
The Replication Crisis is Less of a “Crisis” in Lakatos' Philosophy of Science

Mark Rubin 
Durham University, UK

22nd December 2023

Citation: Rubin, M. (2023, December 22). The replication crisis is less of a “crisis” in Lakatos' philosophy of science. *MetaArXiv*. <https://doi.org/10.31222/osf.io/2dz9s>

Abstract

Popper's (1983, 2002) philosophy of science has enjoyed something of a renaissance in the wake of the replication crisis, offering a philosophical basis for the ensuing science reform movement. However, adherence to Popper's approach may also be at least partly responsible for the sense of “crisis” that has developed following multiple unexpected replication failures. In this article, I contrast Popper's approach with Lakatos' (1978) approach and a related approach called *naïve methodological falsificationism* (NMF; Lakatos, 1978). The Popperian approach is powerful because it is based on logical refutation, but its theories are noncausal and, therefore, lacking in scientific value. In contrast, the Lakatosian approach considers causal theories, but it concedes that these theories are not logically refutable. Finally, the NMF approach subjects Lakatosian causal theories to Popperian logical refutations. However, its approach of temporarily accepting a *ceteris paribus* clause during theory testing may be viewed as scientifically inappropriate, epistemically inconsistent, and “completely redundant” (Lakatos, 1978, p. 40). I conclude that the replication “crisis” makes the most sense in the context of the Popperian and NMF approaches because it is only in these two approaches that replication failures represent logical refutations of theories. In contrast, replication failures are less problematic in the Lakatosian approach because they do not logically refute theories. Indeed, in the Lakatosian approach, replication failures can be legitimately ignored or used to motivate theory development.

Keywords: metascience; Popper; philosophy of science; Lakatos; replication crisis; theory testing

Popper's philosophy of science has enjoyed something of a renaissance in the wake of the replication crisis. His approach is regarded as “useful for both understanding and remediating the replication crisis” (O'Donohue, 2021, p. 236), and the ensuing science reform program has been described as “distinctly Popperian” (Derksen, 2019, p. 460; see also Flis, 2019). Certainly, many aspects of Popper's work are useful in science in general and in relation to the replication crisis and science reform in particular. However, concerns about multiple unexpected replication failures may also be more relevant in the Popperian approach than in other approaches. From this perspective, adherence to the Popperian approach may have accentuated the sense of a replication

“crisis.” Accordingly, it is worth considering how other philosophies of science might characterise multiple unexpected replication failures.

In this article, I explore this issue by contrasting Popper's (1983, 2002) approach with Lakatos' (1978) approach and a related approach called *naïve methodological falsificationism* (Lakatos, 1978). I conclude that the replication crisis may be less of a “crisis” for Lakatosians.¹

Popper argued that theories are only scientific if they are logically falsifiable (e.g., Popper, 1983, pp. xix-xx), meaning that they have the potential to be falsified through a process of logical deduction. Lakatos (1978) disagreed. In a paper that Feyerabend (1975) described as “one of the most important achievements of twentieth-century philosophy,” Lakatos (1970, 1978) proposed that “exactly the most admired scientific theories simply fail to forbid any observable state of affairs” (Lakatos, 1978, p. 16, italics omitted). Popper (1974) responded that, “were the [Lakatosian] thesis true, then my philosophy of science would not only be completely mistaken, but would turn out to be completely uninteresting” (p. 1005). To understand the reasons for this disagreement, I begin by considering the different ways in which Popper and Lakatos conceptualized scientific theories.

What is a Scientific Theory?

A *strictly universal statement* or law such as “all swans are white” can serve as a scientific hypothesis (Popper, 2002, p. 38). This hypothesis can be used to deduce a “negative” prediction, such as “there will be no non-white swans at this time and place.” This “nonexistential proposition” is a “specialization of a universal law...to a particular space-time region”; (Popper, 1974, p. 998). However, there are no non-white swans in lots of space-time regions (Popper, 2002, p. 83), and “we cannot search the whole world in order to make sure that nothing exists which the law forbids” (Popper, 2002, p. 49). Consequently, the universal statement must be combined with the *initial conditions* of a specific space-time region in order to deduce a “potential falsifier” in that particular region. This potential falsifier takes the form of a “basic statement” (a singular existential statement that refers to a specific space-time region) describing an intersubjectively observable event such as “there is a black swan in this location at this time” (see also Popper, 1983, p. xx; Popper, 1974, p. 997; Popper, 2002, p. 38, p. 83). Acceptance of this potential falsifier during hypothesis testing will then logically refute the hypothesis that “all swans are white.”

