Philosophy of Science (2024), **00**, 1–11
doi:10.1017/psa.2024.2

PSA PHILOSOPHY OF SCIENCE ASSOCIATION
PHILOSOPHY OF SCIENCE

SYMPOSIA PAPER

What is the Replication Crisis a Crisis Of?

Uljana Feest

Leibniz Universität Hannover, Germany
Email: feest@philos.uni-hannover.de

(Received 23 April 2023; revised 02 October 2023; accepted 05 October 2023)

Abstract

In recent debates about the replication crisis, two positions have been dominant: one that focuses on methodological reforms and one that focuses on theory building. This paper takes up the suggestion that there might be a deeper difference in play, concerning the ways the very subject matter of psychology is construed by opposing camps, i.e., in terms of stable effects versus in terms of complexity. I argue that each gets something right, but neither is sufficient. My analysis suggests that the context sensitivity of the psychological subject matter needs to be front and center of methodological and theoretical efforts.

1. Introduction

It has become a commonplace that psychology entered a crisis sometime in the second decade of this century. The crisis was triggered by the recognition that seemingly established experimental results could not be replicated, a fact that has given rise to a high degree of stimulating methodological self-reflection within psychology and has attracted philosophical attention as well. Roughly, we can distinguish between two types of responses to the replication crisis, both of which see the ubiquity of replication failures as symptomatic of a deeper problem. The first views the replication crisis as rooted in the prevalence of questionable research practices (e.g., p-hacking and retrospective hypothesis fitting), which give rise to non-replicable results. Scholars in this debate, sometimes associated with the meta-science movement, have focused on ways in which psychological research can be regulated, e.g., by calling for the preregistration of experiments.[1] Another group of scholars takes the narrow focus on (the replicability of) experimental effects itself to be part of a larger problem, namely a relative sparsity of sustained theoretical work in psychology. In turn, this has given rise to some efforts to develop methodologies of theory construction and to think more generally about what theoretical work in psychology might look like.

[1] https://www.cos.io/

Both of these discussions expose problems with the epistemic practices of psychology, but look for the root problem of the crisis in different places. This motivates different answers to the question of what the replication crisis is a crisis of, and (consequently) what kinds of measures should be taken to resolve it. This paper argues that the two diagnoses are mutually compatible, but that there is a deeper question at stake: Rigorous methods of experimental design, data analysis, and theory construction will only be fruitful if applied to the right questions about the right (kinds of) objects. I will explore questions about the "right" questions and objects in psychology by taking as a point of departure Morawski's (2021) suggestion that differing responses to the replication crisis are rooted in different conceptions of the psychological subject matter.

Section 2 analyzes Morawski's (2021) characterization of the difference between "reformers" and "challengers" to consider her suggestion that they differ (among other things) with regard to the ways in which they construe the psychological subject matter: in terms of effects versus in terms of complexity and context sensitivity. Highlighting that replicability is about generating data that allow inferences to specific phenomena, I will argue that both "effect seekers" and "complexity mongers" are confronted with similar epistemic problems. Section 3 argues that psychology needs a more sustained look at what psychological theories are *about*. Section 4 presents an answer to this question, which highlights the importance of studying the context sensitivity of psychological objects in its own right.

2. Reformers and challengers: Competing takes on the replication crisis

Morawski (2021) has recently suggested that differing assessments of the gravity of the replication crisis may be due to differing background assumptions about the psychological subject matter. She divides the community into two groups, "reformers" and "challengers," and argues that *reformers* emphasize the importance of uncovering stable effects, whereas *challengers* view the subject matter of psychology as complex and context sensitive. I refer to proponents of the first position as "effect seekers" and the second as "complexity mongers." While these are, of course, caricatures, they are useful for my analytical purposes in this paper. This section disambiguates the notions of "effect" and "complexity" to get an analytical grip on some issues underlying the replication crisis.

