On Being a Scientist

Committee on the Conduct of Science
National Academy of Sciences

NATIONAL ACADEMY PRESS
Washington, D.C. 1989

The nature of scientific research

Is There a Scientific Method?

Throughout the history of science, some philosophers and scientists have sought to describe a single systematic method that can be used to generate scientific knowledge. For instance, one school of thought, dating back at least to Francis Bacon in the seventeenth century, points to observations as the fundamental source of scientific knowledge. According to this view, scientists must cleanse their minds of preconceptions, sitting down before nature "as a little child," as the nineteenth-century biologist Thomas H. Huxley described it. By gathering facts without prejudice, a scientist will eventually arrive at the correct theory.

Some scientists may believe in such a picture of themselves and their work, but carrying this approach into practice is impossible. Nature is too amorphous and diverse for human beings to observe without having some ideas about what they are observing. Scientific understanding is made possible through the interplay of mental constructs and sensory impressions. Scientists may be able to suspend some prior theoretical or thematic preconceptions to view nature from a new perspective, but they cannot view the physical world without any perspective.

Other formulations of the "scientific method" have been proposed over the years, but many scientists regard such blanket descriptions of what they do with suspicion. Perhaps from a distance science can be organized into a coherent framework, but in practice research is as varied as the approaches of individual researchers. Some scientists postulate many hypotheses and systematically set about trying to weed out the weaker ones. Others describe their work as asking questions of nature: "What would happen if . . . ? Why is it that . . . ?" Some researchers gather a great deal of data with only vague ideas about the problem they might be trying to solve. Others develop a specific hypothesis or conjecture that they then try to verify or refute with carefully structured observations.

Rather than following a single scientific method, scientists use a body of methods particular to their work. Some of these methods are permanent features of the scientific community; others evolve over time or vary from discipline to discipline. In a broad sense, these methods include all of the techniques and principles that scientists apply in their work and in their dealings with other scientists. Thus, they encompass not only the information scientists possess about the empirical world but the knowledge scientists have about how to acquire such information.

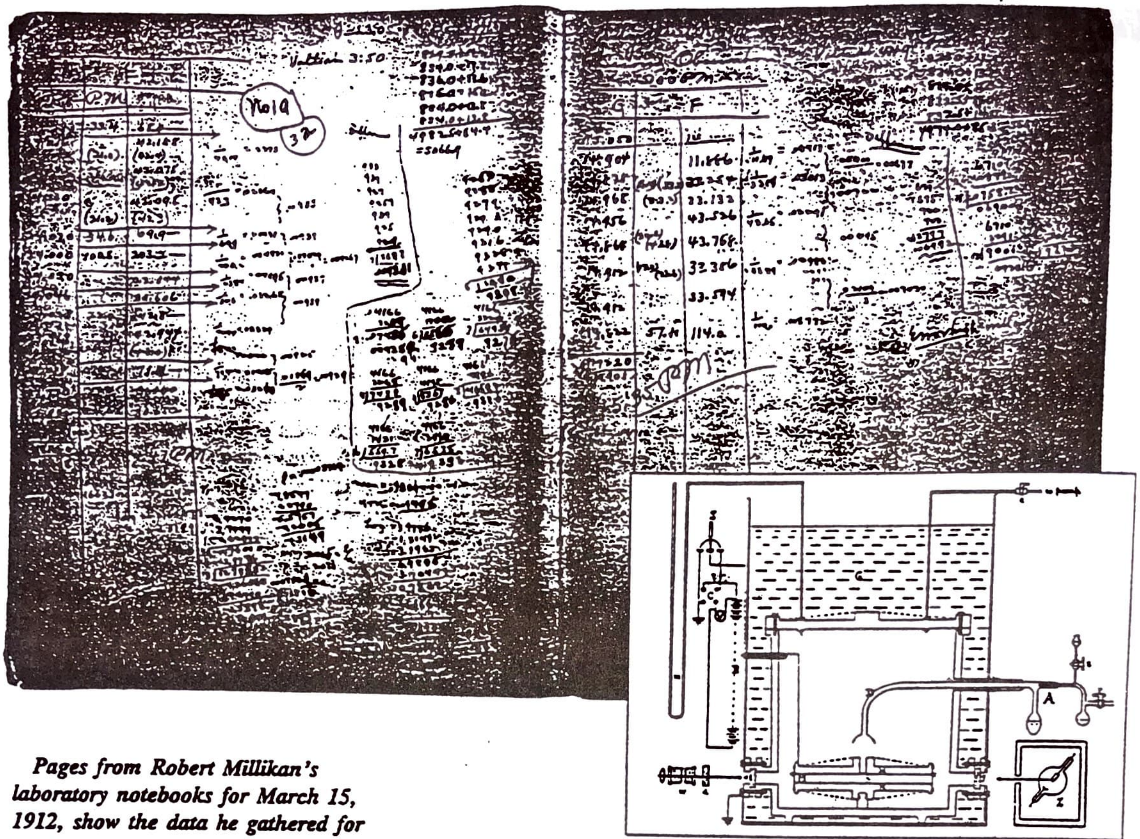
"Scientists are people of very dissimilar temperaments doing different things in very different ways. Among scientists are collectors, classifiers and compulsive tidiers-up; many are detectives by temperament and many are explorers; some are artists and others artisans. There are poet-scientists and philosopher-scientists and even a few mystics."

