RESEARCH · PHILOSOPHY AND PARADIGMS

Karl Popper and Falsifiability as a Criterion of Science

Karl Popper's argument that scientific theories must be open to refutation, how it changed thinking about evidence, the main criticisms it faced, and what it still offers researchers today.

ModQu research article · 2,713 words · 25 references (APA 7th edition)

Few ideas from the philosophy of science have travelled as far beyond philosophy as Karl Popper's principle of falsifiability. It is cited by scientists defending their methods, by courts deciding which expert testimony to admit, and by critics dismissing claims as unscientific. The core idea is simple: a theory is scientific only if some possible observation could show it to be false. Behind that simple statement lies an argument about how knowledge grows, why confirming evidence is weaker than it looks, and what separates science from other kinds of belief. This article sets out Popper's argument, the problems it was meant to solve, the criticisms it attracted and the lessons it still holds for practical research.

The problem of induction

Popper's starting point was a problem identified by the Scottish philosopher David Hume in the eighteenth century. Science seems to work by induction: observing many cases and generalising from them. Every swan we have seen is white, so we conclude that all swans are white. But no number of observations can logically guarantee a general law, because the next case might be different. The sun has risen every morning in recorded history, yet that record does not prove it will rise tomorrow. Inductive conclusions always go beyond their evidence (Chalmers, 2013).

Many philosophers of Popper's time, particularly the logical positivists of the Vienna Circle, tried to solve this problem by developing a logic of confirmation in which evidence could make a theory more probable. Popper took a different path. In *Logik der Forschung*, published in 1934 and translated into English as *The Logic of Scientific Discovery* (Popper, 1959), he argued that science does not need induction at all. It needs only deduction, applied in the opposite direction.

The asymmetry between confirmation and refutation

Popper's key observation was a logical asymmetry. No number of white swans can prove that all swans are white, but a single black swan can prove that the statement is false. A universal claim can never be verified by finite evidence, but it can be refuted by one counter-example. Science, Popper argued, should exploit this asymmetry. Instead of seeking evidence that supports a theory, researchers should derive predictions from it and try as hard as possible to show that those predictions fail.

On this view, theories are never proven. They are conjectures, bold guesses about how the world works, which survive only as long as they withstand attempts to refute them. Popper (1963) summarised the method in the title of a later book: *Conjectures and Refutations*. Knowledge grows not by piling up confirmations but by eliminating errors. A theory that has survived many severe tests is said to be corroborated, but corroboration is a report on past performance, not a guarantee of truth.

Falsifiability as a demarcation criterion

The asymmetry led Popper to a criterion for distinguishing science from non-science, which he called the problem of demarcation. A theory is scientific, he proposed, if and only if it is falsifiable: if it is logically possible to describe an observation that would contradict it.

Popper (1963) described how he arrived at this criterion in Vienna around 1919. He was struck by a contrast between Einstein's theory of gravitation and the psychological theories of Freud and Adler, and the historical theory of Marx, which were then fashionable. Admirers of the latter theories found confirmation everywhere: every case seemed to fit. Popper came to see this apparent strength as a weakness. A theory that could explain any possible behaviour, including opposite behaviours, told us nothing about what would actually happen.

Einstein's theory was different. It predicted that light from distant stars would be bent by a specific amount as it passed the sun, a prediction that could be checked during a solar eclipse. If the observed deflection had been absent, or close to the smaller value implied by Newtonian theory, Einstein's theory would have been in serious trouble. The 1919 eclipse expeditions reported a result in line with Einstein's figure (Dyson et al., 1920). What impressed Popper was not that the theory was confirmed, but that it had taken a genuine risk.



Figure 1. Claims differ in how much they risk. The more outcomes a claim forbids, the more it says and the easier it is to refute. Examples by ModQu, following Popper (1959).

Falsifiability comes in degrees. A theory that forbids more possible observations has more content, says more about the world and is easier to refute. Popper argued that scientists should prefer bold theories with high content over cautious ones that are hard to test, because bold theories offer more opportunities to learn from failure.