Lakatos (1978) did not disagree with the above reasoning. However, he argued that the hypothesis “all swans are white” is not a scientific theory. According to him, “a proposition might be said to be scientific only if it aims at expressing a causal connection” (pp. 18-19). For example, the proposition “swanness causes whiteness” (Lakatos, 1978, p. 19) represents a scientific theory because it expresses a causal connection (e.g., swan DNA causes white plumage; Karawita et al., 2023).

Contrary to Lakatos (1978), Popper (2002, p. 39) believed that the “principle of causality” should be excluded from science. Hence, he would reject Lakatos' (1978, p. 19) proposal that the statement “swanness causes whiteness” represents a scientific theory. This is not to say that Popper ignored causal explanations. As he explained, “to give a causal explanation of an event means to deduce a statement which describes it, using as premises of the deduction one or more universal laws, together with certain singular statements, the initial conditions” (Popper, 2002, p. 38, italics omitted). Popper gave an example in which (a) the hypothesis is “whenever a thread is loaded with a weight exceeding that which characterizes the tensile strength of the thread, then it will break”; (b) the two initial conditions are “the weight characteristic for this thread is 1lb,” and “the weight put on this thread was 2lbs”; and (c) the (positive) prediction is “this thread will break” (Popper,

2002, p. 38). In this situation, the observation that “this thread broke” and the situation in which it occurred are the *explicandum* or “state of affairs to be explained,” and the theory and its deduced prediction in relation to the initial conditions represent the independently testable explanation or *explicans* (Popper, 1983, p. 132). In addition, the initial conditions describe the “cause,” and the prediction describes the “effect” (Popper, 2002, pp. 38-39).

Critically, however, and in contrast to Lakatos, Popperian hypotheses and theories are *noncausal* universal statements (“all swans are white”) rather than causal connections (“swanness causes whiteness”). As in the above example, initial conditions and predictions may be described as “causes” and “effects” respectively. However, Popper (2002) preferred to avoid these terms, and he was clear that no “principle of causality” should be invoked (p. 39). As he explained, “I shall be content simply to exclude it [the principle of causality], as ‘metaphysical’, from the sphere of science” (Popper, 2002, p. 39; see also Popper, 2002, p. 48). Hence, what is logically refuted in a Popperian theory test is a *noncausal* universal statement rather than a causal connection.²

Lakatos (1978) was concerned that, in the absence of a causal connection, a Popperian theory may be regarded as a “mere curiosity” or “oddity” without any obvious “scientific value” (pp. 18-19). *Why are all swans white?* A Popperian theory does not provide an explicit answer to this question. This situation was unsatisfactory for Lakatos, who argued that “science...must be demarcated from a curiosity shop where funny local - or cosmic - oddities are collected and displayed” (p. 18). Similarly, Pearce (1990) noted that Popper’s “all swans are white” example “has little relevance to science since scientific theories are not generalizations of facts; rather, they involve an understanding of the underlying processes that *cause* certain facts to occur” (p. 47, my emphasis). Lakatosian theories provide this scientific relevance in the form of causal connections: “all swans are white *because* swanness causes whiteness.”

In summary, for Popper, scientific theories must be logically falsifiable, whereas for Lakatos, they must be causal connections. In the Popperian approach, a logically falsifiable universal statement represents both a hypothesis and a theory (Popper, 2002, p. 4, pp. 37-38, p. 48; see also Hager, 2000, p. 5).³ Hence, a logical refutation of a hypothesis is also a logical refutation of a scientific theory. In contrast, the Lakatosian approach provides a conceptual distinction between noncausal hypotheses and causal theories: A noncausal hypothesis (e.g., “all swans are white”) is not a scientific theory because it does not provide a causal connection (e.g., “swanness causes whiteness”). Consequently, for Lakatos, the logical refutation of a hypothesis does not necessarily imply the logical refutation of its associated theory. Indeed, as I discuss next, Lakatos argued that causal theories are not logically refutable (for a similar view, see Putnam, 1991).

