 11th hottest July in 58 years of record keeping. And while China also suffered a heat wave and a drought, Ireland and parts of Western Europe were unusually chilly and wet.

A. James Wagner, an analyst at the Weather Service, acknowledges that during this decade the world has seen the four warmest years of the past century—1980, 1983, 1986, and 1987. "But I don't feel that the evidence is overpowering that this is anything more than a normal fluctuation," he said.

Climatologists have invented a number of models in an attempt to understand fluctuations in the weather. One such model, which seems to mimic the real climate quite realistically, was devised by Dr. Edward Lorenz of the Massachusetts Institute of Technology. This model, which does not take carbon dioxide into account but does reckon on the interactions of the atmosphere with the ocean, exhibits large variations.

"The Lorenz model was run backward on a computer for the equivalent of about 400 years," Mr. Wagner said, "and the large fluctuations it sometimes produced, which were not entirely random but were not cyclical either, were quite startling." The swings, he said, were as much as plus or minus 3.6 degrees Fahrenheit in global temperature from one year to the next. The model sometimes produced clusters in which several years close together were unusually hot—a pattern imitating the real climate of the 1980s.

Analysis

Step 1. The real world object of study is the climate of the earth.

Step 2. The model is the greenhouse model. According to this model, the earth's atmosphere acts like the windows in a greenhouse, trapping light and heat under the atmosphere. Carbon dioxide increases the efficiency with which the atmosphere traps heat.

Step 3. The data include the drought that covered much of the United States during the summer of 1988. Also included is information about temperatures during other years, both recent and in the more distant past.

Step 4. The prediction is that the earth's temperatures should be increasing. However, the model is not developed in sufficient detail to permit precise predictions of how much the temperature should increase by specified dates, or whether the increases would be uniform around the world.

Step 5. The data and the prediction seem to agree. The summer of 1988 was unusually warm, and for most of this decade, at least in the United States, the weather has been quite warm relative to earlier years.

Step 6. The data, however, seem to be relatively likely to have occurred even in the absence of any greenhouse effect. Some climatologists have even developed an alternative model, the Lorenz model, which predicts relatively large fluctuations in the earth's temperature. These last few years could be one of those fluctuations. The data, therefore, are inconclusive regarding the applicability of the greenhouse model to our current climate. That does not mean, however, that the greenhouse effect is not operative or that it will not show up with more dramatic force in the future.

2.11 CRUCIAL EXPERIMENTS

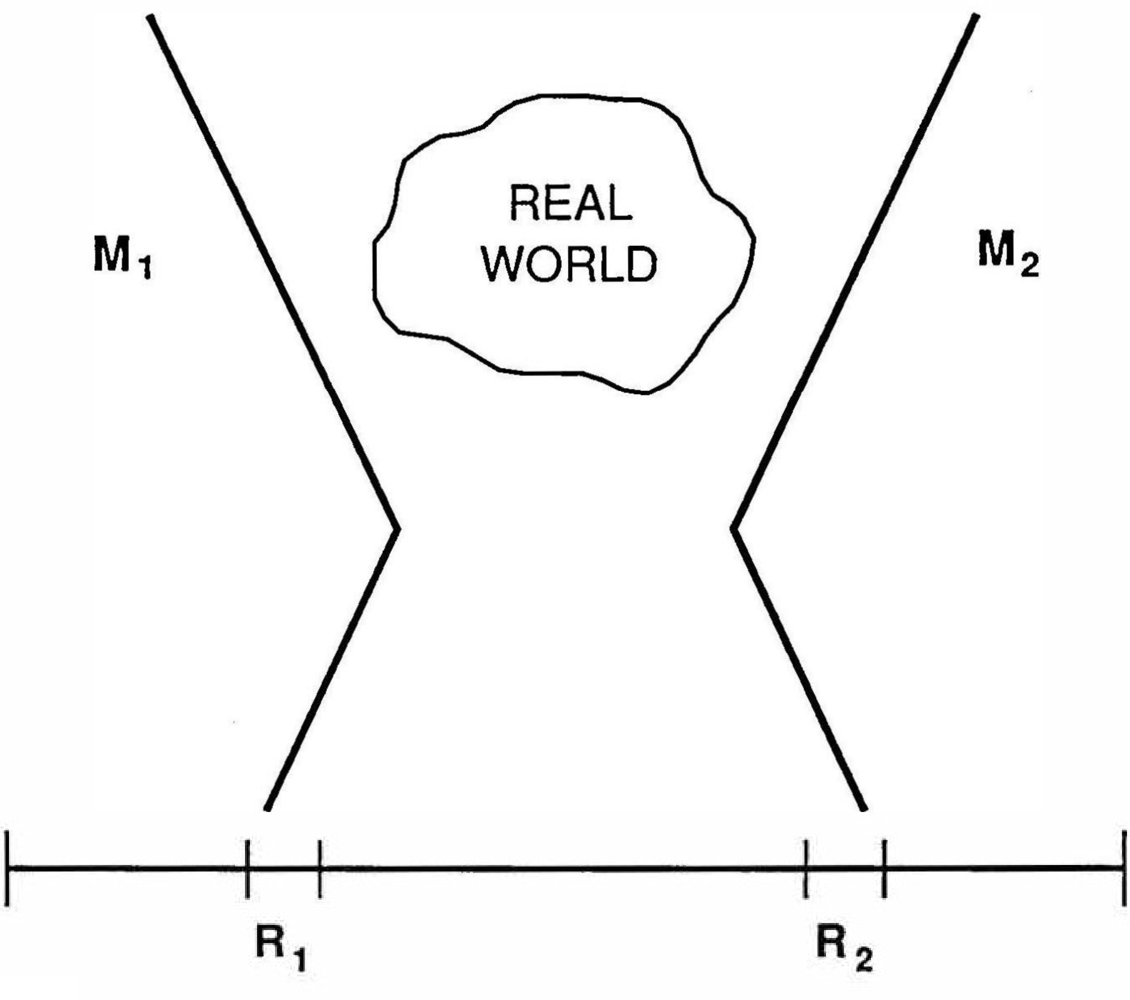
There is one special circumstance in the evaluation of theoretical hypotheses that has fascinated scientists since the seventeenth century. This is an experimental setup that allows one to make a clear choice between two rival models. If the experiment is well designed, one model will come out the winner, and the other the loser. The fate of both is settled in one stroke. Following the terminology of the seventeenth-century scientist, philosopher, and statesman Francis Bacon, experiments of this type are called crucial experiments. So powerful are crucial experiments that scientists, and even historians of science, often reconstruct the past so as to make it seem that a particular experiment was a crucial experiment even though, in fact, no one at the time thought of it that way.

