Finding Wealth in Waste: Irreplicability Re-Examined

Bart Penders* and A. Cecile J. W. Janssens

Irreplicability is framed as crisis, blamed on sloppy science motivated by perverse stimuli in research. Structural changes to the organization of science, targeting sloppy science (e.g., open data, pre-registration), are proposed to prevent irreplicability. While there is an unquestionable link between sloppy science and failures to replicate/reproduce scientific studies, they are currently conflated. This position can be understood as a result of the erosion of the role of theory in science. The history, sociology, and philosophy of science reveal alternative explanations for irreplicability to show it is part of proper, informative and valuable science. Irreplicability need not equate research waste. Sloppy science is the problem, also when results do replicate. Hence, the solution should focus on opposing sloppy research.

1. Introduction

Concerns about the replicability of scientific studies are dominating headlines in science. The attribution of the label “crisis” is contested, including explanations of its origins and proposals to remedy the situation. We argue that a widespread take on the replicability crisis, in which problems with replicability are attributed primarily to sloppy science, is limited and possibly dangerous. We argue that irreplicability is not a problem of the state of science, but a characteristic of the messiness of science and the world it aims to understand.

We discern two mechanisms underlying irreplicability. The first deals with variation in scientific approaches, sampling, methods, and styles of reasoning. The second deals with sub-par research conduct. The first is part of normal science, whereas the latter represents a separate issue, problematic even when it replicates. This article draws from the history, philosophy, and sociology of science, to argue that sloppy science is an insufficient and stigmatizing explanation for non-replicability. We offer three alternative views, including implications for the validity and credibility of science.

The NSF Subcommittee on Replicability in Science defines reproducibility as the ability to duplicate the results of a study using the data from that study with the same or similar analytic strategies, while replication refers to the ability to duplicate the results of a study using new data. This is on replication that we focus our argument here.

2. Issue

Many commentaries on “the replication crisis” point into similar directions: problems with reproducibility follow from poor research design, data collection, or analysis. Most commentaries are hesitant to do so explicitly, but some equate non-replicable science with junk science or argue that fraud and incompetence are to blame. This take on irreplicability blames researchers and their lack of skills or moral character and suggests that it is a problem that will disappear upon fixing sloppy science. Sloppy science and irreplicability are not, however, synonyms. We argue that sloppy science deserves “fixing,” even if it does replicate. In contrast, studies that do not replicate are not necessarily sub-par science; they may offer opportunities to learn from, to improve research.

Sloppy science wastes limited resources and hampers scientific inquiry. It is also harmful to science’s public credibility. When tales of fraud, mischief, and misconduct are dominating press coverage on science, its reliability and perceived trustworthiness risk decline. When the rigor of conduct and analyses is distrusted, the solution is found in requesting more transparency in research design, data collection, and data analysis so that all research steps become traceable and repeatable, allowing replication. The active pursuit and promotion of open science associated with increased transparency, promoted to filter out sloppy science has uncritically been adopted as the cure for irreplicable research. However, open science’s goal cannot and should not be the eradication of irreplicability but rather that of sloppy research.

Irreplicability has grown into a public matter of concern, in parallel to concerns about the state of science and the rigor of inquiry. However, equating irreplicability with sloppy science has severe consequences, especially with scientists working under conditions of insecurity, uncertainty, and constant
pressure. Careers, the trustworthiness of individual experts and studies, and the fate of students are at stake. On a higher aggregate level, it harms the status and credibility of disciplines, as can currently be observed in, for example, social psychology, as well as that of science as a whole.

3. Alternatives

If sloppy science is not the sole driver of irreplicability, where else can its origins be found and understood? Three alternative conceptualizations of science and the roles reserved for irreplicability are offered.

