

Conduct, Misconduct and the Structure of Science

Many plausible-sounding rules for defining ethical conduct might be destructive to the aims of scientific inquiry

James Woodward and David Goodstein

In recent years the difficult question "what constitutes scientific misconduct?" has troubled prominent ethicists and scientists and tied many a blue-ribbon panel in knots. In teaching an ethics class for graduate and undergraduate students over the past few years, we have identified what seems to be a necessary starting point for this debate: the clearest possible understanding of *how science actually works*. Without such an understanding, we believe, one can easily imagine formulating plausible-sounding ethical principles that would be unworkable or even damaging to the scientific enterprise.

Our approach may sound so obvious as to be simplistic, but actually it uncovers a fundamental problem, which we shall try to explore in this article. The nature of the problem can be glimpsed by considering the ethical implications of the earliest theory of the scientific method. Sir Francis Bacon, a contemporary of Galileo, thought the scientist must be a disinterested observer of nature, whose mind was cleansed of prejudices and preconceptions. As we shall see, the reality of science is radically different from this ideal. If we expect to find scientists who are disinterested observers of nature we are bound to be disappointed, not because scientists have failed to measure up to the appropriate standard of behavior, but because we have tried to apply the wrong standard of behavior. It can be worse: Rules or standards of conduct that seem intuitively appealing can turn out to have results that are both unexpected and destructive to the aims of scientific inquiry.

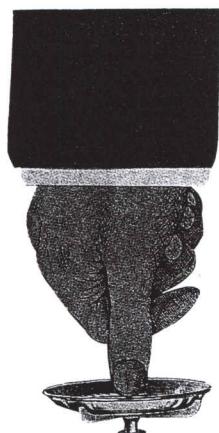
In drafting this article, we set out to examine the question of scientific ethics in light of what we know about science as a system and about the motivations of the scientists who take part in it. The reader will find that this exercise unearths contradictions that may be

especially unpleasant for those who believe clear ethical principles derive directly from the principles of scientific practice. In fact, one can construct a wonderful list of plausible-sounding ethical principles, each of which might be damaging or unworkable according to our analysis of how science works.

Ideals and Realities

We can begin where Sir Francis left off. Here is a hypothetical set of principles, beginning with the Baconian ideal, for the conduct of science:

1. Scientists should always be disinterested, impartial and totally objective when gathering data.
2. A scientist should never be motivated to do science for personal gain, advancement or other rewards.
3. Every observation or experiment must be designed to falsify an hypothesis.
4. When an experiment or an observation gives a result contrary to the prediction of a certain theory, all ethical scientists must abandon that theory.
5. Scientists must never believe dogmatically in an idea nor use rhetorical exaggeration in promoting it.
6. Scientists must "lean over backwards" (in the words of the late physicist Richard Feynman) to point out evidence that is contrary to their own hypotheses or that might weaken acceptance of their experimental results.
7. Conduct that seriously departs from that commonly accepted in the scientific community is unethical.
8. Scientists must report what they have done so fully that any other scientist can reproduce the experiment or calculation. Science must be an open book, not an acquired skill.



James Woodward and David Goodstein are professors at Caltech, where, in addition to the Research Ethics course they teach together, Woodward teaches philosophy and Goodstein teaches physics. Woodward has served Caltech as Executive Officer for the Humanities; Goodstein serves as vice provost and is the Frank J. Gilloon Distinguished Teaching and Service Professor. Address for both: California Institute of Technology, Pasadena, CA 91125.

Behavior that may seem at first glance morally unattractive can, in a properly functioning system, produce results that are generally beneficial.

9. Scientists should never permit their judgments to be affected by authority. For example, the reputation of a scientist making a given claim is irrelevant to the validity of the claim.
10. Each author of a multiple-author paper is fully responsible for every part of the paper.
11. The choice and order of authors on a multiple-author publication must strictly reflect the contributions of the authors to the work in question.
12. Financial support for doing science and access to scientific facilities should be shared democratically, not concentrated in the hands of a favored few.
13. There can never be too many scientists in the world.
14. No misleading or deceptive statement should ever appear in a scientific paper.
15. Decisions about the distribution of resources and publication of results must be guided by the judgment of scientific peers who are protected by anonymity.

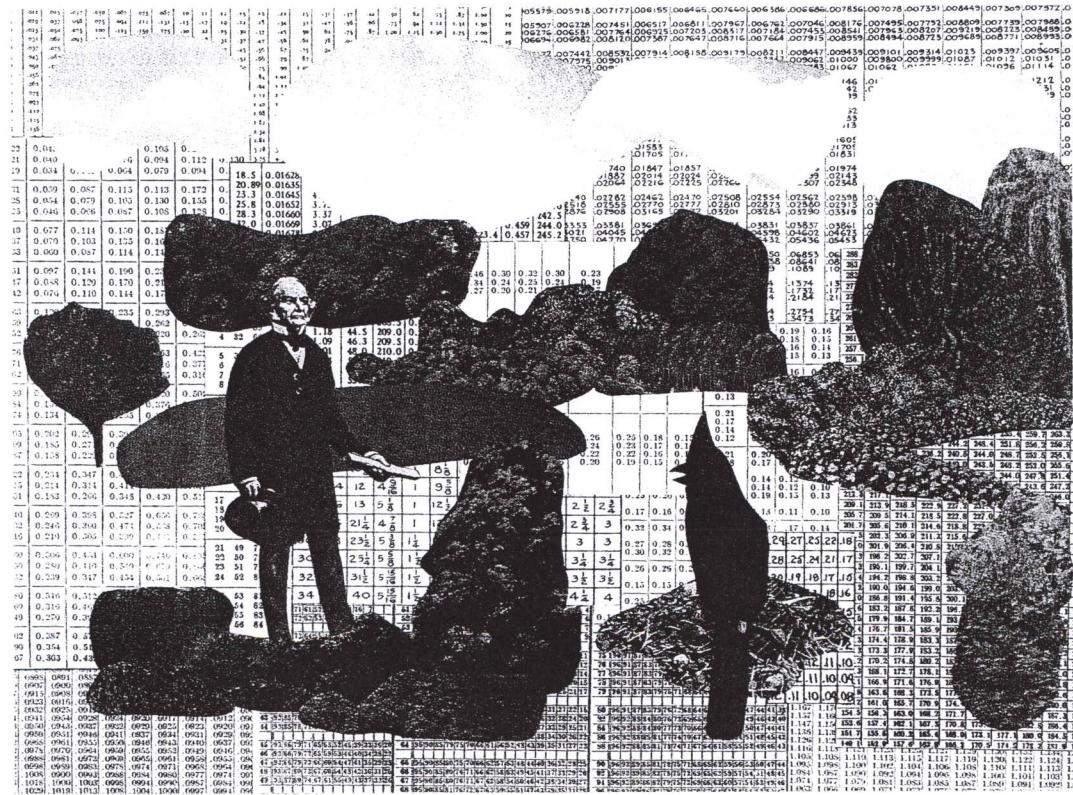
Should the behavior of scientists be governed by rules of this sort? We shall argue that it should not. We first consider the general problems of motivation and the logical structure of science, then the question of how the community of scientists actually does its work, showing along the way why each of these principles is defective. At the end, we offer a positive suggestion of how scientific misconduct might be recognized.