Causal Theories are Not Logically Refutable

There are many causes in the universe, including some that may confound and counteract the particular cause that we are investigating in our study (Johansson, 1980). For example, a genetic factor may cause a swan to be black even though it remains true that swan DNA causes white plumage. The existence and intervention of this counteracting cause would not logically refute our causal theory because it operates independently from our theorized cause. In addition, our theorised cause may be moderated by various factors. For example, swan DNA may only cause white plumage in some environments and not in others. Again, moderator factors do not refute the existence of putative causes; they merely limit their influence.

To acknowledge the potential impact of these confounding, counteracting, and moderating factors, we may attempt to delineate them within an exclusive *ceteris paribus* clause which states that various specified and unspecified causally-relevant factors do not affect our observations

during theory testing (see also Putnam, 1991, p. 137). However, this clause may be incorrect because other relevant factors may, in fact, affect our observations. Hence, we need to acknowledge that a test of a causal theory is also a test of a fallible *ceteris paribus* clause. For example, we don't just test the theory that "swanness causes whiteness"; we test a conjunction of this theory and a *ceteris paribus* clause: "swanness causes whiteness *provided that no other relevant factor is at work*." The observation of a black swan may then refute this proposition because either (a) swanness *does not* cause whiteness or (b) swanness *does* cause whiteness but some other relevant factor has intervened to produce a black swan. The upshot of all this is that, although a black swan can logically refute the noncausal hypothesis that "all swans are white," it cannot logically refute the causal theory that "swanness causes whiteness" because it may instead refute the *ceteris paribus* clause that "no other relevant factor is at work" (see also Putnam, 1991, p. 127).

Naïve Methodological Falsificationism

In summary, the Popperian approach is powerful because it is based on logical refutation, but its weakness is that its theories are noncausal and, therefore, lacking in scientific value (a "mere curiosity"; Lakatos, 1978, p. 19). In contrast, the Lakatosian approach is powerful because it tests causal theories, but its weakness is that its theories are not logically refutable.

Lakatos (1978) noted the possibility of a third approach to theory testing, which he described as *naïve methodological falsificationism* (NMF). From my perspective, NMF hybridizes the Popperian and Lakatosian approaches. It claims *both* the deductive power of Popper's logical refutations *and* the scientific relevance of Lakatos' causal theories. It does so by attempting to logically refute not only noncausal hypotheses ("all swans are white"), but also causal theories ("swanness causes whiteness"). Like the Popperian and Lakatosian approaches, the NMF approach may influence appraisals of replication failures. Hence, I explain how it operates, and I consider its weaknesses.

According to Lakatos (1978), the NMF approach circumvents the logical problems associated with testing causal theories by temporarily and tentatively accepting the *ceteris paribus* clause that "no other relevant factor is at work" during theory testing. In this case, the *ceteris paribus* clause is excluded from the test and the specific causal theory is left as the only remaining statement that can be logically refuted by an anomalous result. As Lakatos (1978) explained, "we may call an event described by a statement *A* an '*anomaly* in relation to a theory *T*' if *A* is a potential falsifier of the conjunction of *T* and a *ceteris paribus* clause but it becomes a potential falsifier of *T* itself after having decided to relegate the *ceteris paribus* clause into 'unproblematic background knowledge'" (p. 26). However, there are three related problems with the NMF approach of accepting *ceteris paribus* clauses as unproblematic (i.e., irrefutable) during theory testing.

(1) Accepting Ceteris Paribus Clauses is Scientifically Inappropriate

Accepting the *ceteris paribus* clause as temporarily "unproblematic" during a causal theory test changes the proposition under test from the logically irrefutable statement that "swanness causes whiteness *provided that no other relevant factor is at work*" to the logically refutable statement that "swanness causes whiteness *and no other relevant factor is at work*." One potential problem with this approach is that the proposition "no other relevant factor is at work" is unrealistic given the potentially infinite range of potential factors to which it refers. For example, even if it is true that "swanness causes whiteness," it is unreasonable to accept that no other factor in the

universe could cause a non-white swan. Consequently, we are left with a choice between (a) testing a proposition that cannot be logically refuted and (b) temporarily accepting a proposition that is unrealistic (for a discussion of a related dilemma, see Reutlinger et al., 2021, Section 4).

