

Aaron Berney

PHIL 430

Professor James Woodbridge

May 8, 2026

Popper's Falsificationism and Shortcoming

Both Karl Popper (1902-1994) and Carl Hempel (1905-1997) were influential individuals in the philosophy of science who were concerned with the operational framework of testability in regards to scientific practice. However, they held sharply contrasting views on the structure of this framework. This divergence in their ways of thinking was especially illustrated in their ideas of what kind of evidential bearing observations can have on a hypothesis. Hempel in his 1966 work *Philosophy of Natural Science* develops an account of scientific testing which is explained as follows: Given a hypothesis H and a number of auxiliary assumptions A , you test H by deriving observational predictions I which are consequential to H (and A) being true—the test implications. The schema is given by a conditional, if H then I . Next, you develop an experiment or wait for a relevant observation in nature which would either align or misalign with I . If observation reveals I , then your hypothesis H and A are supported (not “true” otherwise it would be affirming the consequent). If observation doesn't reveal I to be true then H is refuted (Hempel, Ch.2 & 3 pg. 13 & 29).

Hempel gives a historical example of this scientific testing account through Hungarian physician Semmelweis's inquiry of childbed fever in Vienna General Hospital. Semmelweis observed that between two different wards of the hospital, the number of maternal fatalities were severely different. He wanted to figure out why. As a result, Semmelweis formed a series of conjectures about possible causes such as epidemic atmospheric influences, rough student examinations of patients, childbirth position, the presence of a priest with a bell, and cadaveric matter carried onto physicians from autopsies. He then tested each hypothesis by deriving test implications for each successive hypothesis and checking whether or not they were successful. Eventually the cadaveric matter hypothesis which predicted that washing hands in a

chlorinated lime substance after performing autopsies would decrease the fatality rate of the First Division (later including “putrid matter derived from living organisms”) yielded consequential observations which lent it support (Hempel, Ch.2 pg. 9-13). As acknowledged earlier, Hempel rejects the acceptance of the truth of a hypothesis after yielding positive results due to the fallacy of affirming the consequent.

However, Hempel still believes that positive evidence can still support a hypothesis through inductive reasoning, raising its credibility through successive instances of affirmatory evidence. In Hempel’s view, confirmation is a matter of degree and is strengthened in varying degrees through certain characteristics such as the variety of evidence, the narrowness of the test implications, and successful prediction of novel phenomena (Hempel Ch.4 pg. 40-46).

Popper takes the same stance on affirming the consequent though draws a far more radical conclusion. He insists that there is a deep asymmetry between confirmation and falsification: no amount of observed white swans can decisively verify the hypothesis of “All swans are white.” However, a single instance of a black swan conclusively refutes the hypothesis. Where Hempel would accept inductive support for the white swan hypothesis on the grounds of observing more and more white swans, Popper rejects this case and sees no grounds for believing in the truth or probableness of the swan hypothesis despite recurring white swans. Popper would only acknowledge the fallibility of the theory in the instance of a black swan—he can never accept positive instances as supportive, only that negative instances are decisive.

This is the central point where Popper’s view contrasts from Hempel’s and other’s generally accepted views of science. Most scientists accept the use of induction and the confirmatory framework described by Hempel, that allows for accumulated positive evidence to raise the credibility of a hypothesis. Instead, Popper rejects the use of any inductive logic and insists that science can function without using inductive reasoning. The reason Popper is so passionately against the use of induction is because he doesn’t see any logical path from which the results of a singular observation can be generalized to entire theories. As he puts it “I never assume that we can argue from the truth of singular

statements to the truth of theories. I never assume that by force of 'verified' conclusions, theories can be established as 'true', or even as merely 'probable "' (Popper, Essay 9 pg. 20). Popper reinforces his argument by drawing from an example in the history of science. It is commonly assumed, he notes, that "Newton's dynamics can be induced from Galileo's and Kepler's laws, and it has even been asserted that it can be strictly deduced from them" (Popper, Essay 12 pg. 52). However, Popper denies the logical coherence of this assertion. He instead says that "from a logical point of view, Newton's theory, strictly speaking, contradicts both Galileo's and Kepler's (although these latter theories can of course be obtained as approximations, once we have Newton's theory to work with). For this reason it is impossible to derive Newton's theory from either Galileo's or Kepler's or both, whether by deduction or induction. For neither a deductive nor an inductive inference can ever proceed from consistent premisses to a conclusion that formally contradicts the premisses from which we started. I regard this as a very strong argument against induction" (Popper, Essay 12 pg. 52). In the case of a successive theory, it stands contradictory to any predecessor's theories which means that no accumulation of positive evidence inductively confirming an old theory could ever have developed into Newton's theory.

In place of induction, Popper develops a fully deductive account of science which explains how science proceeds. Scientists propose bold hypotheses (the bolder the better because a bolder theory is fallible in more ways than a less bold one), derive risky test implications of their conjectures, and then try to refute them through empirical testing. Theories which fail their tests are falsified—eliminated from the pool of theories which have not yet been refuted—and those which survive are subject to further attempts to falsify them (Popper, Essays 8 & 9 pg. 3-6 & 20). Popper further explains his falsificationism's methodology by discussing how he views the problem of demarcation. He claims that the criterion, although it still leaves the distinction rough, is falsifiability. A theory can only be considered scientific if it makes claims which in theory can be shown to be false by an observation in the natural world (Popper, Essay 8 pg. 6). He contrasts examples like Einstein's theory of relativity being highly specific and predictive offering it many ways to be falsified vs. Freudian psychology which seems to have a

confirmatory explanation in wake of any observations made. In the case of scientific progress, Popper views it as a process of elimination. A new theory is accepted over a preceding one because it is not falsified by tests which falsified the predecessor and demonstrates greater boldness through riskier predictions (Popper Scientific progress is viewed by Popper in a polarizing way compared to the traditional view. Instead of accumulating truths, we shed false ideas which incrementally leaves us with less and less viable options for the truth. What survives rigorous empirical testing is used provisionally—for the time being—in scientific practice and instrumental applications until it is inevitably falsified and a successor takes its place. (Popper Essay 9, pg. 20).