Motives and Consequences

Many of the provocative statements we have just made raise general questions of motivation related to the issue explicitly raised in principle 2, and it is worth dealing with these up front. We might begin with a parallel: the challenge of devising institutions, rules and standards to govern commerce. In economic life well-intentioned attempts to reduce the role of greed or speculation can turn out to have disastrous consequences. In fact, behavior that may seem at first glance morally unattractive, such as the aggressive pursuit of economic self-interest, can, in a properly functioning system, produce results that are generally beneficial.

In the same way it might appear morally attractive to demand that scientists take no interest in obtaining credit for their achievements. Most scientists are motivated by the desire to discover important truths about nature and to help others to do so. But they also prefer that they (rather than their competitors) be the ones to make discoveries, and they want the recognition and the advantages that normally reward success in science. It is tempting to think that tolerating a desire for recognition is a concession to human frailty; ideally, scientists should be interested only in truth or other purely epistemic goals. But this way of looking at matters misses a number of crucial points.

For one thing, as the philosopher Philip Kitcher has noted, the fact that the first person



to make a scientific discovery usually gets nearly all the credit encourages investigators to pursue a range of different lines of inquiry, including lines that are thought by most in the community to have a small probability of success. From the point of view of making scientific discoveries as quickly and efficiently as possible, this sort of diversification is extremely desirable; majority opinion turns out to be wrong with a fairly high frequency in science.

Another beneficial feature of the reward system is that it encourages scientists to make their discoveries public. As Noretta Koertge has observed, there have been many episodes in the history of early modern science in which scientists made important discoveries and kept them private, recording them only in notebooks or correspondence, or in cryptic announcements designed to be unintelligible to others. The numerous examples include Galileo, Newton, Cavendish and Lavoisier. It is easy to see how such behavior can lead to wasteful repetition of effort. The problem is solved by a system of rewards that appeals to scientists' self-interest. Finally, in a world of limited scientific resources, it makes sense to give more resources to those who are better at making important discoveries.

We need to be extremely careful, in designing institutions and regulations to discourage scientific misconduct, that we not introduce changes that disrupt the beneficial effects that competition and a concern for credit and reputation bring with them. It is frequently claimed that an important motive in a number of re-

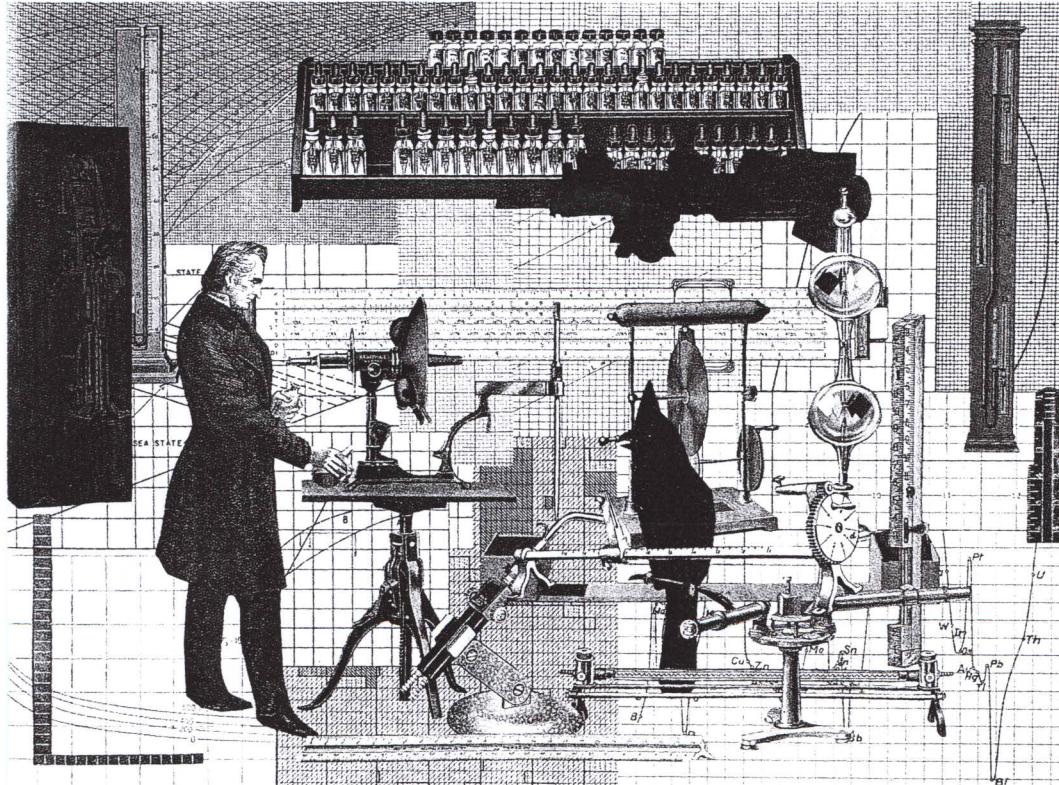
cent cases of data fabrication has been the desire to establish priority and to receive credit for a discovery, or that a great deal of fraud can be traced to the highly competitive nature of modern science. If these claims are correct, the question becomes, how can we reduce the incidence of fraud without removing the beneficial effects of competition and reward?

The Logical Structure of Science

The question of how science works tends to be discussed in terms of two particularly influential theories of scientific method, *Baconian inductivism* and *Popperian falsification*, each of which yields a separate set of assumptions.

According to Bacon's view, scientific investigation begins with the careful recording of observations. These should be, insofar as is humanly possible, uninfluenced by any prior prejudice or theoretical preconception. When a large enough body of observations is accumulated, the investigator generalizes from these, via a process of induction, to some hypothesis or theory describing a pattern present in the observations. Thus, for example, an investigator might inductively infer, after observing a large number of black ravens, that all ravens are black. According to this theory good scientific conduct consists in recording all that one observes and not just some selected part of it, and in asserting only hypotheses that are strongly inductively supported by the evidence. The guiding ideal is to avoid any error that may slip in as a result of prejudice or preconception.

How can we reduce the incidence of fraud without removing the beneficial effects of competition and reward?



Historians, philosophers and those scientists who care are virtually unanimous in rejecting Baconian inductivism as a general characterization of good scientific method. The advice to record all that one observes is obviously unworkable if taken literally; some principle of selection or relevance is required. But decisions about what is relevant inevitably will be influenced heavily by background assumptions, and these, as many recent historical studies show, are often highly theoretical in character. The vocabulary we use to describe the results of measurements, and even the instruments we use to make the measurements, are highly dependent on theory. This point is sometimes expressed by saying that all observation in science is "theory-laden" and that a "theoretically neutral" language for recording observations is impossible.

The idea that science proceeds only and always via inductive generalization from what is directly observed is also misguided. Theories in many different areas of science have to do with entities whose existence or function cannot be directly observed: forces, fields, subatomic particles, proteins and other large organic molecules, and so on. For this and many other reasons, no one has been able to formulate a defensible theory of inductive inference of the sort demanded by inductivist theories of science.