2.1. Disambiguating "effect": Data and (two kinds of) phenomena

Picking up the notion that some researchers (typically those more concerned about replication failures) construe the psychological subject matter in terms of stable effects, it will be helpful to begin by distinguishing between two usages of the term "effect": the first refers to experimental effects (i.e., data), while the second refers to effects that are inferred from experimental effects (i.e., phenomena).[2] Debates about replicability mostly turn on the former, i.e., on the replicability of *experimental effects*, given reasonably similar experiments. Experimental effects, qua *data*, are used to

[2] See Bogen and Woodward (1988) for the classic formulation of the distinction between data and phenomena.

https://doi.org/10.1017/psa.2024.2 Published online by Cambridge University Press

make inferences to statements about *phenomena*. Such statements are best understood as the results of experiments.

I maintain that experimental psychologists will need their experimental data to not only be replicable, but also support the intended conclusion about a given phenomenon. Addressing this latter point first, we can say that researchers aim at experimental effects that serve as reliable evidence for the intended result. I am understanding the term "reliability" as referring to a situation where there is "the right sort of pattern of counterfactual dependence between the data and the conclusions investigators reach on the phenomena themselves" (Woodward 2000, S163). I interpret this to mean that data are reliable evidence for a specific claim only if they stand in the right kind of a relationship to the phenomenon that we draw inferences about.[3] We will need to say more about what "in the right kind of relationship" means, but it seems clear that reliability is a stronger requirement than "mere" replicability, though presumably replicability is a necessary condition for reliability.

Given what was just argued, it seems that both effect seekers and complexity mongers ought to be concerned if they fail to generate replicable experimental effects. So, what are we to make of the suggestion that complexity mongers are less worried about replication failures than effect chasers? To address this question, let's consider the nature of the phenomena that psychologists try to make inferences to. Here, a second distinction becomes relevant, namely, between *two kinds of phenomena* that psychologists are interested in, and thus between two kinds of experimental results they might wish to establish. The first concerns the existence of real-world behavioral (stimulus–response) effects, which are similar to the ones found in the experiment. The second concerns the existence of some feature of the psychological subject matter that cannot be immediately observed in the lab and that is not similar to experimental effect. Feest (2011) refers to such unobservable effects as "hidden phenomena."

Examples of the former are alleged effects such as social priming, power posing, or the Mozart effect. In those cases, researchers attempt to create experimental effects and treat those effects as evidence for a similar effect that exists outside the lab. An example of the latter is provided by facial feedback research, i.e., the (putative) phenomenon that there is a feedback mechanism between smiling and experienced positive emotions. The hypothesis that this phenomenon is real was tested by Strack et al. (1988), in an experiment that required subjects to hold pencils in their mouth (in a way that simulated the facial muscles required for smiling) and subsequently measured the intensity of humorous emotions experienced when reading a funny cartoon. The resulting data seemed to confirm the facial feedback hypothesis. Clearly, though, researchers who perform the latter kind of experiments do not intend the circumstances under which the data are generated to be similar to situations under which facial feedback might be triggered in the real world.

The crucial point here is that in both kinds of cases researchers make inferences from experimental effects (data) to the effects of interest (phenomena). The difference is that the effects of interest are located in different places: In the former

[3] The notion of data reliability has been discussed in relation to neuroscientific experiments (Sullivan 2009), but has not received much attention in the philosophy of psychology so far.

https://doi.org/10.1017/psa.2024.2 Published online by Cambridge University Press

Figure 1. Schematic representation of experiment that targets a simple stimulus–response mechanism.

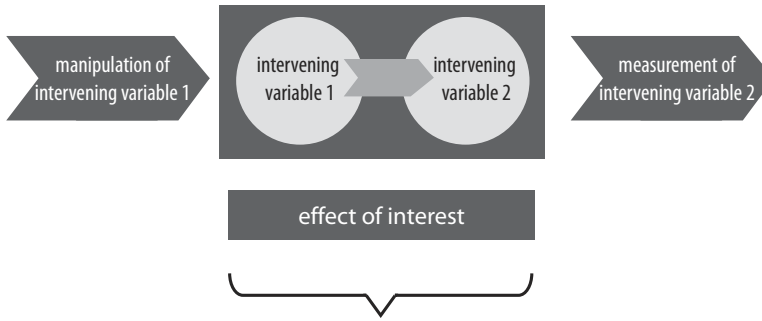

Figure 2. Schematic representation of experiment that targets an "internal" mechanism.

kind of scenario, the effects of interest are stimulus–response effects; in the latter, they are effects internal to the organism (see figures 1 and 2, respectively).

