

Falsification in Social Science Method and Theory



Bryan Benham

University of Utah, Salt Lake City, Utah, USA

Charles Shimp

University of Utah, Salt Lake City, Utah, USA

Glossary

Bayesian inference A use of Bayes' theorem relating conditional and unconditional probabilities, by which it has sometimes been hoped to interrelate objective and subjective probabilities.

falsification Any method by which scientific claims are evaluated by empirical disconfirmation.

induction Any method for confirming general scientific laws or theories by appealing to the accumulation of specific experimental observations.

methodological pluralism The position according to which diverse methods, perhaps having different ontological implications, may nevertheless be legitimate means of discovering scientific truths.

positivism A philosophy of science according to which scientific observation is independent of theoretical commitments and science does and should rest on a secure empirical foundation.

Quine–Duhem thesis The position according to which the meaning or truth of statements cannot be determined individually, but only holistically.

scientific method Any means by which scientific truths are reliably obtained and scientific claims are evaluated.

underdetermination The condition in which two or more mutually inconsistent theories can equally account for the same data and for which evidence alone does not determine theory selection.

Falsification originated in the problem of determining scientific truth. Answers based on induction and on the confirmation of hypotheses were discovered to have theoretical and practical problems. Falsification was proposed as an alternative answer, in terms of which science

can rest on a secure foundation, provided that its hypotheses and theories are required to be subject to possible empirical falsification. Contemporary views of falsification vary greatly. In the philosophy of science, it is seen chiefly in logical and historical terms. In introductory social science texts on methodology, it is often seen as a way to reduce scientists' self-imposed biases that reflect theoretical preferences. Accordingly, in texts on rigorous, quantitative methodology, falsification tends to be praised either explicitly or implicitly as an essential part of the scientific method. Alternatively, in texts on qualitative research methods, it is sometimes described as a largely unhelpful or even counterproductive method. In social science practice, there is often an appeal to informal falsification, along with parsimony, as key evaluative criteria by which theories and empirical claims about truth can be evaluated. In theories of choice, decision making, and rational judgment, humans are seen as hypothesis-testing organisms who, in some contexts, may use an intuitive form of falsification. In mathematical statistics, many alternatives have been developed in response to the somewhat counterintuitive, negative logic of falsification. Similarly, the philosophical context from which falsification is viewed has broadened, so that to some scholars, positivism, within which falsification developed, is often seen as only a historical phase in the development of the philosophy of science. To other scholars, the scientific method is seen as successfully applying to the physical sciences but not to the social sciences. Most radically, some scholars see the scientific method within which falsification is defined as an unworkable method and in any case as a political tool. This constructionist and historicist view is compatible with the increasing use of narratives and other qualitative methods in sociological,

anthropological, and psychological research methods. How the tension between this approach and the traditional scientific method will be resolved remains to be seen.

Introduction and Overview

Falsification is often viewed as central to the scientific method, essential for correct experimental design, and thus required for scientific progress. It is a view strongly associated with Karl Popper's argument that a genuine scientific claim must be subject to a critical empirical test. According to this view, science ideally proceeds by formulating hypotheses that are subject to empirical testing in such a way that the results could, in principle, show the hypotheses to be false. The critical empirical test should have two conceivable outcomes, one that is compatible with the hypothesis, and one that disconfirms it. Hypotheses that are not falsifiable in principle may be seen, according to this view, as having some potential heuristic value, but having no factual scientific status. Such unfalsifiable hypotheses belong merely to prescientific or pseudoscientific endeavors. From this perspective, falsification is designed as a tool to ensure that scientific claims are grounded in empirical observation, rather than in entrenched theoretical or ideological commitments. Falsification is an attempt to overcome various methodological problems in earlier views of science that were based in induction. These problems include worries about the effectiveness of induction, deriving from the underdetermination of theory by confirming evidence. Falsification, however, faces its own set of problems. Among these is that falsification does not accurately describe the historical practice of scientific discovery or justification. Neither can it account for how evidence can, in practice, falsify any hypothesis at all, given that any hypothesis can always be saved by ad hoc hypotheses, or even by the outright rejection of seemingly falsifying evidence. Perhaps a more sophisticated version of falsification can avoid some or all of these problems; on purely logical grounds, however, falsification seems no better than induction as a means for promoting scientific progress.