The difficulties facing inductivism as a general conception of scientific method are so well known that it is surprising to find authoritative characterizations of scientific misconduct that appear to be influenced by this conception. Consider the following remarks by Suzanne Hadley, at one time acting head of what used to be called the Office of Scientific Integrity (now the Office of Research Integrity),

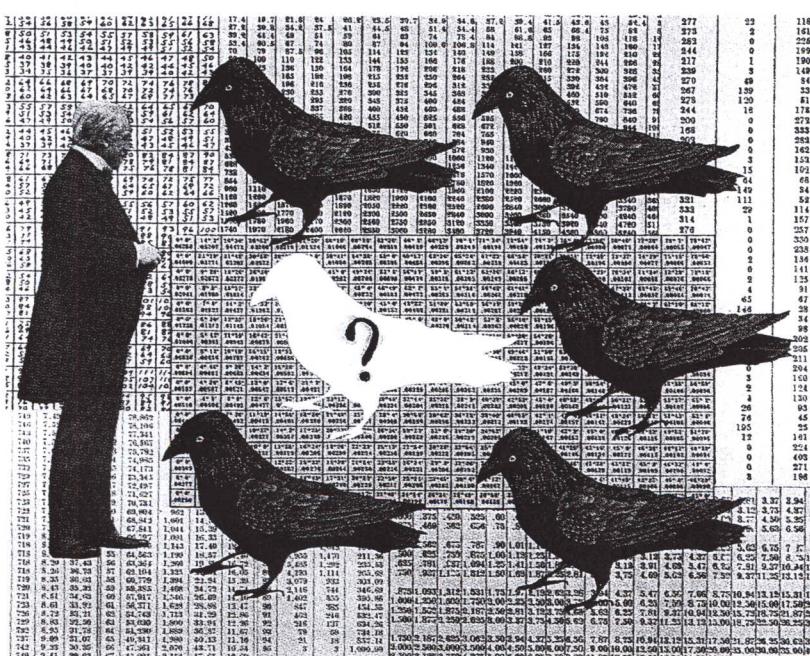
the arm of the U.S. Public Health Service charged with investigating allegations of scientific misconduct. In a paper presented at the University of California, San Diego, in October 1991, Hadley wrote: "...it is essential that observation, data recording, and data interpretation and reporting be veridical with the phenomena of interest, *i.e.*, be as free as humanly possible of 'taint' due to the scientist's hopes, beliefs, ambitions, or desires." Elsewhere she writes: "Anything that impinges on the veridical perception, recording and reporting of scientific phenomena is antithetical to the very nature of science." She also says, "...it is the human mind, which based on trained observations, is able to form higher-order conceptions about phenomena."

Hadley's view may not be as rigidly inductivist as these remarks imply. She adds, "I hasten to say that I am not suggesting that a scientist can or should be relegated to a mechanistic recording device." But this more nuanced view is not allowed to temper excessively her oversight responsibility as a government official: "The really tough cases to deal with are the cases closest to the average scientist: those in which 'fraud' is not clearly evident, but 'out of bounds' conduct is: data selection, failure to report discrepant data, over-interpretation of data...."

The idea that data selection and overinterpretation of data are forms of misconduct seem natural if one begins with Hadley's view of scientific method. A less restrictive view would lead to a different set of conclusions about what activities constitute misconduct.

Although relatively few contemporary scientists espouse inductivism, there are many scientists who have been influenced by the falsificationist ideas of Karl Popper. According to falsificationists, we test a hypothesis by deducing from it a testable prediction. If this prediction turns out to be false, the hypothesis from which it is deduced is said to be falsified and must be rejected. For example, the observation of a single nonblack raven will falsify the hypothesis H: "All ravens are black." But if we set out to test H and observe a black raven or even a large number of such ravens we cannot, according to Popper, conclude that H is true or verified or even that it is more probable than it was before. All that we can conclude is that H has been tested and has not yet been falsified. There is thus an important asymmetry between the possibility of falsification and the possibility of verification; we can show conclusively that an hypothesis is false, but not that it is true.

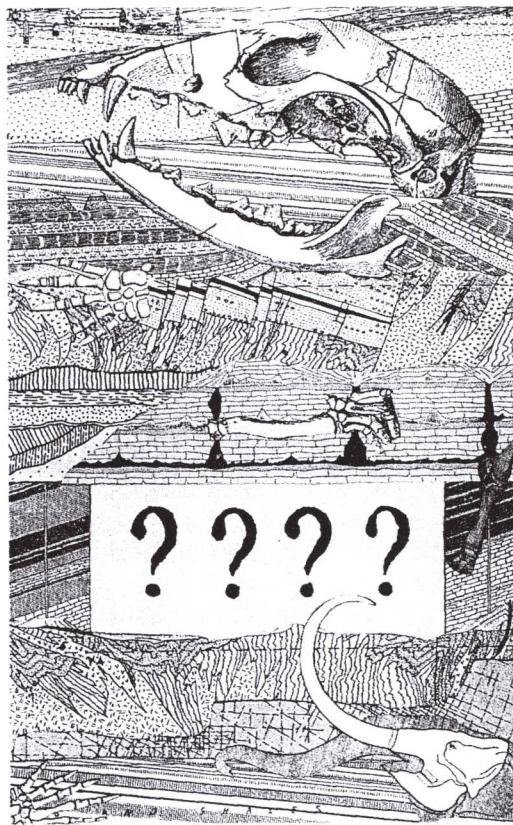
Because of this asymmetry it is a mistake to think, as the inductivist does, that good science consists of hypotheses that are proved or made probable by observation, whereas bad science does not. Instead, according to Popper, good



science requires hypotheses that might be falsified by some conceivable observation. For example, the general theory of relativity predicts that starlight passing sufficiently close to the sun will be deflected by a certain measurable amount E . General relativity is a falsifiable theory because observations of starlight deflection that differ substantially from E are certainly conceivable and had they been made, would have served to falsify general relativity. By contrast, writes Popper, Freudian psychology is unfalsifiable and hence unscientific. If, for example, a son behaves in a loving way toward his mother, this will be attributed to his Oedipus complex. If, on the contrary, he behaves in an hostile and destructive way, this will be attributed to the same Oedipus complex. No possible empirical observation constitutes a refutation of the hypothesis that the son's behavior is motivated by an Oedipus complex.