Of course, scientists often condition their tests on unrealistic counterfactual models on the assumption that “all models are wrong, but some are useful” (Box & Draper, 1987, p. 424; see also Popper, 2002, p. 72). So, the real issue is not whether researchers believe that their models are wrong or unrealistic, but whether they believe that their models are *appropriate for their purposes* (Rubin, 2020, p. 8). In the NMF approach, a proposition with an accepted *ceteris paribus* clause is regarded as appropriate because it fulfils the purpose of allowing a logical refutation of a causal theory. In contrast, in the Lakatosian approach, such a proposition is scientifically inappropriate because it prevents a consideration of whether other relevant factors have influenced the test result. To be clear, Lakatosians may make the unrealistic assumption that no other relevant factor is operating when they test a causal theory. However, unlike, NMF researchers, they never *accept* this assumption as *irrefutable* during their test. They are *always* open to the possibility that other causes have affected their test result.

(2) Accepting Ceteris Paribus Clauses is Epistemically Inconsistent

The NMF decision to accept the *ceteris paribus* clause as irrefutable, even on a tentative and temporary basis, is inconsistent with a scientist's epistemic obligation to specify their doubt and ignorance about the potential influence of other relevant factors in their investigations (e.g., Feynman, 1955; Firestein, 2012; Merton, 1987; Rubin, 2023b, pp. 20-21). The *ceteris paribus* clause represents this doubt and ignorance. It is where scientists acknowledge both their “known unknowns” (what they know they don't know – their “specified ignorance”; Merton, 1987) and their “unknown unknowns” (what they don't know they don't know – their unspecified ignorance; Rubin, 2023a). Consequently, accepting a *ceteris paribus* clause as “unproblematic” during a causal theory test flies in the face of scientific humility.

An NMF researcher's decision to temporarily accept the *ceteris paribus* clause is also inconsistent with (a) their future research activities and (b) their colleagues' ongoing research activities. How can a scientist accept that no other relevant factor is influential during their theory test and then go on to test the influence of some of those factors in their future work? Similarly, how can they accept a *ceteris paribus* clause as “unproblematic” when, all around them, their colleagues are busily investigating the influence of the factors it contains? As Meehl (1990, p. 111) explained,

for the *ceteris paribus* clause to be literally acceptable in most psychological research, one would have to make the absurd claim that whatever domain of theory is being studied (say, personality dynamics), all other domains have been thoroughly researched, and all the theoretical entities having causal efficacy on anything being manipulated or observed have been fully worked out! If that were the case, why are all those other psychologists still busy studying perception, learning, psycholinguistics, and so forth?⁴

These various inconsistencies may be dismissed by arguing that researchers temporarily abandon the role of “scientist” and instead adopt the role of a “quality controller” who accepts the background knowledge of their test and automatically (logically) refutes products (theories) that do not meet the stated criteria (Rubin, 2020). However, this role-switching account does not resolve the problem of “epistemic inconsistency” (Rubin, 2020, p. 7): A logical refutation that is obtained in the role of quality controller becomes an illogical refutation when the quality controller

returns to the role of scientist and begins, once again, to doubt the validity of the *ceteris paribus* clause.