The Structure of Crucial Experiments

Figure 2.13 provides a schematic picture of a crucial experiment. Some bit of the real world is put through the apparatus, which produces a reading on a scale. The reading is the data. There are two rival models of the material under investigation. For the moment, we will simply call these $\mathbf{M_1}$ and $\mathbf{M_2}$, respectively. The experiment is cleverly designed so that if $\mathbf{M_1}$ fits the material in the apparatus, it is very likely that the apparatus would produce a reading in region $\mathbf{R_1}$ of the scale, and very unlikely that it would produce a reading in region $\mathbf{R_2}$ of the scale. Similarly, the design ensures that if $\mathbf{M_2}$ fits the material in the apparatus it

Figure 2.13

A schematic representation of a crucial experiment.



is very likely that the experiment would yield a reading in the region $\mathbf{R_2}$ of the scale, and very unlikely it would produce a reading in region $\mathbf{R_1}$. So $\mathbf{M_1}$ predicts a reading in region $\mathbf{R_1}$, and $\mathbf{M_2}$ predicts a reading in region $\mathbf{R_2}$.

The beauty of this design comes across when one considers that the result of the experiment is to be used to decide which model best fits the world. The strategy is that if the experiment produces a reading in region $\mathbf{R_1}$, then one chooses $\mathbf{M_1}$ as the best-fitting model. Likewise, if the reading is in region $\mathbf{R_2}$, one chooses $\mathbf{M_2}$ as the best-fitting model.

To see why this is an effective strategy, consider the matrix shown in Figure 2.14. Because there are two models to choose between, and two different choices that can be made, the whole experiment has four, not just two, possible final results. Two of these results are *correct* choices. These are choosing $\mathbf{M_1}$ as the best-fitting model when, in fact, $\mathbf{M_1}$ is the best-fitting model, or choosing $\mathbf{M_2}$ as the best-fitting model when, in fact, $\mathbf{M_2}$ is the best fitting model. The other two possible results are clearly *incorrect.* These are choosing $\mathbf{M_1}$ as the best-fitting model when, in fact, $\mathbf{M_2}$ fits best, or choosing $\mathbf{M_2}$ as the best-fitting model when, in fact, $\mathbf{M_1}$ fits best.

The experiment is designed with the presumption that one of the models provides an adequate fit and the other does not. So, initially, there are two possibilities. We can consider each in turn. Suppose, first, that $\mathbf{M_1}$, in fact, provides the better fit. The design ensures that the likely result of the experiment will be a reading in region $\mathbf{R_1}$. In this case, one correctly chooses $\mathbf{M_1}$ as the best-fitting model. On the other hand, suppose that $\mathbf{M_2}$, in fact, provides the better fit. The design ensures that the likely result of the experiment will be a reading in region $\mathbf{R_2}$. In this case, one correctly chooses $\mathbf{M_2}$ as the best-fitting model. Either way, one is very likely to make a correct choice and very unlikely to make an incorrect choice. The only open question is which of the two possible correct choices one will make. That depends on the data.

In actual practice, experiments that are designed to be crucial experiments often do not work out. The data obtained turn out to be readings in

Figure 2.14

A matrix representing the choice between two theoretical models in a crucial experiment.