In the social sciences, falsification has both defenders and detractors. Proponents accept falsification as an essential part of the scientific method. This is sometimes explicitly stated in introductory social science texts, and it is often implicit in experimental design and in the context of hypothesis testing in inferential statistics, where "negative logic" may be used to test a null hypothesis according to which a variable is hypothesized to have no effect. Opponents, on the other hand, point to the methodological and conceptual problems mentioned previously. These problems with falsification have

motivated the development of many alternative methodologies, some of them explicitly designed to preserve some of the logic of induction. In addition, some detractors argue that applying falsification to the social sciences is fundamentally illegitimate because it relies more on the quantitative methods and logic of the physical sciences, whereas its requirements are not met by the social sciences that do or should rely on qualitative and interpretive methods. Despite these criticisms, falsification retains a substantial presence in social science method and theory. This persistence of falsification in the social sciences raises a number of questions regarding the sociology of science as well as the possibility of a legitimate methodological pluralism in the social sciences.

Origins of Falsification

Induction and Scientific Method

Falsification originated in part as a response to various problems raised by earlier views of scientific methods based on induction. Induction has the intuitively desirable property that data that agree with a scientific hypothesis support or confirm that hypothesis. Few claims about scientific method seem more natural to a beginning science student. For example, suppose it is predicted that it will rain if the temperature falls below the dew point. To many students, it seems reasonable that there is confirmation of the prediction when the temperature falls below the dew point and it indeed starts to rain. From the perspective of induction, science progresses by the formulation of a hypothesis and the collection of sufficient empirical data that agree with the hypothesis. When accumulated evidence consistent with the hypothesis is sufficient, the hypothesis is confirmed. Although this account is oversimplified, insofar as induction exists in sophisticated forms, the goal of induction may be simply stated as the confirmation of scientific claims by the accumulation of data that agree with those claims.

Problems with Induction

Both theoretical and practical problems have been identified with induction. Consider the problem that arises if scientific method were based on induction: A universal empirical claim such as "All crows are black," or "Rate of information processing is a linear function of uncertainty," could never be conclusively verified, because it is unclear what finite amount of evidence could ever confirm its truth. Similarly, a probabilistic empirical claim faces the same problem. No finite amount of evidence could conclusively confirm the objective accuracy of a claim such as, "Most crows are black," or "Rate of information processing varies randomly but is on the

average a linear function of uncertainty.” In social science practice, of course, statements such as these are made in the context of additional assumptions that specify experimental and observational requirements. Nevertheless, the problem remains the same. But more damaging is that induction in general requires inferences from a sample of evidence. As Hume observed in 1739, the only reason to accept an inductive generalization as reliable is based on accepting a principle of reasoning—the future will be like the past, similar causes have similar effects, etc.—which is inherently either not subject to empirical confirmation or employs an inductive method to justify itself. The justification for induction therefore is subject to vicious circularity. Thus, if science rests on induction, it would appear that science involves a fundamental irrationality. Even if this skepticism about induction could somehow be satisfactorily overcome, there would remain the problem of underdetermination. Evidence underdetermines hypotheses in the sense that hypotheses or theories established by an accumulated amount of inductive support, aside from not being conclusively verified, are not uniquely confirmed. For any finite set of data, there is no unique hypothesis that is consistent with that data. It is always possible that two or more theories or hypotheses, each logically inconsistent with the others, will be equally consistent with the supporting data. That is to say, inductive support by data, by itself, never logically suffices to permit a scientist to choose between competing theories that are equally consistent with the data. Other criteria, including parsimony, explanatory power, esthetic qualities, perceived heuristic power, perceived social utility, and political and social values, are needed to supplement an inductive method of choosing between theories. But these additional evaluative criteria are obviously based on exactly the kinds of arbitrary, personal, subjective, potentially irrational human values science conventionally hopes to avoid.