2.2. *Disambiguating "complexity"*

If I am right with my above analysis, it seems that my initial labels ("effect chasers" and "complexity mongers") are misplaced since both groups of scholars are interested in effects (both at the experimental level and as targets of their inferences). Nonetheless, I think that the distinction between effect seekers and complexity mongers points to an issue worth exploring. To this end, the current section takes a closer look at the notion of *complexity*.

Why do complexity mongers (even if forced to recognize the importance of replicable experimental effects) resist the suggestion that the replication crisis can be fixed by forcing researchers to implement stricter standards of hypothesis testing? The answer is that while failure to replicate an experimental effect is reason for concern, complexity mongers are less inclined to attribute such failures to questionable research practices alone. Instead, they emphasize the possibility of *other contributing factors*. Specifically, researchers who use their data to make inferences to internal phenomena (see figure 2) bring to the table a heightened sensibility for the difficulty of generating data that are not only *replicable*, but also *reliable*. As already indicated, for experimental effects to function as reliable data for a specific

https://doi.org/10.1017/psa.2024.2 Published online by Cambridge University Press

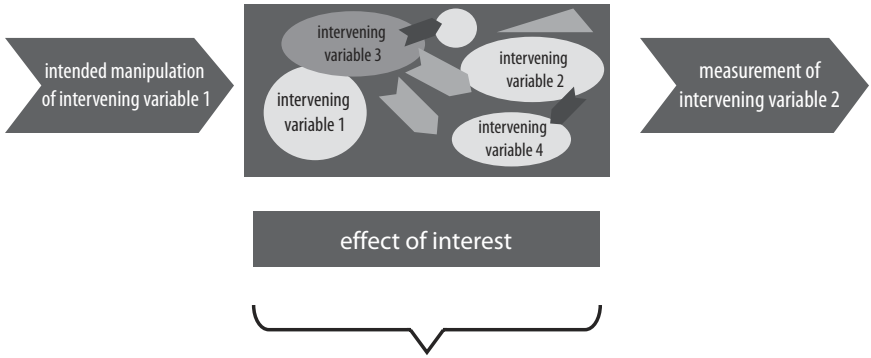

Figure 3. Schematic representation of potential distortions by other intervening variables that can make data unreliable as indicators of specific internal mechanisms.

experimental inference, they need to stand "in the right kind of counterfactual relationship" to the phenomenon described by the conclusion. We can unpack this to mean that data can only be regarded as reliable evidence for a specific claim (e.g., that there is a feedback mechanism between smile muscles and experienced emotions) if the experimental manipulation in fact triggered the effect of interest, and if the experimental data in fact measure the effect of interest.

It is clear that the requirement of reliability calls for an undistorted causal path between experimental manipulation and experimental measurement, such that the data are not confounded. It also seems very plausible that evidence about "internal phenomena" can easily be confounded by other internal phenomena, which are not easily controlled or even recognized (figure 3).

The distinction between *replicability* and *reliability* may be counterintuitive to readers immersed in the methodological literature in psychology, where the term "reliability" is sometimes equated with the ability to achieve the same effect when rerunning a test or experiment or test. However, this misses an important distinction, namely that between having replicable data and having data that support the conclusion to an effect of interest. On a charitable interpretation, complexity mongers are sensitive to this difference, because appreciating the internal complexity of biological organisms makes them aware of the many ways in which experimental data might not be reliable vis-à-vis the intended experimental results.