	$\mathbf{M_1}$ BEST FITS THE REAL WORLD	$\mathrm{M_2}$ BEST FITS THE REAL WORLD
CHOOSE $\mathbf{M_1}$ AS BEST-FITTING MODEL	**CORRECT**	**INCORRECT**
CHOOSE $\mathrm{M_2}$ AS BEST-FITTING MODEL	**INCORRECT**	**CORRECT**

neither of the regions predicted by the two models. Unless there was some mistake in the execution of the experiment or in the derivation of the predicted readings, one can only conclude that neither of the two models adequately represents the real world material. The initial presumption that one of the two models fits has to be rejected.

Even though definitive crucial experiments are relatively rare, the possibility of devising such experiments provides a clear and simple, idealized model of scientific reasoning at its best. Understanding this model of scientific reasoning can help one evaluate the less than ideal cases one meets every day.

An Analysis of a Crucial Experiment

You should attempt to evaluate the following report before reading the sample analysis that follows.

Mutations

It has long been known that changes in the genes of organisms can occur. Such changes are commonly called "mutations." In the 1940s, it was not yet known how mutations occur. Part of the answer to this question was given by a famous experiment performed by Max Delbrück and Salvador Luria.

It had recently been learned that some types of viruses (bacteriophages, or phages) can attack and kill some types of bacteria. It is relatively easy to grow bacteria in covered dishes containing nourishment in which bacteria generally thrive. These are called "bacteria cultures."

Delbrück and Luria discovered that in some bacteria cultures a few of the bacteria survive attacks by phage viruses. Moreover, descendants of the surviving bacteria tend also to survive phage attacks. This shows that the genes of some of the bacteria had undergone mutations that made them resistant to the phage virus, and that these resistant bacteria passed their mutant genes on to their offspring.

The question remained as to whether the mutations that made the bacteria resistant were caused by the attacking virus itself, or whether they merely happened by chance. The experiment at issue was designed to answer this question regarding the cause of the mutations.

Delbrück and Luria considered what would happen if a number (say twenty) of bacteria cultures, each with a similar small number of bacteria, were allowed to grow for a short time, then all were injected with the same quantity of phage virus, and then were allowed to grow some more. If the phages were producing the mutations, they argued, then all the bacteria cultures should end up with roughly the same number of resistant bacteria.

On the other hand, if the mutations were arising by chance, it follows that those bacteria cultures in which the chance mutation happened to occur early in the experiment would end up with many more mutant bacteria than those cultures in which the mutation happened to occur late in the experiment. The earlier mutant bacteria would have a longer time to multiply. Those cultures in which the chance mutation happened to occur at some intermediate time would end up with an intermediate number of mutant bacteria. If it is a matter of pure chance when the mutation occurs, one would, therefore, expect that by the end of the experiment there would be a large variation in the numbers of mutant bacteria in the different bacteria cultures.

Delbrück and Luria prepared a number of bacteria cultures, then introduced the phage virus, and later found that the actual number of resistant bacteria differed widely from one bacteria culture to the next.

Analysis

One way to evaluate such reports would be to run through our standard program twice, once for each model. It is quicker, and more enlightening, however, to do just one analysis. But in this case many of the steps will have two answers. Also, the way this particular report is written makes it natural to consider the predictions before the data, thus putting Step 4 ahead of Step 3.

Step 1. The real world object of study is the process by which mutations arise in bacteria.

Step 2. There are two models: (1) according to the causal model, the mutations are caused by the action of the phage virus on the genes of the bacteria, and (2) according to the chance model, the mutations in the genes of the bacteria occur by chance, independent of any action by the phage viruses.

Step 4. There are two predictions, one for each model: (1) the causal hypothesis yields the prediction that the number of resistant bacteria will be roughly the same in all of the bacteria cultures, and (2) the chance hypothesis yields the prediction that there will be a large variation in the numbers of mutant bacteria in the different bacteria cultures.

Step 3. They found that the actual number of resistant bacteria differed widely from one bacteria culture to the next.

Step 5. The data (1) agree with the prediction derived from the chance hypothesis, and (2) disagree with the prediction derived from the causal hypothesis. The data, therefore, provide good evidence that the causal hypothesis is mistaken.

Step 6. The only rival to the chance hypothesis mentioned in the report is the causal hypothesis that yields a different prediction from the chance hypothesis. It is difficult even to imagine another hypothesis. So it does not appear likely that there would have been a large variation in the number of resistant bacteria if the chance hypothesis were false. The data, therefore, provide good evidence that the chance hypothesis is correct.

EXERCISES

Analyze these reports following the six-point program for evaluating theoretical hypotheses developed in the text. Number and label your steps. Be as clear and concise as you can, keeping in mind that you must say enough to demonstrate that you do know what you are talking about. A simple "yes" or "no" is never a sufficient answer. Many of these reports are taken directly from recent magazine or newspaper articles and are presented here with only minor editing.