Falsification as a Solution to the Problems of Induction

Falsification aims to overcome these problems with induction. According to falsification, the hallmark of scientific methodology is not that it uses observation or empirical evidence to verify or confirm its hypotheses. After all, many “nonscientific” practices, e.g., astrology, also employ this strategy. Rather, according to falsificationists, what makes science unique is that its claims are open to empirical falsification. What makes a generalization such as “All crows are black” genuinely scientific is not that there is a huge amount of observational evidence in its support, but that we know what type of evidence would count decisively against it; namely, the observation of only one nonblack crow would falsify it. Although such

universal claims may never be conclusively verified, they can, in principle, be conclusively falsified. According to this view then, empirical evidence should be used to test scientific hypotheses, not to confirm them. More importantly, falsification is possible without the use of induction, thus avoiding the skeptical problems just described, and to some extent the problems associated with the underdetermination of theory by evidence. In this way, falsification is seen as an improvement over induction. Accordingly, falsification has become a widely accepted feature of accepted scientific method.

Problems with Falsification

Falsification, however, has its own serious logical and methodological problems. First, falsification as described by Popper does not appear to describe accurately historical scientific practice, either of scientific discovery or theory justification. The history of science reveals cases when observational data seemingly incompatible with a hypothesis did not lead to the abandonment of that hypothesis, and for good reason. Copernican theory, for example, initially failed to predict the trajectories of falling terrestrial bodies or why objects do not fly off the surface of the spinning Earth. (Early critics argued, for example, that if Earth spins, as Copernican theory claims, then objects dropped from a tower should fall some distance from the tower base, but this is not what is observed.) Such potentially falsifying observations did not, however, lead to the rejection of Copernican theory. Contemporary social science reveals many other such examples (see later). Second, it is not obvious either in principle or in practice how to determine what counts as a falsifying observation for a specified hypothesis. In the philosophy of science, this problem is described by the Quine–Duhem thesis, according to which hypotheses are not tested in isolation from background theory. That is to say, hypotheses are always tested in clusters consisting of the explicitly tested hypothesis and a complex background theory, which in turn consists of definitions, concepts, conventional methods, instrumentation, and so on. The result of this type of holistic view of hypothesis testing is that an instance of falsifying evidence does not necessarily falsify the hypothesis explicitly being tested. Falsifying evidence, or evidence seemingly incompatible with the tested hypothesis, can only demonstrate that the cluster of claims (the hypothesis plus background theory, etc.) used in the experiment is, at worst, inconsistent. The evidence does not indicate which particular claim to discard. Thus, it is relatively easy to save a hypothesis by rejecting another claim in a background theory, or to propose auxiliary hypotheses not in the original set that would explain the inconsistency. Thus, deciding when to abandon a hypothesis, or when to choose one theory over another, is not merely a matter of seeing falsifying evidence.

Instead, other criteria are needed. Some candidates include considerations of parsimony (e.g., Occam's razor), consistency with already accepted theory, productivity, and predictive power, each of which has its own methodological problems. More insidious problems arise for theory selection if scientific method is viewed from a historicist or sociological view, according to which science involves metatheoretical, arbitrary, possibly irrational, and ideological or political features. From this view, falsification does not provide a methodology from which to escape these idiosyncratic features of scientific research methodology.

Falsification in Contemporary Social Science Research

The Persistence of Falsification

Despite the problems with falsification, it retains a prominent position in social science. One reason why falsification, in one form or another, continues to be used may be that it so effectively captures a widespread view of science: Virtually everyone seems to agree that, in some sense, empirical scientific claims should be subject to refutation by empirical observation. In theory evaluation, in research methods, and especially in statistical methods, falsification is often held up as an essential component of scientific method. Falsification remains prominent even in cases in which it is the object of criticism, in part for reasons already described and in part because some assert that it does not apply constructively to qualitative and interpretive methods that are increasingly employed in the social sciences. The proper place for falsification in social science research is therefore still a matter of some debate.