According to Popper, bad scientific behavior consists in refusing to announce in advance what sorts of evidence would lead one to give up an hypothesis, in ignoring or discarding evidence contrary to one's hypothesis or in introducing ad hoc, content-decreasing modifications in one's theories in order to protect them against refutation. Good scientific method consists in putting forward highly falsifiable hypotheses, specifying in advance what sorts of evidence would falsify these hypotheses, testing the hypotheses at exactly those points at which they seem most likely to break down and then giving them up should such evidence be observed. More generally (and moving somewhat beyond the letter of Popper's theory) we can say that to do science in a Popperian spirit is to hold to one's hypothesis in a tentative, nondogmatic fashion, to explore and draw to the attention of others various ways in which one's hypothesis might break down or one's experimental result may be invalid, to give up one's hypothesis willingly in the face of contrary evidence, to take seriously rather than to ignore or discard evidence that is contrary to it, and in general not to exaggerate or overstate the evidence for it or suppress problems that it faces. Richard Feynman, in a commencement address at Caltech some years ago, recommended a recognizably Popperian attitude in the following remarks:

[There is an] idea that we all hope you have learned in studying science in school—we never explicitly say what this is, but just hope that you catch on by all the examples of scientific investigation.... It's a kind of scientific integrity, a principle of scientific thought that corresponds to a kind of utter honesty—a kind of leaning over backwards. For example, if you're doing an experiment, you should report everything that you think might make it invalid—not only what you



think is right about it; other causes that could possibly explain your results; and things you thought of that you've eliminated by some other experiment, and how they worked—to make sure the other fellow can tell they have been eliminated.

....In summary, the idea is to try to give all the information to help others to judge the value of your contribution, not just the information that leads to judgment in one particular direction or another.

These views form the basis of principles 3–6 above.

Although falsificationism has many limitations (see below), it introduces several corrections to inductivism that are useful in understanding how science works and how to characterize misconduct. To begin with, falsificationism rejects the idea that good scientific behavior consists in making observations without theoretical preconceptions. For Popper, scientific activity consists in attempting to falsify. Such testing requires that one have in mind a hypothesis that will indicate which observations are relevant or worth making. Rather than something to be avoided, theoretical preconceptions are essential to doing science.

Inductivists attach a great deal of weight to the complete avoidance of error. By contrast, falsificationists claim that the history of science shows us that all hypotheses are falsified sooner or later. In view of this fact, our aim should be to detect our errors quickly and to learn as

For science to advance, scientists must be free to be wrong.



It is sometimes a good strategy to hold onto a hypothesis even when it seems to imply an observational consequence that looks to be false.

efficiently as possible from them. Error in science thus plays a constructive role. Indeed, according to falsificationists, putting forward a speculative “bold conjecture” that goes well beyond available evidence and then trying vigorously to falsify it will be the strategy that enables us to progress as efficiently as possible. For science to advance, scientists must be free to be wrong.

Despite these advantages, there are also serious deficiencies in falsificationism, when it is taken as a general theory of method. One of the most important of these is sometimes called the Duhem-Quine problem. We claimed above that testing a hypothesis H involved deriving from it some observational consequence O . But in most realistic cases such observational consequences will not be derivable from H alone, but only from H in conjunction with a great many other assumptions A (auxiliary assumptions, as philosophers sometimes call them). For example, to derive an observational claim from a hypothesis about rates of evolution, one may need auxiliary assumptions about the processes by which the fossil record is laid down. Suppose one hypothesizes that a certain organism has undergone slow and continuous evolution and derives from this that one should see numerous intermediate forms in the fossil record. If such forms are absent it may mean H is false, but it may also be the case that H is true but fossils were preserved only in geological deposits that were laid down at widely separated times. It is possible that H is true and that the reason that O is false is that A is false.

One immediate result of this simple logical fact is that the logical asymmetry between falsification and verification disappears. It may be true, as Popper claims, that we cannot conclusively verify a hypothesis, but we cannot conclusively falsify it either. Thus, as a matter of method, it is sometimes a good strategy to hold onto a hypothesis even when it seems to imply an observational consequence that looks to be false. In fact, the history of science is full of examples in which such anti-Popperian behavior has succeeded in finding out important truths about nature when it looks as though more purely Popperian strategies would have been less successful.

Anti-Popperian strategies seem particularly prevalent in experiments. In doing an experiment one’s concern is often to find or demonstrate an effect or to create conditions that will allow the effect to appear, rather than to refute the claim that the effect is real. Suppose a novel theory predicts some previously unobserved effect, and an experiment is undertaken to detect it. The experiment requires the construction of new instruments, perhaps operating at the very edge of what is technically possible, and the use of a novel experimental design,

which will be infected with various unsuspected and difficult-to-detect sources of error. As historical studies have shown, in this kind of situation there will be a strong tendency on the part of many experimentalists to conclude that these problems have been overcome if and when the experiment produces results that the theory predicted. Such behavior certainly exhibits anti-Popperian dogmatism and theoretical “bias,” but it may be the best way to discover a difficult-to-detect signal. Here again, it would be unwise to have codes of scientific conduct or systems of incentives that discourage such behavior.

Social Structure

Inductivism, falsificationism and many other traditional accounts of method are inadequate as theories of science. At bottom this is because they neglect the psychology of individual scientists and the social structure of science. These points are of crucial importance in understanding how science works and in characterizing scientific misconduct.

Let us begin with what Philip Kitcher has called the division of cognitive labor and the role of social interactions in scientific investigation. Both inductivism and falsificationism envision an individual investigator encountering nature and constructing and assessing hypotheses all alone. But science is carried out by a community of investigators. This fact has important implications for how we should think about the responsibilities of individual scientists.

Suppose a scientist who has invested a great deal of time and effort in developing a theory is faced with a decision about whether to continue to hold onto it given some body of evidence. As we have seen, good Popperian method requires that scientists act as skeptical critics of their own theories. But the social character of science suggests another possibility. Suppose that our scientist has a rival who has invested time and resources in developing an alternative theory. If additional resources, credit and other rewards will flow to the winner, perhaps we can reasonably expect that the rival will act as a severe Popperian critic of the theory, and vice versa. As long as others in the community will perform this function, failure to behave like a good Popperian need not be regarded as a violation of some canon of method.

There are also psychological facts to consider. In many areas of science it turns out to be very difficult, and to require a long-term commitment of time and resources, to develop even one hypothesis that respects most available theoretical constraints and is consistent with most available evidence. Scientists, like other human beings, find it difficult to sustain commitments to arduous, long-term projects

if they spend too much time contemplating the various ways in which the project might turn out to be unsuccessful.

A certain tendency to exaggerate the merits of one's approach, and to neglect or play down, particularly in the early stages of a project, contrary evidence and other difficulties, may be a necessary condition for the success of many scientific projects. When people work very hard on something over a long period of time, they tend to become committed or attached to it; they strongly want it to be correct and find it increasingly difficult to envision the possibility that it might be false, a phenomenon related to what psychologists call *belief-perseverance*. Moreover, scientists like other people like to be right and to get credit and recognition from others for being right: The satisfaction of demolishing a theory one has laboriously constructed may be small in comparison with the satisfaction of seeing it vindicated. All things considered, it is extremely hard for most people to adopt a consistently Popperian attitude toward their own ideas.

Given these realistic observations about the psychology of scientists, an implicit code of conduct that encourages scientists to be a bit dogmatic and permits a certain measure of rhetorical exaggeration regarding the merits of their work, and that does not require an exhaustive discussion of its deficiencies, may be perfectly sensible. In many areas of science, if a scientist submits a paper that describes all of the various ways in which an idea or result might be defective, and draws detailed attention to the contrary results obtained by others, the paper is likely to be rejected. In fact, part of the intellectual responsibility of a scientist is to provide the best possible case for important ideas, leaving it to others to publicize their defects and limitations. Studies of both historical and contemporary science seem to show that this is just what most scientists do.