Criticisms of falsificationism

Popper's account was influential, but it faced serious objections, some of which he accepted and some of which he resisted.

No hypothesis is tested alone

The French physicist Pierre Duhem had argued at the start of the twentieth century that a physical hypothesis is never tested in isolation, because deriving a prediction requires many auxiliary assumptions about instruments, initial conditions and background theory (Duhem, 1954). Quine (1951) generalised the point: when a prediction fails, logic tells us only that something in the whole set of assumptions is wrong, not which part. A researcher can always save a favoured hypothesis by blaming an auxiliary assumption instead.

Popper recognised this and responded with a methodological rule rather than a logical one: scientists should not protect theories with ad hoc adjustments that reduce their testability. Critics noted that this turns falsification into a matter of judgement and convention rather than strict logic.

Scientists do not behave like falsificationists

Kuhn (1962) argued from the history of science that scientists working within an established paradigm do not abandon their theories when they meet contrary evidence. They treat anomalies as puzzles, assume that the fault lies with the experiment or the auxiliary assumptions, and continue. Newton's theory, for example, lived for decades with an unexplained anomaly in the orbit of Mercury. On Kuhn's account, this tenacity is not a failure of rationality but a condition of productive normal science; a theory abandoned at the first difficulty would never be developed far enough to show its strengths.

Research programmes rather than single theories

Lakatos (1970) tried to reconcile Popper's logic with Kuhn's history. He proposed that the unit of appraisal is not a single theory but a research programme: a hard core of central commitments, protected by a belt of auxiliary hypotheses that can be modified when predictions fail. A programme is progressive if its modifications lead to new predictions that are then confirmed, and degenerating if the modifications merely explain away failures after the fact. Rejection is reasonable when a degenerating programme faces a progressive rival.