(3) Accepting Ceteris Paribus Clauses is “Completely Redundant”

NMF researchers might argue that their acceptance of the *ceteris paribus* clause is only tentative and temporary, and that they will bring the clause back into question after their theory test. This position is consistent with Popper (1974, p. 1009), who argued that the *logical refutation* of a theory does not necessarily imply that researchers should subsequently *reject* the theory in practice and stop working on it. According to Popper, the *rejection* of a theory “will depend among other things, on what alternative theories are available” (p. 1009). But if this is the case, then what is the function of logical refutations during theory testing? Why should we temporarily and tentatively “accept” alternative theories as part of an unproblematic *ceteris paribus* clause in order to force a logical refutation if we are only going to bring these alternative theories back into consideration when deciding whether to reject (stop working on) that theory? Instead, why not consider the logical refutation of noncausal *hypotheses* (e.g., “all swans are white”) in the context of explanations provided by *both* fallible causal theories (“swanness causes whiteness”) *and* alternative causal theories within the fallible *ceteris paribus* clause (“some other relevant factor is at work”) and then come to a tentative conclusion in a process of inference to the best explanation (e.g., Haig, 2009)? Lakatos (1978) had a similar view, describing the NMF decision to temporarily accept the *ceteris paribus* clause during theory testing as “completely redundant” (p. 40).

Summary

The NMF approach represents a powerful hybrid of the Popperian and Lakatosian approaches because it results in the logical refutation of causal theories. However, the NMF decision to accept the *ceteris paribus* clause as temporarily irrefutable is problematic for three reasons. First, it is scientifically inappropriate because it prevents a consideration of other relevant factors as having a potential influence on the test result. Second, it results in epistemic inconsistency because researchers accept propositions during theory testing that they subsequently doubt. Third, it is “completely redundant” (Lakatos, 1978, p. 40) because the logical refutation of a theory does not necessarily lead to its rejection in practice.

Implications for the Replication Crisis

The three approaches that I have considered each have their strengths and weaknesses. The Popperian approach can logically refute noncausal theories. However, the scientific value of these theories is unclear. The Lakatosian approach considers causal theories. However, their causal nature makes them logically irrefutable. Finally, the NMF approach aims for the best of both worlds by logically refuting causal theories. However, it does so by temporarily accepting *ceteris paribus* clauses, and this approach may be characterised as scientifically inappropriate, epistemically inconsistent, and completely redundant. Table 1 provides a summary of these three approaches.

According to my analysis, failed replications should be more impactful in the Popperian and NMF approaches than in the Lakatosian approach because it is only in the former two approaches that failed replications logically refute theories.⁵ Consequently, the revelation of multiple unexpected replication failures should be more concerning in the Popperian and NMF approaches because unexpected refutations threaten the integrity of the theory testing process.

Table 1
Comparison of the Popperian, Lakatosian, and NMF Approaches

Approach	Popperian	Lakatosian	NMF
Type of theory	Noncausal	Causal	Causal
Example theory	“All swans are white”	“Swanness causes whiteness”	“Swanness causes whiteness”
Accept ceteris paribus clause during testing?	N/A*	No	Yes
Logical refutation of theory?	Yes	No	Yes

*Popper (1974) explained that “no ceteris paribus clause is necessary” (Footnote 75, pp. 1186-1187).

More specifically, they imply that inadequate theories have slipped through the refutation net due to problems with the theory testing process, such as nonsevere tests, poor methodology, and publication bias. From this perspective, the appropriate response is to improve the theory testing process by, for example, tightening up the deductive derivation chain from theory to prediction, using more severe tests, using more rigorous methodology, conducting more direct replications, and reducing publication bias.

In contrast, in the Lakatosian approach, replication failures are less impactful because they logically refute only noncausal hypotheses; not causal theories. Of course, it remains important to use high quality methodology when undertaking hypothesis tests. However, in the Lakatosian approach, a logically refuted hypothesis does not imply a logically refuted theory. Consequently, replication failures should have a less destructive impact on theories than in either the Popperian or NMF approaches.

Lakatosians have two legitimate responses to direct replication failures. First, they may temporarily ignore such failures on the understanding that they may represent unimportant refutations of unspecified parts of the ceteris paribus clause rather than refutations of the “hard core” of the theory (Lakatos, 1978, p. 89, p. 48). Second, in the absence of an accepted ceteris paribus clause, even direct replications may vary from original studies in important ways (i.e., they are *variable replications* rather than *equivalent replications*; Rubin, 2020). Consequently, following a “negative heuristic” (Lakatos, 1978, p. 48), Lakatosians may get “creative” (Lakatos, 1978, p. 99) and explain replication failures by referring to potentially relevant differences between the original and replication studies (i.e., “hidden moderators”) in order to generate new, falsifiable, “auxiliary hypotheses” (Lakatos, 1978, p. 33) that qualify the “hard core” of the theory (e.g., Lakatos, 1978, p. 179; see also Putnam, 1991, pp. 125-126 & p. 130; for similar reasoning, see Popper, 2002, pp. 56 & 62). For example, they might continue to believe that, in general, “swanness causes whiteness,” but add the auxiliary hypothesis that this causal relation is moderated by location: “swanness causes whiteness, apart from in Australia, where swanness causes blackness” (Karawita et al., 2023; a “boundary condition,” Putnam, 1991, pp. 126-127).