Even though I have explained the problem of data reliability in relation to the internal complexity of the organism, the problem also arises for those who "only" aim to make inferences from experimental stimulus–response effects to the existence of stimulus–response effects in the real world. Confounders do not have to be internal to the organism, as figure 4 illustrates: When an experimenter manipulates an organism, they treat the data as the effect of that manipulation. However, there might be uncontrolled variables in the experimental environment. Furthermore, the experimental stimulus might be described in a way that does not pick out the causally efficacious aspect. In such cases, the resulting data are unreliable vis-à-vis the intended conclusion. In other words, those interested in (mere) stimulus–response

https://doi.org/10.1017/psa.2024.2 Published online by Cambridge University Press

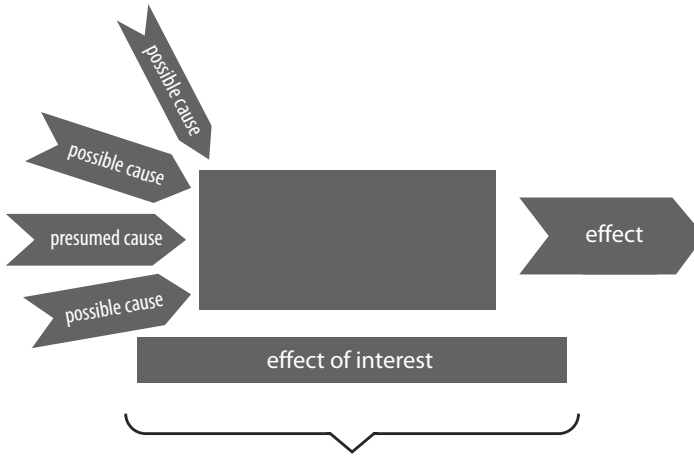

Figure 4. Schematic representation of conceivable environmental confounders affecting data reliability.

effects need to be just as worried about unreliable data as those interested in hidden effects.[4]

3. Theory to the rescue?

The analysis in the previous section has established that both effect seekers and complexity mongers need to be concerned with stable effects (on the level of data and on the level of phenomena; section 2.1). It has also revealed that both need to reckon with the complexity that can threaten data reliability (section 2.2). In other words, effect chasers rightly emphasize the importance of replicable data. Complexity mongers rightly point to the difficulties of generating such data. I have unpacked this latter point to refer to both (a) the difficulty of generating experimental data that can be reproduced, and (b) the difficulty of generating data that allow for the intended inferences (i.e., data that are *reliable* as evidence for specific hypotheses about a given phenomenon). I side with the complexity mongers in arguing that this second point, in particular, cannot be resolved by improving replicability alone.

My analysis converges with recent methodological writings in psychology, which have also pointed out that replicable effects in and of themselves, even if they could be achieved more easily, would not be sufficient for claims about phenomena: while "methodological and statistical solutions to the replication crisis will … help ensure solid stones … they don't help us build the house" (Muthukrishna and Henrich 2019, 1–2). One conclusion from this seems to be that one needs something of a blueprint for "the house," i.e., a theory, or at least a sketch of a theory. As such, this and other writings allude to the point that an adequate response to the replication crisis will

[4] Data reliability can also be affected by mistaken assumptions about the population that is sampled by the experimental subjects. This aspect will not be pursued here.

https://doi.org/10.1017/psa.2024.2 Published online by Cambridge University Press

require theoretical work in addition to methodological reforms. Relevant recent work includes attention to theory-building methodology (Borsboom et al. 2021), discussions of what theories might look like in psychology (e.g., van Rooij and Baggio 2021) as well as the role of formal methods as a way to constrain conceptual vagueness in hypothesis testing (Fried 2020a; Devezer et al. 2021).

Relatedly, Scheel et al. (2021) have pointed out that when psychologists test hypotheses by means of experiments, it is often hard for different researchers to agree on their correct interpretation, because the "derivation chain" (Meehl 1990) between theory, hypotheses, and data is underspecified. This observation fits well with recent attention to the problem of underdetermination in psychological experiments (Uygun Tunç and Tunç 2023; Oude Maatman 2021). It also speaks directly to my concern about *data reliability*, as I have been using the concept here. Addressing the underdetermination of phenomena by data amounts to attempting to improve data reliability. It seems clear that this is closely related to understanding—and physically implementing—the derivation chain between theory and data. The question is what kind of research is need to accomplish this.