—Peter B. Medawar, *The Art of the Soluble*, London: Methuen, 1967, p. 132

The Treatment of Data

One goal of methods is to coax the facts, untainted by human bias, from a scientific investigation. In retrospect, this may seem a straightforward process, a simple application of accepted scientific practices to a specific problem. But at the forefronts of research, neither the problem nor the methods used to solve it are usually well-defined. Instead, experimental techniques are pushed to the limit, the signal is difficult to separate from the noise, and unknown sources of error abound. In such an uncertain and fluid situation, picking out reliable data points from a mass of confusing and sometimes contradictory observations can be extremely difficult.

One well-known example of this difficulty involves the physicist Robert Millikan, who won the Nobel Prize in 1923 for his work on the charge of the electron. In the 1910s, just as most physicists were coming to accept the



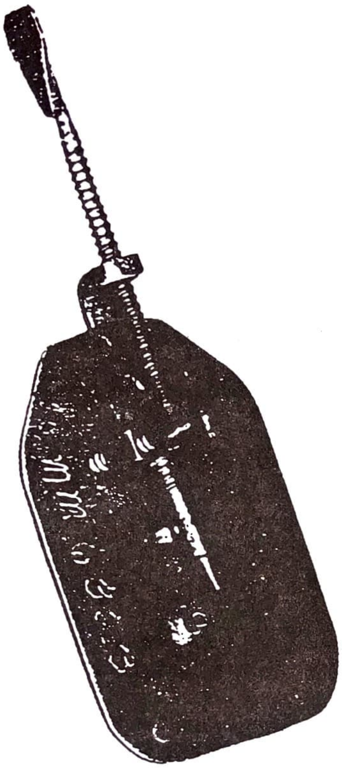
Pages from Robert Millikan's laboratory notebooks for March 15, 1912, show the data he gathered for two oil drops, along with his rough calculations for the uncorrected charge of the electron.

Below: Diagram from a 1913 paper showing Millikan's experimental apparatus. The oil drops were generated by an atomizer (A), and their motions were observed between the horizontal plates of a charged condenser (M and N).

existence of the electron, Millikan carried on a protracted and sometimes heated dispute with the Viennese physicist Felix Ehrenhaft over the magnitude of the smallest electrical charge found in nature. Both men based their findings on the movements of tiny charged objects—oil drops, in Millikan's case—in electric fields. Ehrenhaft used all the observations he made without much discrimination and eventually concluded that there was no lower limit to the size of an electrical charge that could exist in nature. Millikan used only what he regarded as his "best" data sets to establish the magnitude of the charge and argue against the existence of Ehrenhaft's "subelectrons." In other words, Millikan applied methods of data selection to his observations that enabled him to demonstrate the unitary charge of the electron.