If this analysis is correct, there is a real danger that by following proposals (like that advocated by Hadley) to include within the category of "out-of-bounds conduct" behavior such as overinterpretation of data, exaggeration of available evidence that supports one's conclusion or failure to report contrary data, one may be proscribing behavior that plays a functional role in science and that, for reasons rooted deep in human psychology, will be hard to eliminate. Moreover, such proscriptions may be unnecessary, because interactions between scientists and criticisms by rivals may by themselves be sufficient to remove the bad consequences at which the proscriptions are aimed. Standards that might be optimal for single, perfectly rational beings encountering nature all by themselves may be radically deficient when applied to actual scientific communities.

Rewarding Useful Behavior

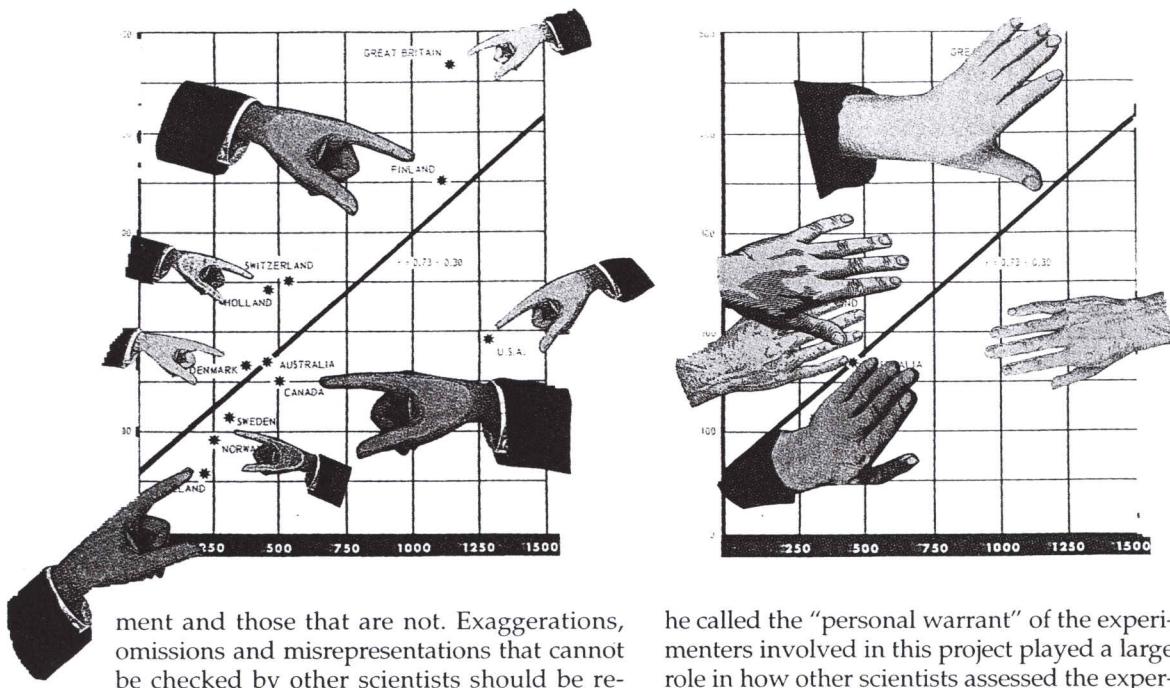
From a Popperian perspective, discovering evidence that merely supports a hypothesis is easy to do and has little methodological value; therefore one might think it doesn't deserve much credit. It is striking that the actual distribution of reward and credit in science reflects a very different view. Scientists receive Nobel prizes for finding new effects predicted by theories or for proposing important theories that are subsequently verified. It is only when a hypothesis or theory has become very well established that one receives significant credit for refuting it. Unquestionably, rewarding confirmations over refutations provides scientists with incentives to confirm theories rather than refute them and thus discourages giving up too quickly in the face of contrary experimental results. But, as we have been arguing, this is not necessarily bad for science.

Conventional accounts of scientific method (of which there are many examples in the philosophical literature) share the implicit assumption that all scientists in a community should adopt the same strategies. In fact, a number of government agencies now have rules that define as scientific misconduct "practices that seriously deviate from those that are commonly accepted within the scientific community..." (see principle 7). But rapid progress will be more likely if different scientists have quite different attitudes toward appropriate methodology. As noted above, one important consequence of the winner-takes-all (or nearly all) system by which credit and reward are allocated in science is that it encourages a variety of research programs and approaches. Other features of human cognitive psychology—such as the belief-perseverance phenomenon described above—probably have a similar consequence. It follows that attempts to characterize misconduct in terms of departures from practices or methods commonly accepted within the scientific community will be doubly misguided: Not only will such commonly accepted practices fail to exist in many cases, but it will be undesirable to try to enforce the uniformity of practice that such a characterization of misconduct would require. More generally, we can see why the classical methodologists have failed to discover "the" method by which science works. There are deep, systematic reasons why all scientists should not follow some single uniform method.

Our remarks so far have emphasized the undesirability of a set of rules that demand that all scientists believe the same things or behave in the same way, given a common body of evidence. This is not to say, however, that "anything goes." One very important distinction has to do with the difference between claims and behavior that are open to public assess-

A certain tendency to exaggerate the merits of one's approach, and to neglect or play down contrary evidence and other difficulties, may be a necessary condition for the success of many scientific projects.





Standards that might be optimal for single, perfectly rational beings encountering nature all by themselves may be radically deficient when applied to actual scientific communities.

ment and those that are not. Exaggerations, omissions and misrepresentations that cannot be checked by other scientists should be regarded much more harshly than those that can, because they subvert the processes of public assessment and intellectual competition on which science rests. Thus, for example, a scientist who fabricates data must be judged far more harshly than one who does a series of experiments and accurately records the results but then extrapolates beyond the recorded data or insists on fitting some favored function to them. The difference is the fact that in the case in which there is no fabrication nothing has been done to obstruct the critical scrutiny of the work by peers; they can look at the data themselves and decide whether there is support for the conclusions. By contrast, other scientists will not be able to examine firsthand the process by which the data have been produced. They must take it on trust that the data resulted from an experiment of the sort described. Fabrication should thus be viewed as much more potentially damaging to the process of inquiry and should be more harshly punished than other forms of misrepresentation.

Science as Craft

Contemporary scientific knowledge is so vast and complex that even a very talented and hardworking scientist will be able to master only an extremely small fragment well enough to expect to make contributions to it. In part for this reason scientists must rely heavily on the authority of other scientists who are experts in domains in which they are not. A striking example of this is provided by the sociologist Trevor Pinch's recent book, *Confronting Nature*, which is a study of a series of experiments that discovered in the solar-neutrino flux far fewer neutrinos than seemed to be predicted by accepted theory. Pinch found that what

he called the "personal warrant" of the experimenters involved in this project played a large role in how other scientists assessed the experimental results. According to Pinch, other scientists often place at least as much weight on an experimentalist's general reputation for careful, painstaking work as on the technical details of the experiment in assessing whether the data constitute reliable evidence.