Hence, Lakatosian scientists consider conditional causal statements, whereas Popperian scientists test universal noncausal statements.

Based on this idea of iterative theory modification, Lakatos (1978, p. 34) argued that scientists should move away from the appraisal of *single* theories and towards the appraisal of *series* of theories in *research programs*. According to Lakatos, a research program is “a series of theories, T_1, T_2, T_3, \dots where each subsequent theory results from adding auxiliary clauses to (or from semantical reinterpretations of) the previous theory in order to accommodate some anomaly, each theory having at least as much content as the unrefuted content of its predecessor” (p. 33). Research programs are then assessed in terms of whether they are *progressive* or *degenerative*. In a progressive research program, the new theories accommodate previous anomalies and make new successful predictions. In a degenerating program, however, the new theories only accommodate previous anomalies, and their new predictions remain unsupported (Lakatos, 1978, p. 34, p. 179).

Alternative Views

Commenting on Lakatos' (1978) approach, Zwaan et al. (2018) proposed that “replications are an instrument for distinguishing progressive from degenerative research programs” (p. 2). However, this proposal seems inconsistent with Lakatos' approach (for a similar conclusion, see Fletcher, 2021, p. 4). Lakatosian research programs require studies that test *new* (previously untested) hypotheses of *new* effects based on *new* (modified) theories. Hence, they do not imply the *direct replications* that Zwaan et al. advocate. In addition, a Lakatosian research program's negative heuristic forbids the refutation of a theory's hard core (e.g., “swanness causes whiteness”; Lakatos, 1978, p. 48; see also Putnam, 1991, p. 131). Hence, Lakatosian research programs do not even imply *conceptual replications* (i.e., studies that aim to refute the same theoretical hard core under different conditions). Instead, progressive research programs modify and develop theories that (a) accommodate previous anomalies and (b) make successful new predictions (e.g., “swanness causes whiteness, apart from in Australia, where swanness causes blackness”). Following Feest (2019, p. 901), the term “exploration” seems more appropriate than “replication” in this context. Furthermore, and contrary to Zwaan et al., it is the results of *innovative new studies*, rather than either direct or conceptual replications, that allow us to distinguish progressive research programs from degenerative ones.

Earp and Trafimow (2015) also considered the replication crisis in relation to auxiliary hypotheses, *ceteris paribus* clauses, and Lakatos' (1978) approach. Similar to Zwaan et al. (2018), they proposed that repeated failures of direct replications by different researchers should gradually decrease confidence in an original study's positive result (Earp & Trafimow, 2015, p. 8). Again, however, from a Lakatosian perspective, our confidence in the theoretical hard core that is used to explain a study's positive result should be unaffected by numerous failed replications of that result. Instead, it is our confidence in the progressiveness of a broader research program that should be reduced following the falsification of auxiliary hypotheses that are used to explain the replication failures.

Summary

In summary, multiple unexpected replication failures should be more concerning in the Popperian and NMF approaches because they imply that the theory testing process is not sufficiently rigorous to screen out inadequate theories. In this respect, scientists' adherence to the Popperian and NMF approaches may be at least partly responsible for the sense of a replication “crisis.” In contrast, multiple replication failures are less concerning in the Lakatosian approach

because (a) causal theories are not the subject of logical refutations, (b) scientists are used to working “in an ocean of anomalies” (Lakatos, 1978, p. 53), and (c) replication failures represent opportunities for theory development rather than cues for theory abandonment (for an illustration, see Sweller, 2023).