Exercise 2.1

Einstein's Impossible Ring: Found

A phenomenon first predicted by Albert Einstein in 1936, and then dismissed by him as something that would be hopeless to look for, has now been found. Astronomers conducting a survey of radio sources at the Very Large Array (VLA) radio telescope near Socorro, New Mexico, have discovered an object in the constellation Leo that has been imaged into a complete ring by the gravitational-lensing effect.

"Of course, we'd all heard of Einstein rings," says team member Jacqueline Hewitt of the Haystack Observatory in Massachusetts. "But when I saw it come up on the computer screen, I thought at first it was a problem with the [VLA's image analysis] software."

It was not. Yet her skepticism was understandable. According to Einstein's general theory of relativity, gravitational lensing would happen when light or radio waves from a distant galaxy or quasar pass by a massive foreground object on the way to Earth. The object's gravity would deflect the radiation and thus produce one or more distorted images of the source. A number of such images have actually been found during the past decade. As Einstein himself pointed out, however, the image can only form a complete ring if a source and the lensing object are precisely lined up with the earth—which seems absurdly improbable.

Except that there it was on Hewitt's computer screen—radio source MG1131 + 0456, a tiny oval about 2 arc seconds across with elongated bright spots at either end. In subsequent observations, Hewitt and her colleagues were able to rule out the possibility of its being a supernova remnant, or any other such ring-like structure. Moreover, using a regular optical telescope, they obtained optical images of a candidate for the imaging mass: a 22nd magnitude object whose shape and other characteristics are those of a large elliptical galaxy.

Exercise 2.2

Why is the World Full of Large Females?

Animals come in a vast range of sizes, from the tiniest zooplankton to the largest whale. Absolute body size has a crucial influence on a species' life history, affecting such factors as metabolic rate, longevity, and territorial range. And, within a species, relative body size—females compared with males—is important in behavioral ecology terms too. In most species in the world, females are larger than males, although this rule applies more to groups, such as insects, fishes, amphibians, and reptiles, than it does to mammals and birds. Nevertheless, the largest animal that has ever lived is a female: the female blue whale.

Why females should attain a larger body size than males has long fascinated biologists. Darwin had an explanation for it, namely: "Increased size must be in some manner of more importance to the females . . . and this, perhaps, is to allow the production of a vast number of ova." This so-called, fecundity-advantage model "has achieved the status of conventional wisdom," says Richard Shine of the University of Sydney, Australia. The model appeals through its simplicity and its consistency with many empirical observations. However, it had not been formally tested, says Shine, a deficiency he has recently repaired. He finds that even though the model may apply in some species it is by no means universal.

It is no easy task, of course, to solve the question of why one sex may be bigger than the other in a particular species, not least because there are two partners in the game. Specifically, the female might be the bigger of the sexes because of the kind of selective advantage that Darwin proposed; but it is equally true that if males evolve small body size for some different, adaptive reason, then the same pattern of body size dimorphism would apply. Several biological factors are likely to be operating in any particular case, and this should always be borne in mind when looking for the *factor.*

Shine elected to test the model in something of a roundabout fashion, thereby hoping to avoid confounding variables that might affect body size in different directions. He measured male-female body size differences in a series of lizard species, some of which produce variable clutch sizes while in others the clutch size is constant. "If the main selective pressure for large female size is an associated increase in fecundity," says Shine, "the species with invariant clutch sizes would have no such advantage and females should tend to be smaller (relative to males)."

It turns out that in anoline iguanids, which produce a single egg, the proportion of species in which the female is larger than the male is about the same as in other iguanids in which clutch size is variable. "The same tends to be true for other lizards with invariant clutch sizes," says Shine. "These data, involving at least seven separate phylogenetic lineages of lizards, appear to falsify the main prediction of the fecundity-advantage model."

Exercise 2.3

Memory Transfer

In the 1950s, biologists and psychologists began speculating that memory works by changing and storing certain chemicals in the brain. That is, they began developing chemical models of memory. In the 1960s, it occurred to some psychologists at the University of Michigan that, if memory is a matter of stored chemicals, one might be able to transfer memory from one organism to another simply by transferring chemicals from the brain of one to the brain of another. Experiments were done on both worms and rats. In the rat experiments, a group of rats was taught to get a small drink of milk by pressing a lever on the opposite side of their cage from where the milk was dispensed. This is difficult for rats to learn. The average time it took a typical group of rats to learn it was about 25 hours. Then, chemicals from the brains of the trained rats were extracted and injected into the brains of other similar rats. Still other rats were injected with material from the brains of untrained rats. The injected rats, they reasoned, should learn faster if they received injections from the previously trained rats. The rats that received injections from trained rats learned the trick in an average of only 3 hours. The others averaged around the usual 25 hours.

Exercise 2.4

Green Antarctica

Most of us think of Antarctica as a Dantean wasteland, locked—past, present, and future—in the frozen, white grip of perpetual polar winter. Until recently, geologists have held much the same view. But new findings indicate that Antarctica has gone through periods during which it was warm enough to actually turn green.