One reason why such appeals to personal warrant play a large role in science has to do with the specialized character of scientific knowledge. There is, however, another related reason which is of considerable importance in understanding how science works and how one should think about misconduct. This has to do with the fact that science in general—and especially experimentation—has a large "skill" or "craft" component.

Conducting an experiment in a way that produces reliable results is not a matter of following algorithmic rules that specify exactly what is to be done at each step. As Pinch put it, experimenters possess skills that "often enable the experimenter to get the apparatus to work without being able to formulate exactly or completely what has been done." For the same reason, assessing whether another investigator has produced reliable results requires a judgment of whether the experimenter has demonstrated the necessary skills in the past. These facts about the role of craft knowledge may be another reason why the general rules of method sought by the classical methodologists have proved so elusive.

The importance of craft in science is supported by empirical studies. For example, in a well-known study, Harry Collins investigated a number of experimental groups working in Britain to recreate a new kind of laser that had been successfully constructed elsewhere. Collins found that no group was able to reproduce a working laser simply on the basis

of detailed written instructions. By far the most reliable method was to have someone from the original laboratory who had actually built a functioning laser go to the other laboratories and participate in the construction. The skills needed to make a working device could be acquired by practice "*without necessarily formulating, enumerating or understanding them.*" Remarks on experimental work by working scientists themselves often express similar claims, not notwithstanding principle 8 above. If claims of this sort are correct, it often will be very difficult for those who lack highly specific skills and knowledge to assess a particular line of experimental work. A better strategy may be to be guided at least in large part by the experimenter's general reputation for reliability.

These facts about specialization, skill and authority have a number of interesting consequences for understanding what is proper scientific conduct. For example, a substantial amount of conduct that may look to an outsider like nonrational deference to authority may have a serious epistemological rationale. When an experimentalist discards certain data on the basis of subtle clues in the behavior of the apparatus, and other scientists accept the experimentalist's judgment in this matter, we should not automatically attribute this to the operation of power relationships, as is implied by principle 9 in our list above.

A second important consequence has to do with the responsibility of scientists for the misconduct or sloppy research practices of collaborators. It is sometimes suggested that authors should not sign their names to joint papers unless they have personally examined the evidence and are prepared to vouch for the correctness of every claim in the paper (principle 10). However, many collaborations bring together scientists from quite different specializations who lack the expertise to evaluate one another's work directly. This is exactly why collaboration is necessary. Requiring that scientists not collaborate unless they are able to check the work of collaborators or setting up a general policy of holding scientists responsible for the misconduct of coauthors would discourage a great deal of valuable collaboration.

Understanding the social structure of science and the operation of the reward system within science also has important ethical implications. We consider three examples: the Matthew effect, the Ortega hypothesis and scientific publication.

Matthew vs. Ortega

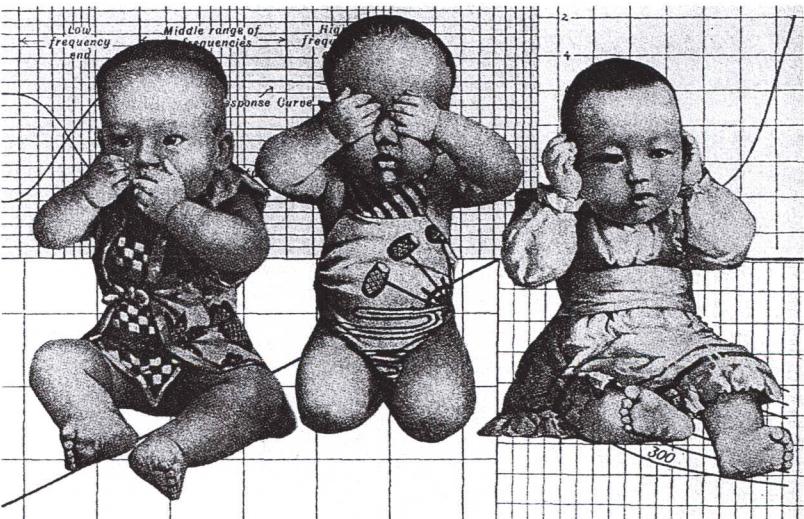
The sociologist Robert K. Merton has observed that credit tends to go to those who are already famous, at the expense of those who are not. A paper signed by Nobody, No-

body and Somebody often will be casually referred to as "work done in Somebody's lab," and even sometimes cited (incorrectly) in the literature as due to "Somebody *et al.*" Does this practice serve and accurately depict science, or does the tendency to elitism distort and undermine the conduct of science?

It is arguable that what Merton called the Matthew Effect plays a useful role in the organization of science; there are so many papers in so many journals that no scientist has time to read more than a tiny fraction of those in even a restricted area of science. Famous names tend to identify those works that are more likely to be worth noticing. In certain fields, particularly biomedical fields, it has become customary to make the head of the laboratory a coauthor, even if the head did not participate in the research. One reason for this practice is that by including the name of the famous head on the paper, chances are greatly improved that the paper will be accepted by a prestigious journal and noticed by its readers. Some people refer to this practice as "guest authorship" and regard it as unethical (as would be implied by principle 11 above). However, the practice may be functionally useful and may involve little deception, since conventions regarding authorship may be well understood by those who participate in a given area of science.

"In the cathedral of science," a famous scientist once said, "every brick is equally important." The remark (heard by one of the authors at a gathering at the speaker's Pasadena, California home) evokes a vivid metaphor of swarms of scientific workers under the guidance, perhaps, of a few master builders erecting a grand monument to scientific faith. The speaker was Max Delbrück, a Nobel laureate often called the father of molecular biology. The remark captures with some precision the scientists' ambivalent view of their craft. Delbrück never for an instant thought the bricks

A substantial amount of conduct that may look to an outsider like nonrational deference to authority may have a serious epistemological rationale.



If the elitist view is right, then since science is largely financed by the public purse, it is best for science and best for society to restrict our production to fewer and better scientists.

he laid were no better than anyone else's. If anything, he regarded himself as the keeper of the blueprints, and he had the fame and prestige to prove it. It was exactly his exalted position that made it obligatory that he make a ceremonial bow to the democratic ideal that many scientists espouse and few believe. In fact it is precisely the kind of recognition that Delbrück enjoyed that propels the scientific enterprise forward.

The view expressed by Delbrück has been called the Ortega Hypothesis. It is named after Jose Ortega y Gasset, who wrote in his classic book, *The Revolt of the Masses*, that

...it is necessary to insist upon this extraordinary but undeniable fact: experimental science has progressed thanks in great part to the work of men astoundingly mediocre. That is to say, modern science, the root and symbol of our actual civilization, finds a place for the intellectually commonplace man and allows him to work therein with success. In this way the majority of scientists help the general advance of science while shut up in the narrow cell of their laboratory like the bee in the cell of its hive, or the turnspit of its wheel.

This view (see principle 12) is probably based on the empirical observation that there are indeed, in each field of science, many ordinary scientists doing more or less routine work. It is also supported by the theoretical view that knowledge of the universe is a kind

of limitless wilderness to be conquered by relentless hacking away of underbrush by many hands. An idea that is supported by both theory and observation always has a very firm standing in science.