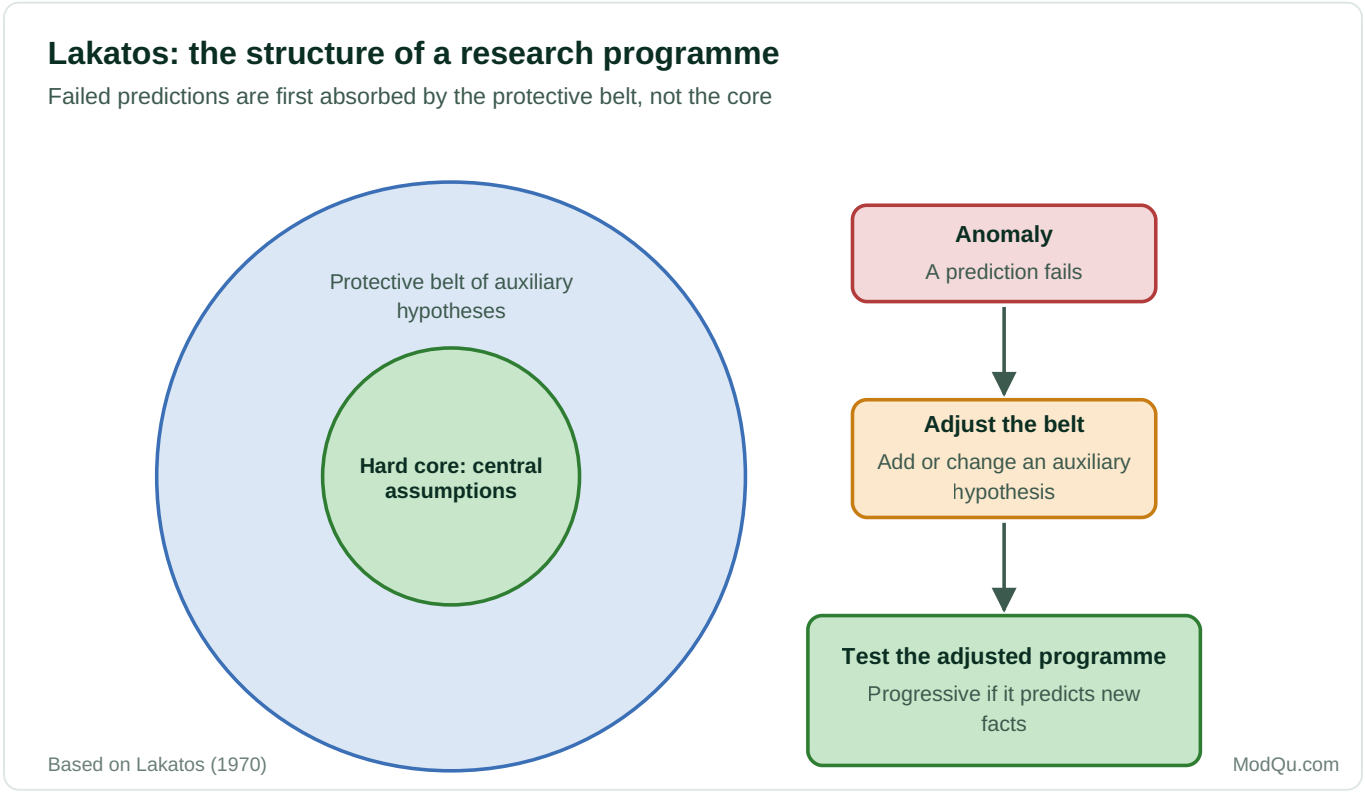


Figure 2. A Lakatosian research programme. Anomalies are absorbed by adjusting the protective belt; the programme is progressive only if the adjustments lead to new confirmed predictions. Based on Lakatos (1970); diagram by ModQu.

Against method

Feyerabend (1975) went further, arguing that no methodological rule, falsification included, had been followed consistently in the history of science, and that some of the greatest advances, such as Galileo's defence of Copernican astronomy, required breaking such rules. His provocative conclusion was that the only principle that does not inhibit progress is "anything goes". Few have accepted that conclusion, but Feyerabend's historical examples showed that the relationship between method and progress is more complex than any simple rule suggests.

Demarcation questioned

Some philosophers concluded that the search for a sharp line between science and non-science had failed. Laudan (1983) argued that no proposed criterion, including falsifiability, picks out all and only the activities we regard as scientific, and that the question should be replaced by asking whether particular claims are well supported. Others, including the contributors to a volume edited by Pigliucci and Boudry (2013), have argued that demarcation remains important in practice, for example when distinguishing medicine from quackery, even if it rests on several criteria rather than one.

Popper's treatment of psychoanalysis was also challenged. Grünbaum (1984) argued that some Freudian claims were in fact falsifiable, and that the real problem was that many had been tested and failed, rather than that they were untestable.

The Bayesian alternative

A further line of criticism came from philosophers who argued that Popper had dismissed confirmation too quickly. On a Bayesian view, evidence does raise or lower the probability of a hypothesis, and the size of the change depends on how surprising the evidence would be if the hypothesis were false. Howson and Urbach (2006) argued that this framework explains why some confirmations matter more than others, which Popper's account struggled to do. A successful risky prediction, such as the eclipse result, strongly supports a theory because it would have been very unlikely otherwise; a confirmation that was bound to occur adds almost nothing. Interestingly, this conclusion is close to Popper's own intuition about severe tests, expressed in the language of probability that he rejected. Much of the modern debate concerns which framework captures that intuition better, rather than whether the intuition is right.

Corroboration is not proof

Popper was careful to distinguish corroboration from confirmation. A corroborated theory has survived serious attempts to refute it, but that record gives no logical guarantee about the future. This raises an awkward question that critics pressed: if corroboration says nothing about future success, why should engineers rely on well-tested theories when building bridges? Popper's answer was that it is rational to act on the best-tested theory available, because it is the one whose errors we have tried hardest to find. Critics such as Chalmers (2013) have noted that this answer seems to reintroduce a form of inductive reasoning by the back door. The debate remains unresolved, but it shows that even Popper's critics accept the practical value of testing theories as severely as possible before relying on them.