Camped out for six weeks in the highlands of the Transantarctic mountain range, a team led by Peter Webb of Ohio State University found 2-million- to 3-million-year-old fossils of trees, plants, pollens, and spores.

The evidence has led Webb to paint a new, revised picture of an Antarctica that went through periodic cycles of dramatic warming. During those epochs, he says, Antarctica may have looked much like the hardy, low-lying forests currently found in southern Chile, New Zealand, and Tasmania, where dwarf shrubs and small beeches—"They look a little like bonsai trees," he says—cling to rocks and point perpetually in the direction of the strong prevailing winds. Interestingly, Webb has recently seen reports indicating that some of the same fossils have been found in the Arctic. Periodic warming, he concludes, "may be a bipolar phenomenon. It's impressive that similar species developed at both ends of the earth."

Exercise 2.5

Hot Extinction Theories

Many scientists now agree with the notion that an asteroid 10 to 20 kilometers across slammed into Earth 65 million years ago, at a time termed the Cretaceous-Tertiary (K-T) boundary. But the debate rages on about what toll it took on our planet's life forms, and even on whether the result of the impact produced a warming or cooling of the earth's atmosphere.

At a meeting of planetary scientists in Houston, Texas, John D. O'Keefe and Thomas J. Ahrens of the California Institute of Technology put forth the idea that the earth became warmer after the cataclysm. They reason that an impact occurring in shallow seas, or on dry land that had once been under water, would vaporize thick beds of carbonate sediments and thus inject huge quantities of carbon dioxide (CO_2) into the atmosphere. (By current estimates, such sediments have locked up an amount of CO_2 equivalent to between fifty and eighty times the total mass in Earth's present atmosphere.) The CO_2 would then block the escape of Earth's infrared radiation to space, creating a greenhouse effect that would persist for millennia. Calculations have shown that a global warming of only 10 degrees Kelvin would have been sufficient to trigger the K-T extinctions.

To support their model, the Caltech researchers used powerful guns to propel projectiles into calcite ($CaCo_3$) targets at 4 to 6 kilometers per second. Then they measured how much CO_2 was released as a function of impact velocity and extrapolated the results to a global scale. For example, an object 20 kilometers across, striking a 1-kilometer-thick calcite bed at 20 kilometers per second, should double the amount of CO_2 now in the atmosphere.

Admittedly, the O'Keefe-Ahrens concept requires not only that the impactor strike carbonate beds, but also that the beds be *thick enough to release an adequate mass of* CO_2*. Some scientists in Houston questioned the likelihood that all these conditions could be met. Others emphasized that more research is needed to determine the total response of the atmosphere to a major impact.*

Exercise 2.6

A Heresy in Evolutionary Biology

As anyone with even a passing knowledge of evolutionary biology knows, natural selection is a twofold process: the generation of genetic mutation followed by the fixation of variants that are favored by prevailing conditions. In the world of evolutionary biology, one thing has seemed certain: the generation of genetic mutations is a continuous and random process, uninfluenced by external circumstances. However, if John Cairns, Julie Overbaugh, and Stephen Miller of the Harvard School of Public Health are correct in their interpretation of certain experiments with the bacterium Escherichia coli, *that certainty may be on shaky ground.*

In a paper published in Nature, *Cairns and his colleagues aim "to show how insecure is our belief in the spontaneity (randomness) of most mutations." The Harvard researchers describe the results of a handful of experiments which, they suggest, demonstrate that "bacteria can choose which mutations they should produce." Anything more heretical can hardly be imagined. They do add, however, that "this is too important an issue to be settled by three or four rather ambiguous experiments."*

One of the experiments involves taking colonies of E. coli *that are incapable of metabolizing lactose and exposing them to the sugar. If the lactose-utilizing mutants simply arise spontaneously in the population and are then favored by prevailing conditions, then this would lead to one pattern of new colony growth. A distinctly different pattern is produced if, under the new conditions, the rate of production of lactose-utilizing mutants is enhanced. The observation is something of a mixture of patterns, indicating that directed mutation appears to be occurring. "This experiment suggests that populations of bacteria . . . have some way of producing (or selectively retaining) only the most appropriate mutations," note Cairns and his colleagues. They cite two other types of experiment that can also be interpreted in this way.*

Because the randomness of mutation has been so fundamental to evolutionary biology since the 1940s, few researchers have cared to test the notion directly. There are, therefore, no data beyond those from this handful of experiments that might *indicate how general a phenomenon directed mutation might be. Nevertheless, Cairns suspects that it might well turn out to be rather widespread, at least in bacteria. Kent Holsinger, a theoretical population geneticist at the University of Connecticut, says that "if it is general and not just confined to* E. coli *and other bacteria, it could have major implications for evolutionary biology. At the very least, he notes, "there is something going on here that we haven't considered."*

Exercise 2.7

Quartz Discovery Supports Theory That Meteor Caused Dinosaur Extinction

Researchers say they have strong new evidence that the age of dinosaurs ended 65 million years ago when a giant meteorite, or comet, slammed into the earth with the energy of a billion atomic bombs. Scientists with the U.S. Geological Survey office in Denver said that microscopic particles of quartz found in Europe, New Zealand, the Pacific Basin, and elsewhere contain structural cracks associated with the impact of a large body hitting the earth. The mineral debris indicates that a single catastrophic event, and not a series of volcanic eruptions suggested by other scientists, ended the 150-million-year reign of the great lizards, they said. The microscopic fracturing found in the quartz is more like that associated with the pressures of a massive impact than what would have resulted from volcanic activity, they say in a study published in the journal Science.