I agree with the suggestion by Scheel et al. (2021) that the research in question needs to focus on concept formation and exploratory research (including both exploratory experiments and formal modeling), while noting that this does not commit me to a clear-cut distinction between exploratory and confirmatory research (see also Devezer et al. 2021; Rubin and Donkin 2022). However, I argue that a focus on theoretical and exploratory work highlights the more fundamental question of what psychological theories, models, and concepts are actually about. As I have shown above, questions about data reliability are a concern for both effect chasers and complexity mongers. This suggests that there are peculiarities of the psychological subject matter that both sides have to grapple with, independently of their specific theories or research interests.

4. The context sensitivity of the units of psychological analysis

In search of a peculiarity of the psychological subject matter, let's begin with Fried's contention that psychological theories are about phenomena, not about data (Fried 2020b, section 3), followed, a few pages later, by the assertion that "psychological constructs can be thought of as target systems" that "are represented via a theory's structure, which, like the target system, features components and relations among them" (ibid, section 3.2). I would like to point out that there is an important difference between phenomena and systems: Theories can *explain* individual phenomena, but systems can exhibit *multiple phenomena*.[5] Thus, I argue that while any give hypothesis derived from a theory can be about a specific phenomenon, psychological theories are usually not *just* about one specific phenomenon, but about a set of interrelated phenomena that are assumed to jointly constitute the object of research (or, as Fried calls it, the "target system"). Think, for example, of a psychological object like *emotion*. Clearly, this object has multiple phenomena associated with it, including a great variety of behavioral responses to stimuli (stimulus–response effects), but also

[5] Whether psychological objects are indeed neatly separated systems is an open question, but we will ignore this question for now.

https://doi.org/10.1017/psa.2024.2 Published online by Cambridge University Press

internal/hidden phenomena, such as the facial feedback mechanism and emotional experiences.

Once we recognize that objects of psychological investigation are often *systems of phenomena*—or "clusters of phenomena" (Feest 2017)—this adds to our appreciation of the complexity of the psychological subject matter (figure 3). It also raises the question of what is a reasonable way to conceptualize the units that contain, or constitute, such systems, such that they can be differentially affected by environmental factors (figure 4). In this vein, I distinguish questions about the *objects* of psychological research (e.g., emotion) from questions about the *units of analysis* psychologists are interested in. Moreover, I suggest that we understand objects of psychological research as complex cognitive, behavioral, and experiential capacities (Feest 2022b), which are exhibited by individual organisms.[6]

This brings to the fore another claim that Morawski (2021) attributes to "challengers," namely that "psychology's objects are not only sensitive to material conditions of the world, including the laboratory, but also affected by the shifting meanings that individuals derive from contexts both inside and outside the lab" (Morawksi 2021, 4). Darwin's facial feedback hypothesis is a case in point: while Strack et al.'s (1988) study seemed to confirm Darwin's hypothesis, a later replication study did not. Even more recently, it was found that the replication study had in fact introduced a small change that made the effect go away, thus confounding the data, i.e., the fact that they had filmed the study participants during the replication study—see Feest (2022a) for details. The example illustrates the way in which specific phenomena associated with an object can be context sensitive. It also illustrates that it is not obvious that data reliability could be improved by an improved theory of the object (emotion) alone since it seems that the confounder had to do with the awareness of being filmed, not with *emotion*, narrowly construed as the "target system." This is crucial here since it suggests that what makes data unreliable is an ineliminable part of psychological objects (complex cognitive, behavioral, and experiential capacities) and of the units that exhibit the phenomena peculiar to the subject matter (organisms as a whole).

The conclusion I want to draw from the above is that the context sensitivity of the psychological subject matter needs to be at the center of both theoretical and empirical work, not merely as a way of controlling for confounders but also because, in the long run, it is the experiences, cognitions, and behavior of complex organisms in complex environments that psychology needs to focus on.

5. Going forward (instead of a conclusion)

The preceding analysis laid out why I (like many others) don't think that the replication crisis is merely a crisis of failure to apply stringent methodological rules to practices of hypothesis testing and data analysis. While I agree with "complexity mongers" that the crisis is (at least in part) due to the sheer complexity of the subject matter, I have tried to unpack what this means in more specific terms by pointing to the problems of context sensitivity (as a feature of the psychological subject matter)

[6] The notion that psychological objects are capacities is taken from Cummins (1983). I concur with van Rooij and Baggio (2021) that such capacities will likely receive mechanistic explanations.

https://doi.org/10.1017/psa.2024.2 Published online by Cambridge University Press

and data reliability (as a feature of experimental evidence). Crucially, the latter is closely related to the former.