3.1. A Science of Theories

Philosopher Karl Popper developed the now commonly accepted position that scientists should not attempt to verify their claims but instead try to falsify them. Science, he argued, cannot advance solely through value-free observation but must start with a theory that makes precise predictions about the world. Subsequently one does not seek out observations that verify the theory, for they would add little. If we devise the theory that artificial intelligence (AI) will improve medical image recognition, endless observations of improved image recognition cannot exclude an AI failing in a certain subset. Our best contribution to the theory would be the attempt to disprove it,[5] to actively seek out that subset. Failure to falsify theories despite best efforts does not mean that they are true. It only means that they are more likely to be. Truth is always out of reach according to Popper’s critical rationalism.

A failure to replicate would mean that our theory is not able to make predictions that are consistent. The theory then has to be discarded. That does not mean that we do not speak of it again. Rather, intricate knowledge about the theory and about the exact data that disproved it can help scientists generate a new theory that offers better predictions.

According to critical rationalism, rather than blaming sloppy science, a failure to replicate is an opportunity to refine science’s theories, the route to improve the set of predictions science can make. Falsification is the only way for science to advance.

3.2. A Science of Paradigms

Scientific theories delineate ways to understand phenomena and generate different predictions. These theories are important as part of the identity of groups and their investigators. Scientists build their careers and reputations on one or a few theories and their falsification could be devastating. Historian Thomas Kuhn noticed that observations not fitting a theory are not used to disprove it. Rather, reasons are found to ignore or discredit them: the data or experimenter cannot be trusted, circumstances are not comparable, etc. Observations not fitting the theory are treated as an anomaly, not as a reason to discard it.

Kuhn demonstrated that scientists stick with their theories despite evidence to the contrary.[6] These theories set the frame within which research is done: the paradigm. They determine which questions can be asked, which studies are legitimate, and who are legitimate other experts, subscribing to the same paradigm. This situation Kuhn called normal science. Only when two conditions are met, can this change. First, the amount of anomalies has to grow so large that they cannot be ignored anymore. Second, one or more other rival theories must be available to fill the void, should the original theory be discarded. If sufficient anomalies emerge, science faces a time of crisis. In normal science, the boundaries of what one can do and understand are clear, imposed by a dominant theory or paradigm. In a time of crisis, rival theories compete for dominance. Only when one theory emerges as dominant, will the time of normal science return – this time with a new paradigm or theory that sets new boundaries for legitimate questions.

In a situation of normal science, failure to replicate is most likely ignored with little consequence for the theory upon which the studies were based. That may change when such failures aggregate, but only when competing theories are available. In a time of crisis, a successful replication may serve the function of helping scientists choose sides between rival paradigms, aiding the victory of one.

3.3. A Science of People

Drawing from Kuhn’s initial attempts at a social history of science, other sociologists and historians initiated a research program called the sociology of scientific knowledge, which conceptualizes science as a cultural and human endeavor in which groups and individuals compete for attention. Accordingly, scientists play status games and compete for credibility and prestige while sociologists have studied the ensuing “credibility economy” accompanying (scientific) knowledge. “In sociological terms of art, an individual’s belief (or an individual’s claim) was contrasted to collectively held knowledge. The individual’s belief did not become collective – and so put out of knowledge – until and unless it had won credibility. No credibility no knowledge.”[7]

To sociologists, the struggle to achieve this credibility is the key characteristic that can explain why certain claims achieve the status of knowledge and others do not. This requires consensus-building, a political process: finding allies, resources, and support in greater numbers and at a greater speed than one’s competitors. Replication of a study is not something that adds a lot of credibility to one’s work, so it is a relatively rare activity – it offers little status reward. However, when claims or theories compete for credibility in science, everything counts. This means that successful or failed replications can contribute to the power of one consensus-building campaign while destabilizing another. If the competing claim or theory and their backers have enough power, a single failed replication will not tilt the scales. If the competing theories yield comparable power, it might. In such political struggles, accusations of sloppy science are not solely diagnoses or explanations. Rather they are a political strategy influencing consensus formation.[7]
4. Discussion

Irreplicability can be understood as a core part of the scientific process (see Table 1). To Popper, it allows the discarding and improving of the collective of theories upon which science rests. To Kuhn, irreplicability is initially ignored, but as the reservoir of concerns fills, it still is part of the initiation of a scientific crisis, from which novel paradigms are born. The sociology of scientific knowledge shows how scientists compete, and labels such as sloppy science or irreplicability carry reputational weight and are used accordingly. None of these views on science point to sloppy science to explain failures to replicate. All of them show how irreplicability contributes to science’s knowledge dynamics, to identify the limits of science’s understanding, the place where the predictive power of our theories or paradigms ends. Reliably identifying these limits and learning from them is an immense source of value to science. Accordingly, irreplicability directly contributes value to science, provided this value is not offset by sloppy science.