Diverse Forms of Usage

There are many ways in which falsification is currently used, among which are the following five. First, to gain some small empirical perspective on the extent to which social science texts agree that falsification has a useful role, the present authors informally reviewed various introductory texts on scientific method in psychology and sociology (including statistical methodology). Falsification appears in several guises. (Excluded from consideration are historically related meanings, such as when one scientist criticizes another for having falsified or misrepresented something.) According to most of the introductory texts reviewed, most prominently in psychology texts, science involves a method that relies on observation, empirical data, and hypothesis testing. Some of these texts refer to a self-correcting process of peer review in which, sooner or later, false claims will be laid to rest. Some

texts explicitly state that science does not address basic moral or religious issues because such issues involve claims that are not testable or falsifiable. Sociology texts were found somewhat more often than psychology texts to advocate various qualitative methods not involving rigorous hypothesis testing or falsification. Second, statistical methods often explicitly rely on falsification when, for example, the sampling distribution of a statistic of theoretical interest is derived to permit a null hypothesis to be tested. Such a test is motivated by a commitment to falsification, namely, that empirical data will show the hypothesis is falsified, i.e., is unacceptably improbable. If the null hypothesis is falsified in this sense, it is widely believed that scientific progress has been made. Third, falsification appears in its most basic form when a scientist concludes (perhaps most often in the case involving a theory contrary to the scientist's own theoretical predisposition) that the outcome of a single experiment falsifies an opponent's theoretical position. In accordance with the problems with falsification described previously, the theorist whose theoretical position Nevin ostensibly falsified was moved only to generalize and strengthen it, not to reject it. Fourth, on a grander scale, falsification appears in the idea of a "critical experiment" or "critical test." This idea was central to classic contests in the 1930s and 1940s between behavioral and Gestalt learning theorists. Hull and Tolman conducted numerous experiments, each designed to show that the competing theory was wrong. Again, in accordance with one of the problems with falsification described previously, neither side seemed much fazed by this series of critical tests wherein each side successively claimed their opponent's position had been falsified. More recently, the "cognitive revolution" is often said to have shown a basic inadequacy of behaviorism, and to have led to the death of behaviorism. Although cognitive psychologists appear not to claim explicitly that behaviorism was "falsified," the practical consequences seem similar: Behaviorism is declared dead, presumably on the basis of empirical results. As in the case of Hull and Tolman, many behaviorists continue to live happily, however, with their commitment to behaviorism, and care little that cognitive psychologists believe behaviorism is dead. Fifth, falsification as part of the logic of hypothesis testing appears in theories of human and nonhuman animal naturalistic decision making. That is, humans and animals have been studied as intuitive statisticians, especially in terms of the rationality and optimality of their decisions.

Alternative Methods

Not all social scientists are willing to accept falsification as an essential part of scientific method, for reasons previously described. Others reject falsification and associated views about rigorous hypothesis testing

because they believe those views derive from methods appropriate to the physical sciences, but not necessarily to the social sciences. These scientists believe that the subject matter of the physical sciences allows for falsification only because of the quantitative and observational methods available to them. These methods, they argue, are not available to social scientists, because human behavior can be adequately understood only by using hermeneutic methods. They therefore worry that transferring the commitments of falsification to the social sciences may be illegitimate and ultimately damaging to the social sciences. These and other concerns have motivated the development of alternative methods and design criteria that avoid some of the problems raised by falsification. Some of these alternative methods are designed to restore a more confirmatory and inductionist approach to statistical analyses. These methods include techniques such as confidence intervals and other forms of Bayesian inference, and the development and evaluation of mathematical and computer simulation theories, some of which have been developed without appeal to falsification. In addition, case studies, narratives, and other qualitative methods are often not committed to falsification or hypothesis testing. Falsificationists tend to worry that some of these alternative methods merely raise again the problems falsification was designed to avoid.

The Broader Impact of Falsification

Philosophy and Sociology of Science

Falsification is often seen as coming out of a positivist view of science, according to which progress is real and cumulative, and which employs methods that are empirical and objective, and which has a self-correcting process in the form of peer review. The critique of falsificationism in the philosophy of science has focused much-needed attention on metatheoretical issues in the practice of science, rather than on idealized logical reconstructions of scientific method. The end result is a richer appreciation for the historical and sociological features of scientific theorizing. Although falsificationism is rejected by philosophers of science as an adequate account of scientific method, it motivates important questions in the sociology and rationality of science.