Critical rationalism beyond science

Popper extended his account of knowledge into a general philosophy he called critical rationalism. Its central claim is that we cannot justify our beliefs with certainty, but we can subject them to criticism and learn from our mistakes. In *The Open Society and Its Enemies*, written during the Second World War, Popper (1945) applied this idea to politics. He argued that societies, like theories, should be open to criticism and capable of correcting their errors without violence, and he attacked philosophies of history that claimed to know the inevitable direction of social change. The same attitude, he believed, should govern institutions of all kinds: policies should be treated as experiments, their outcomes monitored, and their failures acknowledged rather than concealed.

For organisations, this has an obvious parallel. A business that treats its strategies as conjectures, measures their results honestly and abandons those that fail is applying critical rationalism, whether or not it uses the term. One that defends its strategies against all evidence, or measures only what is likely to look good, is not.

Falsification in modern research practice

Despite these criticisms, Popper's influence on research practice remains strong, sometimes in forms he would not have endorsed.

Null hypothesis testing

Statistical significance testing is often described as Popperian, because researchers try to reject a null hypothesis. Meehl (1967, 1978) argued that this is largely an illusion. In most social science studies, the null hypothesis is a claim of exactly zero effect, which is almost never true; with a large enough sample, it will always be rejected. Rejecting it therefore puts the researcher's substantive theory at little risk. In physics, by contrast, a theory predicts a specific value, and better measurement makes the test harder to pass. Meehl concluded that soft psychology had made slow progress partly because its tests were too weak to refute anything, and urged researchers to derive precise, risky predictions in the way Popper intended.

Mayo (1996) developed this idea into a philosophy of severe testing: a hypothesis is well supported by evidence only if the test would very probably have revealed it to be false if it were false. Gelman and Shalizi (2013) argued that good Bayesian practice also contains a falsificationist element, because models should be checked against data and rejected when they fit badly.

Publication bias and undead theories

If refutation drives progress, then failed predictions must be reported. In practice, journals have tended to publish confirmations and neglect null results. Ferguson and Heene (2012) argued that this has filled psychology with a "graveyard of undead theories" that have never been properly refuted because the studies that failed to support them were not published. Simmons et al. (2011) showed how flexible analysis choices allow almost any hypothesis to find support. Both problems defeat the Popperian ideal, because they make theories immune to evidence by controlling which evidence is seen.

Pre-registration, in which predictions and analysis plans are recorded before data collection, can be read as an attempt to restore genuine risk. Nosek et al. (2018) argue that it makes clear which findings were predicted and which were discovered after the fact.

Strong inference

Platt (1964) proposed a working method closely related to Popper's: devise alternative hypotheses, design experiments that would exclude one or more of them, and repeat. The value of an experiment, on this view, lies in what it rules out. This approach is as useful in applied research as in laboratory science. A team investigating why sales fell can list competing explanations and design analyses whose results would differ depending on which is true.

Falsifiability in the courtroom

Popper's criterion has also entered law. In *Daubert v. Merrell Dow Pharmaceuticals* (1993), the United States Supreme Court set out factors for judges to use when deciding whether expert scientific testimony is admissible. The first was whether the theory or technique can be, and has been, tested, and the Court cited Popper and Hempel (1966) in support. The decision brought a philosophical criterion into routine legal practice, although commentators have noted that judges apply it with varying degrees of understanding.

What falsifiability offers applied researchers

For a researcher designing a survey, a market study or an evaluation, Popper's ideas translate into practical habits.

State predictions before seeing the data. If a hypothesis is formulated after the results are known, the results cannot test it. Recording predictions in advance, even informally, keeps the line between testing and exploring clear.

Make predictions specific. "The campaign will improve brand perceptions" can be satisfied by almost any result. "Unaided awareness among 18 to 34-year-olds will rise by at least five percentage points within three months" can fail, and therefore tells us more when it succeeds.