Bruce Bohor, Peter Modreski, and Eugene Foord said the "shocked quartz" is found in the same sediment layers that contain unusually high levels of iridium, a metal common in asteroids, meteors, and comets. The researchers said the latest findings bolster the controversial, 10-year-old theory of Nobel Prize-winning physicist Luis Alvarez and his geologist son, Walter, that a single catastrophic event led to a great extinction of life on Earth. Geological evidence appears to show a massive extinction of dinosaurs beginning about 65 million years ago, but scientists have been unsure of the reason why.

The Alvarez theory says the impact of the extraterrestrial body released energy equivalent to that of 6 billion Hiroshima atomic bombs and threw up a giant cloud of debris that encircled the globe and diminished sunlight for months, if not years. Climate cooling, resulting from the dust blocking sunlight, resulted in the death of dinosaurs and many other types of animal and plant life, according to the theory.

The theory is based upon finding up to 600 times normal levels of iridium in clay deposits from the period, and Bohor said, in *a telephone interview, that the same iridium concentrations have been found at every site of the telltale quartz particles. The even distribution of the shock quartz, and certáin minerals combined with it, point to a big comet or meteorite striking a continental area in the Northern Hemisphere, he said. A body six miles wide hitting the earth at 45,000 miles per hour, as calculated by the Alvarez theory, could have blasted debris high enough into the atmosphere to account for the worldwide shock quartz distribution, Bohor said.*

High pressure associated with volcanic activity can fracture quartz crystals, and proponents of the volcanic theory of dinosaur extinction say this is the source of the shock crystals found in sediments from the period. However, Bohor and his colleagues said quartz fractures caused by impacts are distinct from those resulting from other pressures. "When a meteorite strikes the earth, the mass and speed of impact cause a shock," Bohor said. "This shock wave bounces around in different directions as it hits other objects and comes back into the crystal to produce multiple sets of fracture features unique in impact shocks."

The report said multiple sets of fractures were evident in quartz particles found at several sites: Stevns Klint and Nye Klov in Denmark; Petriccio and Pontedazzo in Italy; Caravaca in Spain; Woodside Creek in New Zealand; and a core taken from the north central Pacific Ocean basin. The samples matched those previously found in Montana and Wyoming, Bohor said, and Soviet scientists recently found the same type of quartz on the east side of the Caspian Sea in the Soviet republic of Turkmen.

Exercise 2.8

The Expanding Universe

One of the most interesting discoveries of the twentieth century is that the universe is expanding; that is, the galaxies are all moving away from each other. This discovery stimulated the creation of numerous models of systems in which such expansion would take place. Two of these models were widely regarded as possibly representing the structure of the real universe. One was an explosion model (the Big Bang theory) in which all matter is originally concentrated in one place and explodes outward. The other is a steady-state model, in which subatomic bits of matter are created out of nothing and eventually move outward, leaving each region of space with the same total amount of matter for all time.

If the universe is an exploding system, it follows that the density of matter (the number of galaxies per cubic light year) gets *less and less the farther away from the original explosion one gets. If the universe is a steady-state system, on the other hand, the density of matter should be exactly the same everywhere. This remains true wherever in the universe one happens to be. Whichever direction one looks, the density of the most distant galaxies should be less if the explosion model is correct.*

To decide which of these two models best fits the real universe, what we need to do is measure the density of matter in the most distant regions of space. In recent years, radio telescopes have made it possible to make such measurements. These measurements show a clear decrease in the density of the most distant observable galaxies.

Exercise 2.9

New Observations Reveal Cosmic Mystery

In the 1920s, it was discovered that the stars exist in large spiral- or platter-shaped clusters we now call galaxies. Each galaxy contains millions of stars with most near the center and fewer out toward the edges. Our sun and solar system are now thought to be roughly one-third the way in from the edge of the galaxy we call the Milky Way.

Since the discovery of galaxies, astronomers and astrophysicists have naturally wondered how they work, and in particular, what keeps them together. The standard idea has been that a galaxy is held together by the force of gravity. If this is right, then it is possible to calculate the motions of various stars within the galaxy. Unfortunately, it has not until now been possible to measure the motions of stars with sufficient accuracy to determine whether the calculated motions are correct.