Endnotes

1. Personally, I favour the Lakatosian approach over the Popperian or naïve methodological falsificationist approaches.
2. In my view, Lakatosian “causal connections” represent what Popper (2002) described as “strictly or purely existential statements (or ‘there-is’ statements)” (p. 47, italics omitted; e.g., “there is at least one case in which swanness causes whiteness”). Strictly existential statements cannot be falsified by basic statements (Popper, 2002, p. 48). Consequently, Popper treated them as “metaphysical” (p. 48). Confusingly, Popper (1983, p. 288) used the term “causal hypotheses” to refer to “non-probabilistic” hypotheses as opposed to probabilistic hypotheses. However, these non-probabilistic hypotheses did not imply Lakatosian causal connections. It is also worth noting that Popper (2002) replaced the metaphysical principle of causality with a “methodological rule” (p. 39) “always to try to deduce statements from others of higher universality” (p. 107). For example, one might deduce the statement “all swans are white” from the more universal statement that “all birds are camouflaged” and the “initial conditions” of swans’ often snowy habitats (Holt, 2022). Again, however, the key point here is that Popperian tests logically refute *noncausal* universal statements rather than *causal* connections.
3. Popper (2002, pp. 54-55) considered the concept of a “theoretical system” containing different hypotheses of varying levels of universality. However, any strictly universal statement could take the role of both theory and hypothesis. For example, Popper (1983) described the statement “all swans are white” as a theory (e.g., p. xx) and a hypothesis (e.g., p. 343); sometimes on the same page (p. 234).
4. According to Meehl (1990), “common sense tells us that both the importance and the dangerousness of C_p [a ceteris paribus clause] are much greater in psychology than in chemistry or genetics” (p. 111). My own view is that a ceteris paribus clause should remain in doubt during *any* scientific investigation.
5. In practice, Popper (2002) argued that researchers should consider a theory “falsified” when they “discover a *reproducible effect* which refutes the theory” (p. 66, italics in original). However, in response to the question “how often has an effect to be actually reproduced in order to be a ‘reproducible effect’,” he responded “in some cases *not even once*” (Popper, 2002, p. 67, italics in original). Hence, even one-off effects can falsify a theory when they are independently verifiable.

References

- Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces* (1st ed.) Wiley.
- Derksen, M. (2019). Putting Popper to work. *Theory & Psychology*, 29(4), 449-465. <https://doi.org/10.1177/0959354319838343>
- Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, 6, Article 621. <https://doi.org/10.3389/fpsyg.2015.00621>