The Ortega Hypothesis was named by Jonathan and Steven Cole when they set out to demolish it, an objective they pursued by tracking citations in physics journals. They concluded that the hypothesis is incorrect, stating:

It seems, rather, that a relatively small number of physicists produce work that becomes the base for future discoveries in physics. We have found that even papers of relatively minor significance have used to a disproportionate degree the work of the eminent scientists....

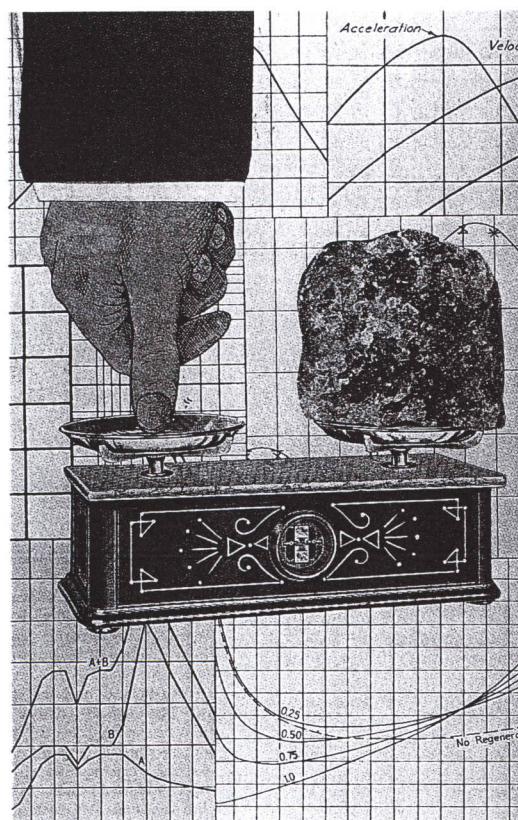
In other words, a small number of elite scientists produce the vast majority of scientific progress. Seen in this light, the reward system in science is a mechanism evolved for the purpose of identifying, promoting and rewarding the star performers.

One's view of the Ortega Hypothesis has important implications concerning how science ought to be organized. If the Ortega Hypothesis is correct, science is best served by producing as many scientists as possible, even if they are not all of the highest quality (principle 13). On the other hand, if the elitist view is right, then since science is largely financed by the public purse, it is best for science and best for society to restrict our production to fewer and better scientists. In any case, the question of whether to produce more or fewer scientists involves ethical issues (what is best for the common good?) as well as policy issues (how to reach the desired goal).

Peers and Publication

In a classic paper called "Is the scientific paper a fraud?" Peter Medawar has argued that typical experimental papers intentionally misrepresent the actual sequence of events involved in the conduct of an experiment, the process of reasoning by which the experimenter reached various conclusions and so on. In general, experimentalists will make it look as if they had a much clearer idea of the ultimate result than was actually the case. Misunderstandings, blind alleys and mistakes of various sorts will fail to appear in the final written account.

Papers written this way are undoubtedly deceptive, at least to the uninitiated, and they certainly stand in contrast to Feynman's exhortation to "lean over backward." They also violate principle 14. Nevertheless, the practice is virtually universal, because it is a much more efficient means of transmitting results than an accurate historical account of the scientist's activities would be. Thus it is a simple



fact that, contrary to normal belief, there are types of misrepresentation that are condoned and accepted in scientific publications, whereas other types are harshly condemned.

Nevertheless, scientific papers have an exalted reputation for integrity. That may be because the integrity of the scientific record is protected, above all, by the institution of peer review. Peer review has an almost mystical role in the community of scientists. Published results are considered dependable because they have been reviewed by peers, and unpublished data are considered not dependable because they have not been. Many regard peer review to be (as principle 15 would suggest) the ethical fulcrum of the whole scientific enterprise.

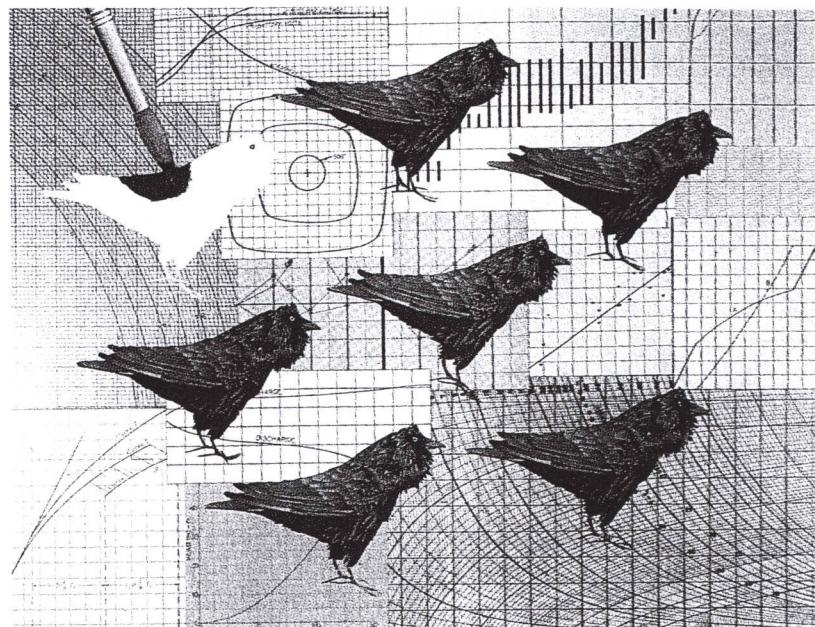
Peer review is used to help determine whether journals should publish articles submitted to them, and whether agencies should grant financial support to research projects. For most small projects and for nearly all journal articles, peer review is accomplished by sending the manuscript or proposal to referees whose identities will not be revealed to the authors.

Peer review conducted in this way is extremely unlikely to detect instances of intentional misconduct. But the process is very good at separating valid science from nonsense. Referees know the current thinking in a field, are aware of its laws, rules and conventions, and will quickly detect any unjustified attempt to depart from them. Of course, for precisely this reason peer review can occasionally delay a truly visionary or revolutionary idea, but that may be a price that we pay for conducting science in an orderly way.

Peer review is less useful for adjudicating an intense competition for scarce resources. The pages of prestigious journals and the funds distributed by government agencies have become very scarce resources in recent times. The fundamental problem in using peer review to decide how these resources are to be allocated is obvious enough: There is an intrinsic conflict of interest. The referees, chosen because they are among the few experts in the author's field, are often competitors with the author for those same resources.

A referee who receives a proposal or manuscript to judge is being asked to do an unpaid professional service for the editor or project officer. The editor or officer thus has a responsibility to protect the referee, both by protecting the referee's anonymity and by making sure that the referee is never held to account for what is written in the report. Without complete confidence in that protection, referees cannot be expected to perform their task. Moreover, editors and project officers are never held to account for their choice of referees, and they can be confident that, should anybody ask, their referees will have the proper credentials to withstand scrutiny.