Recently, however, computers have been used to sharpen the images of stars produced by large telescopes. This new technique has revealed that the stars on the outer edge of the Milky Way are moving much faster than they should according to the standard calculations.

The result has scientists baffled. They have not yet been able to come up with an alternative model that might explain this surprising result.

Exercise 2.10

Scientists Put a New Twist on Creation of the Universe

A Great Wall of galaxies stretching hundreds of millions of light years across the known universe has been discovered by *two Harvard astronomers and threatens long-held, fundamental theories of how the universe came into existence.*

The wall, some 500 million light years long, 200 million light years wide, and 15 million light years thick, consists of more than 15,000 galaxies, each with billions of stars. Described as the "largest single coherent structure seen, so far, in nature," the image of the wall emerged after about a decade of effort to map the structure of the universe in three dimensions. Correspondingly, there are large areas of relative void containing comparatively few galaxies.

The astronomers' findings were published in the journal Science. *While the area surveyed is huge, it is only a small piece of the known universe. The astronomers involved said the survey area is to the universe what Rhode Island is to the earth, a small percent. The survey was done by mapping the "red shifts," or motions, of a large group of galaxies some 200 to 300 light years from Earth.*

The survey, using the 60-inch telescope at the Whipple Observatory in Arizona, was conducted by dividing an area of the universe into thin slices, then noting the position of galaxies within the slices. As more of the slices were completed and layered together, a three-dimensional portrait of the area emerged and slowly revealed the massive wall of galaxies.

If this map of the universe proves to be a reality, it means the long-held theories of a universe evolving from a smooth, superhot plasma following a Big Bang some 15 billion years ago might be wrong. The concentration of galaxies around voids filled with either mysterious dark matter, or nothing, points to a very uneven, lumpy universe that apparently could not evolve out of the smooth beginnings described by current theories.

What, if anything, is in the voids remains unknown and could be a key to understanding the universe. Many scientists believe that 90 to 99 percent of the matter in the universe is dark and has yet to be detected. Whether the huge voids, some as large as 150 million light years in diameter, contain dark matter is not known.

But given their vast size, said Margaret Geller, one of the two Harvard astronomers to make the discovery, "it may make more physical sense to regard the voids as the fundamental, large-scale structures of the universe."

The fundamental theory most threatened by the discovery is known as the standard model. It holds that the universe originated in a Big Bang and has expanded in a fairly uniform manner for the past 15 billion years. It calls for a homogeneous universe that is very much the same in every direction. But the new findings raise questions about how such an uneven system of voids and galaxies could have developed out of a uniform, homogeneous beginning in 15 billion years.

"In my opinion, something is fundamentally wrong in how (scientists believe) the structure of the universe formed," Geller said. "Maybe the universe isn't homogeneous." She said the work is important philosophically because "mapping the universe is about finding its origins. If we can't understand where the universe came from, then we can't understand where we came from."

In their paper detailing the discovery of the Great Wall, Geller and her colleague, John Huchra, hinted that the wall may be much larger than they have measured. The "most sobering" result of their discovery, they said, is that the size of the largest structures they have found "is limited only by the extent of the survey." If the structures and the voids are larger still, then scientists may have to abandon entirely the fundamental idea of a homogeneous universe.

"There is no theory that even comes close to explaining what we are seeing," said Michael Kurtz, also an astronomer at Harvard.

Exercise 2.11

Gravity Waves

Newspapers around the world recently carried reports of a striking new experimental test of Einstein's General Theory of Relativity. According to Einstein's theory, a large rotating mass should give off gravity waves that travel at the speed of light. Several scientists have previously claimed to have detected gravity waves, but none of these claims were very well substantiated. The new experiment was made possible by the discovery of a unique pulsar—a very dense star that emits powerful radio waves in distinctive pulses. This particular pulsar appears to be rotating in orbit around another large but invisible mass (perhaps a black hole). The discovery of this unique pulsar was made using a radio telescope 1000 feet in diameter—the world's largest.

If the observed system indeed fits an Einsteinian model, it should give off gravity waves. Moreover, according to the model, because gravity waves carry off energy from the system, the rate of rotation of the pulsar should be slowing down at the rate of about one ten-thousandth of a second a year. Using the large radio telescope and clocks accurate to fifty-millionths of a second a year, the period of rotation of the pulsar was measured over a period of four years. The most recent measurements show that the pulsar has indeed been slowing down by just about one ten-thousandth of a second a year—as predicted. Even scientists who had previously doubted the existence of gravity waves had *to admit that finding the system to have slowed down by the predicted amount was quite remarkable.*

Exercise 2.12

Comet Source: Close to Neptune

Comets have been considered outsiders, visitors from an enveloping cloud of inactive comets on the far fringes of the solar system. But new computer simulations of how comets can be drawn into the inner solar system eliminate a far-distant, spherical cloud of comets as the source for a major class of comets. The only practicable source in these simulations for the comets that now follow small, quick orbits near the planets, is a flat disk of comets lying just outside the orbit of Neptune. That is a distance of little more than thirty times the distance between the sun and Earth (30 astronomical units) compared to the tens of thousands of astronomical units to the distant comet cloud.