Millikan has been criticized for not disclosing which data he omitted or why he omitted those data. But an examination of his notebooks reveals that Millikan felt he knew just how far he could trust his raw data. He often jotted down in his notebooks what he thought were good reasons for excluding data. However, he glossed over these exclusions in some of his published papers, and by present standards this is not acceptable. Scientists must be willing to acknowledge the limitations on their data if they are not to mislead others about the data's reliability.

General rules for distinguishing *a priori* "good" data from "bad" cannot be formulated with much clarity. Nevertheless, good scientists have methods that they can apply in judging the reliability of data, and learning these methods is one of the goals of a scientific apprenticeship. These methods may be unique to a given situation, depending on how and why a set of observations is being made. Nevertheless, they impose constraints on how those observations can be interpreted. A researcher is not free to select only the data that fit his or her prior expectations. If certain data are excluded, a researcher must have justifiable reasons for doing so.



Anton van Leeuwenhoek discovered microorganisms "all alive in a drop of water" using a microscope such as the one shown here. The screws were used to position the sample in front of the lens, which consisted of a glass bead encased in the metal plate.

The Relation Between Hypotheses and Observations

Attempts to isolate the facts and nothing but the facts in scientific research can raise philosophical as well as methodological problems. One prominent difficulty involves the line of demarcation between hypotheses and observations. For years philosophers have tried to construct purely observational languages free of theoretical constructs, but they have never been completely successful. Even a simple description such as "The temperature in this room is 25 degrees centigrade" contains a host of theoretical underpinnings. The thermometer used to measure the temperature is a complex device subject to its own systematic and random errors. And the quantity being measured is not some fundamental attribute of nature but depends in a complex way on the movements and interactions of gas particles, which are described in terms of the kinetic theory of gases, quantum mechanics, and so on.

The terms used in science also contribute to the interpenetration of hypotheses and observations. For example, Anton van Leeuwenhoek, the seventeenth-century Dutch microscopist, prided himself in describing what he saw through his lenses without any theoretical speculation. However, his descriptions were anything but theory-neutral. When he examined the water standing in the gutter outside his window, some of the microscopic creatures he saw were probably *Euglena*. Today we know that these single-celled organisms contain chlorophyll and are more closely related to plants than animals. But because the creatures moved, van Leeuwenhoek called them "animalcules," not "planticules."

Terms such as "energy," "gross national product," "pion," "black hole," "intelligence quotient," and "gene" are clearly derived from particular theories and obtain much of their meaning from their roles in these theories. But such theoretical terms can take on a life of their own and be gradually transformed into more observational terms. Similarly, as terms become unmoored from their original theories, the potential to misuse or misunderstand them increases.

The Risk of Self-Deception

Awareness of the inroads that theory can make into observations serves as a valuable reminder of the constant danger of self-deception in science. Psychologists have shown that people have a tendency to see what they expect to see and fail to notice what they believe should not be there. For instance, during the early part of the twentieth century one of the most ardent debates in astronomy concerned the nature of what were then known as spiral nebulae—diffuse pinwheels of light that powerful telescopes revealed to be quite common in the night sky. Some astronomers thought that these nebulae were spiral galaxies like the Milky Way at such great distances that individual stars could not be distinguished. Others believed that they were clouds of gas within our own galaxy.

One astronomer in the latter group, Adriaan van Maanen of the Mount Wilson Observatory, sought to resolve the issue by comparing photographs of the nebulae taken several years apart. After making a series of painstaking measurements, van Maanen announced that he had found roughly consistent unwinding motions in the nebulae. The detection of such motions indicated that the spirals had to be within the Milky Way, since motions would be impossible to detect in distant objects.

Van Maanen's reputation caused many astronomers to accept a galactic location for the nebulae. A few years later, however, van Maanen's colleague Edwin Hubble, using the new 100-inch telescope at Mount Wilson, conclusively demonstrated that the nebulae were in fact distant galaxies; van Maanen's observations had to be wrong. Studies of his procedures have not revealed any intentional misrepresentation or sources of systematic error. Rather, he was working at the limits of observational accuracy, and he saw what he expected to see.