Ask what would change your mind. Before a study, identify the results that would count against the expected conclusion. If no possible result would do so, the study is not testing anything.

Treat failures as information. A disconfirmed prediction is not a wasted study. It eliminates an explanation and directs attention elsewhere. Reporting failures, internally if not publicly, prevents organisations from repeating them.

Test against rivals, not against nothing. A finding that the new product concept scores better than "no product" says little. A finding that it scores better than the strongest competing concept, under conditions where it could have lost, says a great deal more. Choosing a demanding comparison is the applied equivalent of choosing a severe test.

Be wary of explanations that fit everything. A segmentation or customer model that explains every observed pattern after the fact, but predicts nothing in advance, deserves the scepticism Popper directed at the theories he encountered in Vienna.

Conclusion

Popper's principle of falsifiability did not settle the philosophy of science. Duhem and Quine showed that refutation is never purely logical, Kuhn and Lakatos showed that scientists rightly persist with theories in the face of anomalies, and Laudan questioned whether any single criterion can separate science from non-science. Yet the core of Popper's insight has survived its critics. Theories earn trust by risking failure, confirming evidence is cheap when it is sought selectively, and knowledge grows most quickly when errors are actively sought out. As Box (1976) put it in a statistical context, progress comes from a continual interplay between conjecture and evidence. For any researcher, the most useful question Popper left behind is a simple one: what result would show that I am wrong?

References

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*(356), 791–799. https://doi.org/10.1080/01621459.1976.10480949

Chalmers, A. F. (2013). *What is this thing called science?* (4th ed.). Hackett.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

Duhem, P. (1954). *The aim and structure of physical theory* (P. P. Wiener, Trans.). Princeton University Press. (Original work published 1906)

Dyson, F. W., Eddington, A. S., & Davidson, C. (1920). A determination of the deflection of light by the sun's gravitational field, from observations made at the total eclipse of May 29, 1919. *Philosophical Transactions of the Royal Society of London. Series A*, *220*, 291–333. https://doi.org/10.1098/rsta.1920.0009

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, *7*(6), 555–561. https://doi.org/10.1177/1745691612459059

Feyerabend, P. (1975). *Against method: Outline of an anarchistic theory of knowledge*. New Left Books.

Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 8–38. https://doi.org/10.1111/j.2044-8317.2011.02037.x

Grünbaum, A. (1984). *The foundations of psychoanalysis: A philosophical critique*. University of California Press.

Hempel, C. G. (1966). *Philosophy of natural science*. Prentice-Hall.

Howson, C., & Urbach, P. (2006). *Scientific reasoning: The Bayesian approach* (3rd ed.). Open Court.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge University Press.

Laudan, L. (1983). The demise of the demarcation problem. In R. S. Cohen & L. Laudan (Eds.), *Physics, philosophy and psychoanalysis* (pp. 111–127). D. Reidel. https://doi.org/10.1007/978-94-009-7055-7_6

Mayo, D. G. (1996). *Error and the growth of experimental knowledge*. University of Chicago Press.

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science, 34*(2), 103–115. https://doi.org/10.1086/288135

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*(4), 806–834. https://doi.org/10.1037/0022-006X.46.4.806

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Pigliucci, M., & Boudry, M. (Eds.). (2013). *Philosophy of pseudoscience: Reconsidering the demarcation problem*. University of Chicago Press.

Platt, J. R. (1964). Strong inference. *Science, 146*(3642), 347–353. https://doi.org/10.1126/science.146.3642.347

Popper, K. R. (1945). *The open society and its enemies*. Routledge.

Popper, K. R. (1959). *The logic of scientific discovery*. Hutchinson.

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. Routledge & Kegan Paul.

Quine, W. V. O. (1951). Two dogmas of empiricism. *The Philosophical Review, 60*(1), 20–43. https://doi.org/10.2307/2181906

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632