Martin Duncan of Lick Observatory, and Thomas Quinn and Scott Tremaine of the University of Toronto, ran their computer model of the solar system to see how some of the peculiar characteristics of the short-period comets, those having orbital periods of 200 years or less, might have originated. Could the cloud of comets, called the Oort cloud, supply objects that behaved like these? It clearly is the source of the 589 known, long-period comets, the ones rarely seen from Earth as they loop by the sun on circuits that require 200 to millions of years to complete. Everything about their motion is consistent with having been jostled out of the Oort cloud by a passing star and into an orbit that passes near the sun.

But the short-period comets do not behave like typical Oort cloud comets gone astray. Most strikingly, the orbits of the short-period comets tend to lie within about 30 degrees of the plane of Earth's orbit, called the ecliptic. And only four of the 121 known, short-period comets have their orbits tilted more than 90 degrees so that they are orbiting in the direction opposite to that of Earth. In contrast, half the comets in the Oort cloud must have such retrograde orbits.

Duncan and his colleagues first checked to see if, as reported fifteen years ago, the four giant planets could select, from comets falling in from the Oort cloud, only those having low-inclination orbits typical of short-period comets. Using a mathematical model that included the sun, the four giant planets, and 5000 comets falling into the vicinity of massive Jupiter from the Oort cloud they ran a simulation on their own souped-up, Sun-3 microcomputer for a total of several months of computer time. They *speeded up the calculation of millions of years of orbital evolution by increasing the mass of the giant planets by a factor of up to forty in the simulation.*

To their surprise, these modelers found that the planets are not selective at all when they deflect comets into new, smaller orbits. The comets from the Oort cloud that achieved periods of less than 200 years formed a cloud of their own, with no preference for orbiting in a disk near the ecliptic. In addition, three-quarters of them had periods greater than fifteen years; only twenty-one of the 121 observed, short-period comets have periods greater than fifteen years.

Ruling out the Oort cloud, the modelers next tried a belt of low-inclination comets near Neptune's orbit. The idea dates back to a suggestion by Gerard Kuiper in 1951 that it would be only natural to find some debris from the formation of the solar system beyond Neptune. Comet-sized objects would not have formed planets there, but neither the giant planets nor passing stars could easily dislodge them, or even smear their disk into a cloud. In fact, an absence of comets there would imply an oddly abrupt, outer edge to the original solar system disk.

When the simulation was run with a disk of comets orbiting between a distance of 50 and 20 to 30 astronomical units, or well inside the orbit of Neptune, the comparison with reality was impressive. The mean oribital distances of both simulated and observed, short-period comets cluster around 3 astronomical units with a lesser tendency to be near 5 astronomical units, the orbital distance of gravitationally influential Jupiter. That is also where the preponderance of maximum orbital distances lies for both sets of comets. Comets in each set also have a tendency to be passing the ecliptic when they are closest to the sun.

Most crucially, about half of the simulated comets, which started out with inclinations of 0 degrees to 18 degrees, retained inclinations of less than 30 degrees. More than 80 percent of observed short-period comets are confined to such a disk. About 8 percent of the simulated comets had their orbits tilted as much as to be in retrograde motion, compared with 3 percent of observed comets and 50 percent of Oort-cloud comets. As is the case with observed comets, simulated, retrograde comets tended to have periods longer than fifteen years.

"You can't get good agreement," says Tremaine, "if you start off with inclinations far out of the ecliptic. It works if, and only if, the source is in the plane of the ecliptic." That moves the hypothesis of a comet belt lying beyond Neptune from being "very plausible to being the only plausible hypothesis." The Oort cloud as a significant source for short-period comets "is now ruled out," he says.

Exercise 2.13

Project

Find a report of the results of an experiment that is relevant to a theoretical hypothesis. You may find an example in a newspaper or newsmagazine. The Sunday supplement to your local newspaper is a good bet. The weekly science section of The New York Times *is particularly good. You might also try some popular sources that specialize in scientific findings, such as* Scientific American, Psychology Today, *or* Science. *When you have found something that you find interesting and substantial enough to work on, analyze the experiment following the standard program for theoretical hypotheses.*

This exercise may be turned into a longer project by looking for other sources of information on the same theory. You might be able to uncover a whole history of experiments relating to the theory, each of which can be analyzed. You may discover other experiments bearing on hypotheses using similar theoretical models. Alternately, you may discover cases in which such hypotheses were refuted. Perhaps more elaborate models of the same type were then developed to replace the discarded models. You can then look to see whether later evidence supported these new hypotheses, and so on.

One of the side effects of this project is that you may get an idea of the different levels of science reporting in various popular sources. Some sources tell you everything you need to know in order to evaluate the reported hypotheses. Others give so little information that you cannot tell whether the evidence supports the hypotheses or not. You are forced to take their word for it. Most sources fall somewhere in between.