Self-deception can take more subtle forms. For example, a researcher may stop a data run too early because the observations conform to expectations, whereas a longer run might turn up unexpected discrepancies. Insufficient repetitions of an experiment are a common cause of invalid conclusions, as are poorly controlled experiments.

Methods and Their Limitations

Over the years, scientists have developed a vast array of methods that are designed to minimize the kinds of problems discussed above. At the most familiar level, these methods include techniques such as double-blind trials, randomization of experimental subjects, and the proper use of controls, which are all aimed at reducing individual subjectivity. Methods also include the use of tools in scientific work, both the mechanical tools used to make observations and the intellectual tools used to manipulate abstract concepts.

The term "methods" can be interpreted more broadly. Methods include the judgments scientists make about the interpretation or reliability of data. They

N RAYS

Self-delusion is not a danger only for individual scientists. Sometimes a number of scientists can get caught up in scientific pursuits that later prove to be unfounded.

One of the most famous examples of such "pathological science" is the history of N rays. In the first few years of the twentieth century, shortly after the discovery of X rays by the German physicist Wilhelm Roentgen, the distinguished French physicist Rene Blondlot announced that he had discovered a new type of radiation. Blondlot named the new radiation N rays after the University of Nancy, where he was professor of physics. The rays were supposedly produced by a variety of sources, including electrical discharges within gases and heated pieces of metal; they could be refracted through aluminum prisms, and they could be detected by observing faint visual effects where the rays hit phosphorescent or photographic surfaces. Within a few years, dozens of papers describing the properties of N rays had been published in journals by eminent scientists.

Other scientists, however, found it impossible to duplicate the experiments. One such scientist was the American physicist Robert W. Wood, who traveled to Blondlot's laboratory in 1904 to witness the experiments for himself. After viewing several inconclusive experiments, Wood was shown an experiment by Blondlot in which N rays generated by a lamp were bent through an aluminum prism and fell on a phosphorescent detector. At one point in the experiment, Wood took advantage of the room's darkness to surreptitiously remove the aluminum prism from the apparatus. Nevertheless, Blondlot continued to detect the visual signals that he believed were caused by N rays.

In an article in Nature published shortly after his visit, Wood wrote that he was "unable to report a single observation which appeared to indicate the existence of the rays." Scientific work on N rays soon collapsed, and previous results were shown to be experimental artifacts or the result of observer effects. Yet Blondlot continued to believe in the existence of N rays until his death in 1930.

also include the decisions scientists make about which problems to pursue or when to conclude an investigation. Methods involve the ways scientists work with each other and exchange information. Taken together, these methods constitute the craft of science, and a person's individual application of these methods helps determine that person's scientific style.

Some methods, such as those governing the design of experiments or the statistical treatment of data, can be written down and studied. (The bibliography includes several books on experimental design.) But many methods are learned only through personal experience and interactions with other scientists. Some are even harder to describe or teach. Many of the intangible influences on scientific discovery—curiosity, intuition, creativity—largely defy rational analysis, yet they are among the tools that scientists bring to their work.

Although methods are an integral part of science, most of them are not the product of scientific investigation. They have been developed and their use is required in science because they have been shown to advance scientific knowledge. However, even if perfectly applied, methods cannot guarantee the accuracy of scientific results. Experimental design is often as much an art as a science; tools can introduce errors; and judgments about data inevitably rest on incomplete information.

The fallibility of methods means that there is no cookbook approach to doing science, no formula that can be applied or machine that can be built to generate scientific knowledge. But science would not be so much fun if there were. The skillful application of methods to a challenging problem is one of the great pleasures of science. The laws of nature are not apparent in our everyday surroundings, waiting to be plucked like fruit from a tree. They are hidden and unyielding, and the difficulties of grasping them add greatly to the satisfaction of success.