Falsification from the Perspective of Evolutionary Theory

Falsification has also raised questions about the extent to which hypothesis testing is a natural or default form of decision making. If it is not, then it is possible that humans, including scientists, may find it difficult to behave

in accordance with the rational logic of scientific hypothesis testing. For example, Tversky and Kahneman have argued, especially in terms of base-rate neglect, that human categorical judgment displays large-scale irrationality. In another series of experiments to determine whether humans are naturally Popperian falsificationists, Wason presented subjects with a selection task aimed at testing an abstract rule. The result was that participants demonstrated a strong bias toward selecting potentially confirming evidence, and not equally available potentially disconfirming evidence. However, others have argued that rationality in these types of experiments must be interpreted in terms of cognitive processes evolution has provided, not in terms of the logic of falsificationist hypothesis testing. Thus, Tooby and Cosmides showed that participants in the Wason selection task actually did use something like falsificationist logic when the rule to be tested was placed in a more naturalistic setting, involving detecting cheaters in a social setting, not as a form of abstract rule testing. These types of experiments are provocative in what they can tell us about the standards of scientific methodology and research design.

Methodological Pluralism

The acceptance of falsification has been so pervasive in some circles that it has caused concern among researchers, who have felt its weaknesses have been underestimated and that domination by a single method is unhealthy. Alternative methods less committed to falsification and more in line with the tradition of induction seem to be finding growing acceptance in some parts of social science. The increasing use of narrative and other qualitative methods in sociological, anthropological, and psychological research raises controversial questions about the feasibility and appropriateness of methodological pluralism. Legitimate scientific method may not be a unified system relying heavily on falsification in the context of hypothesis testing, but may involve a great diversity of methods that may or may not be consistent with one another. How the tension between these competing approaches and the traditional view of scientific method will be resolved remains to be seen.

Acknowledgment

The authors thank Dr. Frank Drews for helpful references to naturalistic decision making as hypothesis testing.

See Also the Following Articles

Bayesian Statistics • Deduction and Induction

Further Reading

- Cosmides, L., and Tooby, J. (1992). Cognitive adaptations for social exchange. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (J. H. Barkow, et al., eds.), pp. 163–228. Oxford University Press, Oxford, England.
- Duhem, P. (1906/1954). *The Aim and Structure of Physical Theory* (translated by P. P. Wiener). Princeton University Press, Princeton, New Jersey.
- Feyerabend, P. K. (1987). *Farewell to Reason*. Verso, London.
- Gigerenzer, G. (1996). The psychology of good judgment: Frequency formats and simple algorithms. *Med. Decis. Making* **16**, 273–280.
- Hanson, N. R. (1969). *Perception and Discovery*. Freeman, Cooper, San Francisco.
- Hume, D. (1739–1740/2000). *Treatise on Human Nature*. Dent, London.
- Kuhn, T. S. (2000). *The Road since Structure*. University of Chicago Press, Chicago.
- Leahey, T. H. (1992). *A History of Psychology*, 3rd Ed. Prentice Hall, Englewood Cliffs, New Jersey.
- Nevin, J. A. (1969). Interval reinforcement of choice behavior in discrete trials. *J. Exp. Anal. Behav.* **12**, 875–885.
- Pickering, A. (1995). *The Mangle of Practice*. University of Chicago Press, Chicago.
- Popper, K. (1935/1959). *The Logic of Scientific Discovery*. Basic Books, New York.
- Quine, W. V. (1951). Two dogmas of empiricism. *Philos. Rev.* **60**, 20–43.
- Shimp, C. P., and Hightower, F. A. (1990). Intuitive statistical inference: How pigeons categorize binomial samples. *Animal Learn. Behav.* **18**, 401–409.
- Stuart, A., Ord, K., and Arnold, S. (eds.) (1999). *Kendall's Advanced Theory of Statistics, Vol. 2A: Classical Inference and the Linear Model*, 6th Ed. Arnold, London.
- Taylor, C. (1994). Interpretation and the sciences of man. In *Readings in the Philosophy of Social Science* (M. Martin and L. C. McIntyre, eds.), pp. 181–211. MIT Press, Cambridge, Massachusetts.
- Tversky, A., and Kahneman, D. (eds.) (1982). *Judgement under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge.
- Zsombok, C. E., and Klein, G. (eds.) (1997). *Naturalistic Decision Making*. Lawrence Erlbaum, Mahwah, New Jersey.