Referees would have to have high ethical standards to fail to take personal advantage of



their privileged anonymity and to make peer review function properly in spite of these conditions. Undoubtedly, most referees in most circumstances do manage to accomplish that. However, the fact is that many referees have themselves been victims of unfair reviews and this must sometimes influence their ability to judge competing proposals or papers fairly. Thus the institution of peer review seems to be suffering genuine distress.

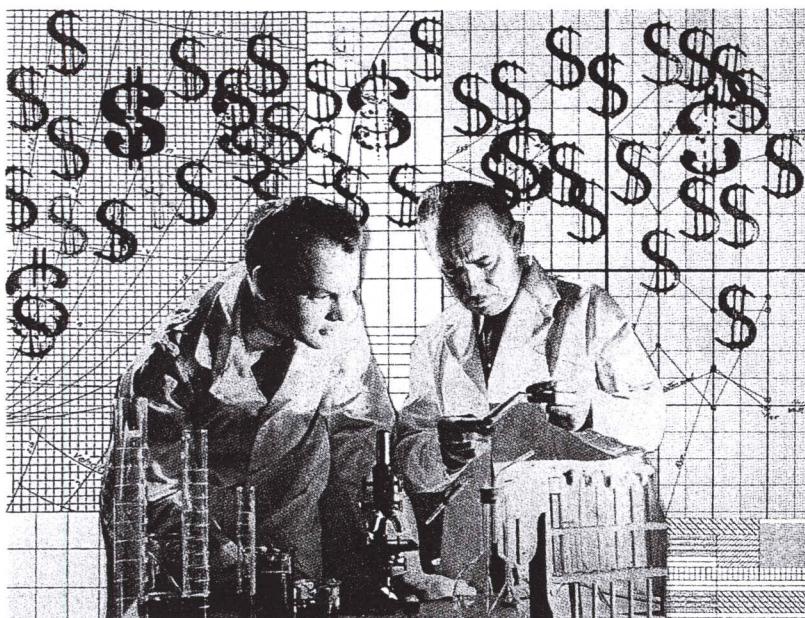
Once again, this analysis shows that science is a complex enterprise that must be understood in some detail before ethical principles can be formulated to help guide it.

Conclusions

We have put forth arguments in this article that indicate why each of the principles listed above may be defective as a guide to the behavior of scientists. However, our repeated admonition that there are no universal rules of scientific conduct does not mean that it is impossible to recognize distinctive scientific misconduct. We would like to conclude with some thoughts on how scientific misconduct might be distinguished from other kinds of misconduct.

We propose that distinctively scientific forms of misconduct are those that require the expert judgment of a panel of scientists in order to be understood and assessed. Other forms of misconduct may take place in science, but they should not constitute scientific misconduct. For example, fabricating experimental data is scientific misconduct, but stealing scientific instruments is not. Similarly, misappropriation of scientific ideas is scientific misconduct, but plagiarism (copying someone else's words) is not. Stealing and plagiarism are serious misdeeds, but there are other well-established means for dealing with them, even

There are types of misrepresentation that are condoned and accepted in scientific publications, whereas other types are harshly condemned.



Fabrication or covert and unwarranted manipulation of data is an example of the kind of deceptive practice that cannot be tolerated because it undermines the mutual trust essential to the system of science.

when they are associated with science or committed by scientists. No special knowledge is required to recognize them.

On the other hand, only a panel of scientists can deal with matters such as data fabrication that require a detailed understanding of the nature of the experiments, the instruments used, accepted norms for presenting data and so on, to say nothing of the unique importance of experimental data in science. In a dispute over an allegation that a scientific idea has been misappropriated, the issues are likely to be so complex that it is difficult to imagine a lay judge or jury coming to understand the problem from testimony by expert witnesses or any other plausible means. Similarly, expert judgment will usually be required to determine whether an experimenter's procedures in selecting or discarding data constitute misconduct—the conventions governing this vary so much across different areas of science that judgments about what is reasonable will require a great deal of expert knowledge, rather than simply the application of some general rule that might be employed by nonscientists.

In the section on the logical structure of science, we drew a sharp distinction between advocacy, which is permitted or encouraged in science, and deception that is not open to public assessment, which is judged very harshly in science. Fabrication or covert and unwarranted manipulation of data is an example of the kind of deceptive practice that cannot be tolerated because it undermines the mutual trust essential to the system of science that we have described. Similarly, misappropriation of ideas undermines the reward system that helps motivate scientific progress. In both cases, a panel of scientists will be required to determine whether the deed occurred and, if so, whether it was done with intent to deceive or with reckless disregard for the

truth. Should these latter conditions be true, the act may be judged to be not merely scientific misconduct but, in fact, scientific fraud.

Acknowledgment

The authors wish to thank Kathy Cooke, Ph.D., for her valuable assistance in thinking through the problems discussed in this article.

Bibliography

- Cole, J., and S. Cole. 1972. The Ortega hypothesis. *Science* 178:368–375.
- Collins, H. 1974. The TEA-set: Tacit knowledge and scientific networks. *Science Studies* 4:165–186.
- Collins, H. 1975. The seven sexes: A study in the sociology of a phenomenon, or the replication of experiments in physics. *Sociology* 9:205–224.
- Duhem, P. 1962. *The Aim and Structure of Physical Theory*. New York: Atheneum.
- Feyerabend, P. 1975. *Against Method: Outline of an Anarchistic Theory of Knowledge*. London: New Left Books.
- Feynman, R. 1985. *Surely You're Joking, Mr. Feynman*. New York: W. W. Norton.
- Galison, P. 1987. *How Experiments End*. Chicago: University of Chicago Press.
- Hadley, S. "Can Science Survive Scientific Integrity?" Paper presented on October 17, 1991, at the University of California, San Diego. Unpublished.
- Kitcher, P. 1990. The division of cognitive labor. *Journal of Philosophy* 87:5–22.
- Koertge, N. 1991. The function of credit in Hull's evolutionary model of science. *Proceedings of the Philosophy of Science Association* 2:237–244.
- Kuhn, T. 1970. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Lakatos, I. 1974. Falsification and the methodology of scientific research programmes. In *Criticism and the Growth of Knowledge*, ed. I. Lakatos and A. Musgrave. Cambridge, U.K.: Cambridge University Press. Pp. 91–196.
- Medawar, P. 1963. Is the scientific paper a fraud? *The Listener* (12 September) 70:377–378.
- Merton, R. 1968. The Matthew Effect in science. *Science* 159:56–63.
- Ortega y Gasset, J. 1932. *The Revolt of the Masses*. New York: W. W. Norton.
- Pinch, T. 1986. *Confronting Nature*. Dordrecht, Holland: P. Reidel.
- Popper, K. 1968. *The Logic of Scientific Discovery*. New York: Harper and Row.
- Popper, K. 1969. *Conjectures and Refutations*. London: Routledge and Kegan Paul.
- Quine, W. V. O. 1961. *From a Logical Point of View*. New York: Harper and Row.



*Originally published in *American Scientist* 84(5); 468-478.

EXPLORING ANIMAL BEHAVIOR

Readings from *American Scientist* THIRD EDITION

EDITED BY

Paul W. Sherman
Cornell University

John Alcock
Arizona State University



Sinauer Associates, Inc. Publishers
Sunderland, Massachusetts

2001