"A large number of incorrect conclusions are drawn because the possibility of chance occurrences is not fully considered. This usually arises through lack of proper controls and insufficient repetitions. There is the story of the research worker in nutrition who had published a rather surprising conclusion concerning rats. A visitor asked him if he could see more of the evidence. The researcher replied, 'Sure, there's the rat.'"

—E. Bright Wilson, Jr., *An Introduction to Scientific Research*, New York: McGraw-Hill, 1952, p. 34

Values in Science

When methods are defined as all of the techniques and principles that scientists apply in their work, it is easier to see how they can be influenced by human values. As with hypotheses, human values cannot be eliminated from science, and they can subtly influence scientific investigations.

The influence of values is especially apparent during the formulation or judgment of hypotheses. At any given time, several competing hypotheses may explain the available facts equally well, and each may suggest an alternate route for further research. How should one select among them?

Scientists and philosophers have proposed several criteria by which promising scientific hypotheses can be distinguished from less fruitful ones. Hypotheses should be internally consistent, so that they do not generate contradictory conclusions. Their ability to provide accurate predictions, sometimes in areas far removed from the original domain of the hypothesis, is viewed with great favor. With disciplines in which prediction is less straightforward, such as geology or astronomy, good hypotheses should be able to unify disparate observations. Also highly prized are simplicity and its more refined cousin, elegance.

The above values relate to the epistemological, or knowledge-based, criteria applied to hypotheses. But values of a different kind can also come into play in science. Historians, sociologists, and other students of science have shown that social and personal values unrelated to epistemological criteria—including philosophical, religious, cultural, political, and economic values—can shape

Paleoanthropology, with its scarce fossil data and emotionally charged subject matter, has often been influenced by social and personal values as well as by epistemological criteria.

scientific judgment in fundamental ways. For instance, in the nineteenth century the geologist Charles Lyell championed the concept of uniformitarianism in geology, arguing that incremental changes operating over long periods of time have produced the Earth's geological features, not large-scale catastrophes. However, Lyell's preference for this still important idea may have depended as much on his religious convictions as on his geological observations. He favored the notion of a God who is an unmoved mover and does not intervene in His creation. Such a God, thought Lyell, would produce a world where the same causes and effects keep cycling eternally, producing a uniform geological history.

The obvious question is whether holding such values can harm a person's science. In many cases the answer has to be yes. The history of science offers many episodes in which social or personal values led to the promulgation of wrong-headed ideas. For instance, past investigators produced "scientific" evidence for overtly racist views, evidence that we now know to be wholly erroneous. Yet at the time the evidence was widely accepted and contributed to repressive social policies.

Attitudes regarding the sexes also can lead to flaws in scientific judgments. For instance, some investigators who have sought to document the existence or absence of a relationship between gender and scientific abilities have allowed personal biases to distort the design of their studies or the interpretation of their findings. Such biases can contribute to institutional policies that have caused females and minorities to be underrepresented in science, with a consequent loss of scientific talent and diversity.

Conflicts of interest caused by financial considerations are yet another source of values that can harm science. With the rapid decrease in time between fundamental discovery and commercial application, private industry is subsidizing a considerable amount of cutting-edge research. This commercial involvement may bring researchers into conflict with industrial managers—for instance, over the publication of discoveries—or it may bias investigations in the direction of personal gain.

The above examples are valuable reminders of the danger of letting values intrude into research. But it does not follow that social and personal values necessarily harm science. The desire to do accurate work is a social value. So

A theory of continental drift was proposed early in the twentieth century, yet most earth scientists rejected the idea until the 1960s, when new evidence concerning the magnetization of seafloor basalts became available.

is the belief that knowledge will ultimately benefit rather than harm humankind. One simply must acknowledge that values do contribute to the motivations and conceptual outlook of scientists. The danger comes when scientists allow values to introduce biases into their work that distort the results of scientific investigations.

The social mechanisms of science discussed later act to minimize the distorting influences of social and personal values. But individual scientists can avoid pitfalls by trying to identify their own values and the effects those values have on their science. One of the best ways to do this is by studying the history, philosophy, and sociology of science. Human values change very slowly, and the lessons of the past remain of great relevance today.