I have proposed a specific account of what the replication crisis is a crisis of (unreliable data in conjunction with a lack of reflection on what are units and objects of psychological analysis). Let me add two disclaimers. First, I am not claiming that the underlying issue I have identified (concerning the subject matter of psychology) is the only issue worth exploring. Second, I have not presented an easy solution to the crisis. I do, however, think my analysis points in two directions. Even though the search for effects will continue to be an important part of psychological research, more efforts are needed to (1) think about how these effects contribute to our overall understanding of the objects of psychological research, and (2) explore how their manifestations are affected by variables internal and external to the organism. Small changes in the experimental design can be confounders relative to a specific intended experimental inference. But looked at from a different perspective, they can also indicate ways in which the object under investigation is context sensitive and thus moderated by the change in question.[7] In this vein, I would press that we regard the context sensitivity of the psychological subject matter as a feature, not only as a bug: The very question of how organisms respond to environmental variations (but also how members of different populations respond differentially to similar environmental conditions) should be central to psychological research efforts.

While I agree with Eronen and Brinkman (2021, 785) that the way forward will include attention to stabilizing phenomena and "strengthening the conceptual basis of psychological theories," the crucial question is how to delineate the corresponding objects and their component phenomena in the first place, and how to ensure that the data (experimental effects) that are generated in support of claims about robust phenomena are reliable. In this regard, I push for a macro-level perspective that takes the behavior of the whole organism into view first. My outlook here is sympathetic to earlier functionalist and ecological approaches to the psychological subject matter (from figures like James and Dewey to Gibson and Brunswik and gestalt psychology). Unlike those earlier approaches, my acknowledgment of internal phenomena and mechanisms as integral to the psychological subject matter recognizes the importance of integrating a solid empirical understanding of (what I have called) stimulus–response effects (figure 1) with mechanistic theorizing (figure 2) and how they are brought about (Hatfield 2021). Individuating the phenomena that are context sensitive in this way is going to be far from trivial (Wajnerman-Paz and Rojas-Libano 2022). However, I concur with de Houwer (2011) that it is important to distinguish the search and characterization of stimulus–response effects from cognitive (i.e., hidden) effects, and to direct conceptual, empirical, and theoretical work at the question of how they are related.[8] In conclusion, I argue that such simultaneous attention to the shape of the psychological subject matter and to the reliability of data is likely to be a crucial component of our response to the replication crisis.

[7] Let me stress that I am not claiming that data that are confounded relative to a specific hypothesis can simply be taken as evidence for a different hypothesis.

[8] I am grateful to Eronen and Brinkman (2021) for directing my attention to the piece by de Houwer (2011).

https://doi.org/10.1017/psa.2024.2 Published online by Cambridge University Press

Acknowledgements. I would like to thank the audience and other members of the PSA symposium for helpful questions and suggestions. I am also particularly grateful to Bart Penders for creating the flow diagrams that appear in this publication.

References

Bogen, James and James Woodward. 1988. "Saving the Phenomena." *The Philosophical Review* 97 (3):303–352.

Borsboom, Denny, Han L. J. van der Maas, Jonas Dalege, Rogier A. Kievit, and Brian D. Haig. 2021. "Theory Construction Methodology: A Practical Framework for Building Theories in Psychology." *Perspectives on Psychological Science* 16 (4):756–766. doi: 10.1177/1745691620969647.

Cummins, Robert. 1983. *Psychological Explanation*. Cambridge, MA: MIT Press.

De Houwer, Jan. 2011. "Why the Cognitive Approach in Psychology Would Profit from a Functional Approach and Vice Versa." *Perspectives on Psychological Science*, 6 (2):202–209.

Devezer, Berna, Danielle Navarro, Joachim Vandekerckhove, and Erkan Ozge Buzbas. 2021. "The Case for Formal Methodology in Scientific Reform." *Royal Society Open Science* 8 (3):200805. https://doi.org/10.1098/rsos.200805.