- Feest, U. (2019). Why replication is overrated. *Philosophy of Science*, 86(5), 895-905. <https://doi.org/10.1086/705451>
- Feyerabend, P. (1975). Imre Lakatos. *The British Journal for the Philosophy of Science*, 26(1), 1-18. <https://doi.org/10.1093/bjps/26.1.1>
- Feynman, R. P. (1955). The value of science. *Engineering and Science*, 19(3), 13-15. <https://calteches.library.caltech.edu/1575/1/Science.pdf>
- Firestein, S. (2012). *Ignorance: How it drives science*. Oxford University Press.
- Fletcher, S. C. (2021). The role of replication in psychological science. *European Journal for Philosophy of Science*, 11, Article 23. <https://doi.org/10.1007/s13194-020-00329-2>
- Flis, I. (2019). Psychologists psychologizing scientific psychology: An epistemological reading of the replication crisis. *Theory & Psychology*, 29(2), 158-181. <https://doi.org/10.1177/0959354319835322>
- Hager, W. (2000). About some misconceptions and the discontent with statistical tests in psychology. *Methods of Psychological Research Online*, 5(1), 1-31. <https://psycnet.apa.org/record/2001-03459-001>
- Haig, B. D. (2009). Inference to the best explanation: A neglected approach to theory appraisal in psychology. *The American Journal of Psychology*, 122(2), 219-234. <https://doi.org/10.2307/27784393>
- Holt, D. W. (2022). Why are snowy owls white and why have they evolved distinct sexual color dimorphism? A review of questions and hypotheses. *Journal of Raptor Research* 56(4), 440-454. <https://doi.org/10.3356/JRR-21-56>
- Johansson, I. (1980). Ceteris paribus clauses, closure clauses and falsifiability. *Zeitschrift für Allgemeine Wissenschaftstheorie*, 11, 16-22. <https://doi.org/10.1007/BF01801276>
- Karawita, A. C., Cheng, Y., Chew, K. Y., Challagulla, A., Kraus, R., Mueller, R. C., ... & Short, K. R. (2023). The swan genome and transcriptome, it is not all black and white. *Genome Biology*, 24, Article 13, 1-24. <https://doi.org/10.1186/s13059-022-02838-0>
- 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.
- Lakatos, I. (1978). *The methodology of scientific research programmes (Philosophical Papers, Volume I)*. Cambridge University Press.
- Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, 1(2), 108-141. https://doi.org/10.1207/s15327965pli0102_1
- Merton, R. K. (1987). Three fragments from a sociologist's notebooks: Establishing the phenomenon, specified ignorance, and strategic research materials. *Annual Review of Sociology*, 13(1), 1-29. <https://doi.org/10.1146/annurev.so.13.080187.000245>
- O'Donohue, W. (2021). Are psychologists appraising research properly? Some Popperian notes regarding replication failures in psychology. *Journal of Theoretical and Philosophical Psychology*, 41(4), 233-247. <https://doi.org/10.1037/teo0000179>
- Pearce, N. (1990). White swans, black ravens, and lame ducks: Necessary and sufficient causes in epidemiology. *Epidemiology*, 1, 47-50. <http://www.jstor.org/stable/20065623>
- Popper, K. (1974). Reply to my critics. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper (Book II)*. Open Court.
- Popper, K. (1983). *Realism and the aim of science: From the postscript to the logic of scientific discovery*. Routledge.

- Popper, K. (2002). *The logic of scientific discovery*. Routledge.
- Putnam, H. (1991). The 'corroboration' of theories. In R. Boyd, P. Gasper, & J. D. Trout (Eds.), *The philosophy of science* (pp. 121-137). MIT Press.
- Reutlinger, A., Schurz, G., Hüttemann, A., & Jaag, S. (2021). Ceteris paribus laws. *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition). <https://plato.stanford.edu/archives/fall2021/entries/ceteris-paribus>
- Rubin, M. (2020). "Repeated sampling from the same population?" A critique of Neyman and Pearson's responses to Fisher. *European Journal for Philosophy of Science*, 10, Article 42, 1-15. <https://doi.org/10.1007/s13194-020-00309-6>
- Rubin, M. (2023a, June 7). The preregistration prescriptiveness trade-off and unknown unknowns in science: Comments on Van Drimmelen (2023). *Critical Metascience: MetaArXiv*. <https://doi.org/10.31222/osf.io/3t7pc>
- Rubin, M. (2023b, December 14). Type I error rates are not usually inflated. *MetaArXiv*. <https://doi.org/10.31222/osf.io/3kv2b>
- Sweller, J. (2023). The development of cognitive load theory: Replication crises and incorporation of other theories can lead to theory expansion. *Educational Psychology Review*, 35, Article 95, 1-20. <https://doi.org/10.1007/s10648-023-09817-2>
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, 41, e120. <https://doi.org/10.1017/S0140525X17001972>

Funding: I declare no funding sources.

Conflict of interest: I declare no conflict of interest.

Biography: I am a professor of psychology at Durham University, UK. In my work on metascience, I have argued that it is not always problematic to engage in questionable research practices such as hypothesising after the results are known and uncorrected multiple testing. I have also criticised some science reforms, such as preregistration and stricter adherence to Neyman-Pearson hypothesis testing. For further information about my work in this area, please visit <https://sites.google.com/site/markrubinsocialpsychresearch/replication-crisis>

Correspondence: Correspondence should be addressed to Mark Rubin at the Department of Psychology, Durham University, South Road, Durham, DH1 3LE, UK. E-mail: Mark-Rubin@outlook.com



Copyright © The Author(s). OPEN ACCESS: This material is published under the terms of the Creative Commons BY Attribution 4.0 International licence (CC BY 4.0 <https://creativecommons.org/licenses/by/4.0/>). This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use.