Judging Hypotheses

Values emerge into particularly sharp relief when a long-established theory comes into conflict with new observations. Individual responses to such situations range between two extremes. At one end of the spectrum is the notion that a theory must be rejected or extensively modified as soon as one of its predictions is not borne out by an experiment. However, history is full of examples in which this would have been premature because not enough was known to make an accurate prediction. A classic example involves Charles Darwin's defense of the theory of evolution. After Darwin presented his theory, physicists argued that the age of the Earth—then calculated to be between 24 million and 100 million years based on the loss of the heat generated by the Earth's formation—could not possibly be long enough for Darwinian evolution to have occurred. Doggedly, although admittedly rather miserably, Darwin hung on. Only after his death was he vindicated. When physicists discovered radioactivity and realized that natural radioactive heating must be included in the Earth's heat budget, there proved to be plenty of time for natural selection to have produced today's species.

On the other hand, history also contains many examples of scientists who held on to an outdated theory after it had been discredited. Human beings have a strong tendency to cling to long-established ideas even in the face of considerable opposing evidence. A trend in the data can always be resisted by citing uncertainties in the observations or by supposing that unknown factors are at work.

Hanging on for a while to a favorite but embattled idea is often a necessity during the initial stages of research. But scientists must also learn to give way in light of new and more insistent evidence. Knowing why an idea is so appealing, or why countervailing evidence is so strongly resisted, can help a person develop this fine sense of discrimination.

Peer Recognition and Priority of Discovery

Human values are also an integral part of the forces that motivate scientists. These forces are numerous and psychologically complex. They include curiosity about the natural or social world, the desire to better the human condition, and a feeling of awe, whether religious or secular, at discerning the workings of nature.

Another important motivating force in science is a desire for recognition by one's peers. One of the greatest rewards scientists can experience is to have their work acknowledged and praised by other scientists and incorporated into their colleagues' research. Sometimes the quest for personal credit can become counterproductive, as when time, energy, or even friendships are lost to priority disputes or ad hominem polemics. But a strong personal attachment to an idea is not necessarily a liability. It can even be essential in dealing with the great effort and frequent disappointments associated with scientific research.

In science, the first person or group to publish a result generally gets the lion's share of credit for it, even if another group that has been working on the problem much longer publishes the same result just a little later. (Actually, priority is dated from when a scientific journal receives a manuscript.) Once published, scientific results become the public property of the research community, but their use by other scientists requires that the original discoverer be recognized. Only when results have become common knowledge are scientists free to use them without attribution.

In deciding when to make a result public, a scientist weighs several competing factors. If a result is kept private, researchers can continue to check its accuracy and use it to further their research. But researchers who refrain from publishing risk losing credit to someone else who publishes first. When considerations such as public acclaim or patent rights are added to the mix, decisions about when to publish can be difficult.

THE HISTORICAL ORIGINS OF PRIORITY

The system of associating scientific priority with publication took shape during the seventeenth century in the early years of modern science. Even then, a tension existed between the need of scientists to have access to other findings and a desire to keep work secret so that others would not claim it as their own. Scientists of the time, including Isaac Newton, were loathe to convey news of their discoveries to scientific societies for fear that someone else would claim priority, a fear that was frequently realized.

To ensure priority, many scientists, including Galileo, Huygens, and Newton, resorted to constructing anagrams describing their discoveries that they would then make known to others. For instance, the law "mass times acceleration equals force" could be disguised as "a remote, facile question scares clams" (though Newton would have constructed his anagrams in Latin). Later, if someone else came up with the same discovery, the original discoverer could unscramble the anagram to establish priority.

The solution to the problem of making new discoveries public while assuring their authors credit was worked out by Henry Oldenburg, the secretary of the Royal Society of London. He won over scientists by guaranteeing rapid publication in the Philosophical Transactions of the society as well as the official support of the society in case the author's priority was brought into question. Thus, it was originally the need to ensure open communication in science that gave rise to the convention that the first to publish a view or a finding, not the first to discover it, gets credit for the discovery.