The false view of irreplicability as a problem solely of sloppy science has philosophical roots as well, in empiricism. In this view, science is built on a collection of value-free observations of the natural world. The larger that set of observations (data) to verify the claim, the more likely the truth of the claim. These observations, however, need to be “pure” and value-free to legitimately form the basis for science. Collections of data, when trusted sufficiently, can generate theories, but only scientists of excellent moral character are uniquely able to observe objectively and value-free. Irreplicability of a study means that one of the studies was either technically imperfect, observation was insufficiently objective (biased), or the studies were too different. This empiricist view of science has lost serious traction in the philosophy of science halfway through the previous century. Simplified, this view remains alive, in popular representations of science, mythologizing scientific capacities and the morally impeccable character of scientists.[9] Empiricism also never disappeared from scientific practice itself, possibly because it offers appealing identities for scientists, deflates the value of theory, and allows scientists to shy away from responsibility.

The deflation of the value of theory helped fuel the conflation of replicability problems and sloppy science. However, as we show, irreplicability need not overlap with sub-par science. Sloppy science is real and exacerbated by broken evaluation cultures. However, because irreplicability can have other, less malign origins, a sole focus on transparency, such as open science, will never “solve” replication problems. Studies of the epistemologies of open science suggest humility is required with respect to its novelty claims and ambitions,[9] while the roots of science’s epistemic crises remain plural.[10]

The irreplicability debate often serves as a proxy for debate on misconduct, fraud, and sloppy science. Unravelling irreplicability from sloppy science means unearthing value from waste. We call on those who participate in the replicability debate to look beyond science as the sum of data and methods and to reassign due value to theory. This will allow scientists and other experts to assess studies and their replication, or non-replication on their own merits. In summary, we should seek to 1) separate irreplicability from sloppy science; 2) continue to oppose sloppy science; and 3) destigmatize irreplicability and embrace its value.

Acknowledgments

The authors report no concrete funding associated with this work aside from the support of the employers.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

irreplicability, organization of science, replication crisis, responsible research, sloppy science, theory

Received: September 6, 2018
Published online: October 12, 2018

[1] K. Bollen, J. T. Cacioppo, R. M. Kaplan, J. A. Krosnick, J. L. Olds, Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science. Report of the Subcommittee on Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences. National Science Foundation, Arlington, VA 2015.

[2] D. Randall, C. Welser, The Irreproducibility Crisis of Modern Science: Causes, Consequences, and the Road to Reform. National Association of Scholars, New York, NY 2018.
[3] C. Laine, S. N. Goodman, M. E. Griswold, H. C. Sox, Ann. Intern. Med. 2007, 146, 450.
[4] S. Leonelli, 2018, Re-thinking reproducibility as a criterion for research quality. Preprint: http://philsci-archivepittedu/id/eprint/14352
[5] K. R. Popper, Conjectures and Refutations. The Growth of Scientific Knowledge. Routledge and Kegan Paul, London 1963.
[6] T. S. Kuhn, The structure of scientific revolutions. University of Chicago Press, Chicago 1970 [1962].
[7] S. Shapin, Perspect. Sci. 1995, 3, 255.
[8] C. L. Veldkamp, C. H. Hartgerink, M. A. van Assen, J. M. Wicherts, Account. Res. 2017, 24, 127.
[9] P. Mirowski, Soc. Stud. Sci. 2018, 48, 171.
[10] A. Saltelli, S. Funtowicz, Futures 2017, 91, 5.