Popper acknowledges the need for some account of describing how a theory has held up through repeated attempts at falsification. For this, he introduces the idea of corroboration. A theory can be compared to another theory based on its level of corroboration, which refers to how many attempts at falsification it has resisted (worth noting that the quality and distinctness of these attempts matter, not just the number of them). The degree of corroboration of a theory is established through its survival of genuine, sincere attempts to falsify it. The harder the test, and the more independently the theory has passed it, the higher its degree of corroboration. Corroboration serves as a purely backwards-looking evaluation of a theory, a record of a theory's performance through empirical rigor. It's distinct from confirmation in that the degree of corroboration of a theory has no weight to the truth or probableness of a theory whereas the degree of confirmation of a theory does imply a level of truth or probableness (Popper, Essay 12 pg. 59). We may rationally prefer a well-corroborated theory for the purposes of further testing and scientific discussion, but not because we have any positive grounds for thinking it is more likely to be true than its rivals. This notion of truth implication and preference is highly radical and causes some major issues which we will discuss in the following section.

Popper's ideology faces numerous challenges which must be taken into account in order to evaluate his falsificationism. The first of which is the Duhem-Quine problem which Popper acknowledges in his essays. No hypothesis is tested in isolation. Any test of a hypothesis H is also a test of background

assumptions, auxiliary hypotheses, and assumptions about instruments and conditions. When a theory sees an observation which is in misalignment with its predictions, we are not able to deductively conclude whether H is false or whether some auxiliary hypothesis is. Strict falsification is therefore rarely available, since scientists can always rescue H by revising the auxiliaries. Popper responds to this by prescribing some refutations of falsifying observations as ad hoc. He claims that a refutation of falsification based on auxiliaries is ad hoc if it introduces no independently testable consequences beyond the original problem it was brought in to solve (Popper, Essay 8 pg. 13). Another challenge was brought up by Hilary Putnam in his work *The 'Corroboration' of Theories* which argues that Popper's idea of corroboration is lacking a crucial mechanism which science needs to function in a practical sense. Take a scenario described by Godfrey-Smith in *Theory and Reality* where engineers are developing a bridge. In the development of the bridge the engineers are relying on Newtonian mechanics to construct it and they are betting on Newton's theories to work for the foreseeable future. These engineers are working with positive, confirmatory grounds of belief while corroboration offers no explanation for why these engineers should rely on any theory for a matter of fact. If corroboration tells us nothing about future performance or—in other words truth or probability of truth—then we are offered no explanation for why engineers should choose to rely on any given theory or why they should choose between a well-corroborated theory vs. a lowly-corroborated theory to construct their bridge (Godfrey-Smith, Ch. 4.4 pg. 82-83). If Popper suggests that we use degrees of corroboration to guide our theory reliance, then this reintroduces induction into the picture which he strictly rejects (Putnam, pg. 3). Another issue Popper faces is with strictly probabilistic claims. Claims about half-lives, decay rates, or population frequencies, cannot be conclusively falsified by any finite series of observations, since improbable outcomes do sometimes occur. Popper's falsification fails to account for the probabilistic predictions of many scientific theories (Godfrey-Smith Ch 4.4 pg. 81).

In my assessment of Popper's falsification structure and the challenges presented against it, I believe they are fatal to Popper's notion of falsification and how science should proceed. However,

Popper's developments on falsification were crucial to the development of science and falsification is a deeply important aspect of how scientific practice operates. Mentioned before, Popper responds to the Duhem-Quine problem with his notion of ad hoc. However it is very difficult to draw a line between what is considered ad hoc and what is considered legitimate when not looking at the conjecture from a retrospective position. What is considered ad hoc and legitimate is not shaped by logical practice alone, it is also fueled by auxiliary assumptions and inductive inferences about what should count as independently testable. With no clear distinction of what can be considered ad hoc and a reliance on inductive reasoning in that decision making process already makes me refute Popper's strictly deductivistic framework. This already can be considered by some to be fatal to Popper's falsificationism. In regards to the other challenges presented in the preceding section, Popper offers no adequate response to them without having to fall back on inductive reasoning. He has no *explanicum* for why we prefer and make decisions to rely on a selection of theories in a competitive theory landscape. Nor does he offer any way of accounting for probabilistic conjectures in a strictly deductive sense. These challenges even if not taken in accumulation are each fatal in their own capacity to a strict view of falsification. However, I do think Popper's falsificationism is partial success in the framing of it as a corrective and with his demarcation distinction. Popper rightly emphasizes the importance of risky predictions, of severe tests, and of a willingness to give up theories when faced with contrary evidence. The demarcation criterion is a useful diagnostic for telling serious scientific work from theories that immunize themselves against any possible refutation. And the logical asymmetry between universal generalizations and their counterexamples is real and worth taking seriously in any account of scientific testing. But as a complete account of scientific reasoning and methodology, it fails in a number of crucial ways if Popper is to fully reject any form of inductive reasoning.