Eronen, Markus and Laura Brinkman. 2021. "The Theory Crisis in Psychology: How to Move Forward." *Perspectives on Psychological Science* 16 (4):779–788.

Feest, Uljana. 2011. "What Exactly is Stabilized When Phenomena are Stabilized?" *Synthese* 182:57–71.

Feest, Uljana. 2017. "Phenomena and Objects of Research in the Cognitive and Behavioral Sciences." *Philosophy of Science* 84 (5):1165–1176.

Feest, Uljana. 2022a. "Data Quality, Experimental Artifacts, and the Reactivity of the Psychological Subject Matter." *European Journal for the Philosophy of Science* 12:13. https://doi.org/10.1007/s13194-021-00443-9.

Feest, Uljana. 2022b. "Progress in Psychology." In *New Philosophical Perspectives on Scientific Progress*, edited by Yafeng Shan, 184–203. New York: Routledge.

Fried, Eiko. 2020a. "Lack of Theory Building and Testing Impedes Progress in the Factor and Network Literature." *Psychological Inquiry* 31(4):271–288. https://psyarxiv.com/zg84s.

Fried, Eiko. 2020b. "Theories and Models: What They Are, What They Are For, and What They Are About." Preprint, https://psyarxiv.com/dt6ev.

Hatfield, Gary. 2021. "Gibson and Gestalt: (Re)presentation, Processing, and Construction." *Synthese* 198 (Suppl 9):2213–2241.

Meehl, Paul E. 1990. "Why Summaries of Research on Psychological Theories are Often Uninterpretable." *Psychological Reports* 66 (1):195–244.

Morawski, Jill. 2021. "How to True Psychology's Objects." *Review of General Psychology* 26 (2):157–171. https://doi.org/10.1177/10892680211046518.

Muthukrishna, Michael and Joseph Henrich. 2019. "A Problem in Theory." *Nature Human Behavior* 3: 221–229. https://doi.org/10.1038/s41562-018-0522-1.

Oude Maatman, Freek. 2021. "Psychology's Theory Crisis, and Why Formal Modelling Cannot Solve It." https://psyarxiv.com/puqvs/.

Rubin, Mark and Chris Donkin. 2022. "Exploratory Hypothesis Tests Can Be More Compelling Than Confirmatory Hypothesis Tests." *Philosophical Psychology*. https://doi.org/10.1080/09515089.2022.2113771.

Scheel, Anne, Leonid Tiokhin, Peter Isager, and Daniël Lakens. 2021. "Why Hypothesis Testers Should Spend Less Time Testing Hypotheses." *Perspectives on Psychological Science* 16 (4):744–755.

Strack, Fritz, Leonard L. Martin, and Sabine Stepper. 1988. "Inhibiting and Facilitating Conditions of the Human Smile: A Nonobtrusive Test of the Facial Feedback Hypothesis." *Journal of Personality and Social Psychology* 54 (5):768–777.

Sullivan, Jacqueline. 2009. "The Multiplicity of Experimental Protocols: A Challenge to Reductionist and Non-Reductionist Models of the Unity of Neuroscience." *Synthese* 167:511–539.

Uygun Tunç, Duygu and Mehmet Necip Tunç. 2023. "A Falsificationist Treatment of Auxiliary Hypotheses in Social and Behavioral Sciences: Systematic Replications Framework." *Metapsychology* 7. https://doi.org/10.15626/MP.2021.2756.

https://doi.org/10.1017/psa.2024.2 Published online by Cambridge University Press

van Rooij, Iris and Giosuè Baggio. 2021. "Theory Before the Test: How to Build High-Verisimilitude Explanatory Theories in Psychological Science." *Perspectives on Psychological Science* 16 (4):682–697. https://journals.sagepub.com/doi/full/10.1177/1745691620970604.

Wajnerman-Paz, Abel and Daniel Rojas-Libano. 2022. "On the Role of Contextual Factors in Cognitive Neuroscience Experiments: A Mechanistic Approach." *Synthese* 200:402. https://doi.org/10.1007/s11229-022-03870-0.

Woodward, James. 2000. "Data, Phenomena, and Reliability." *Philosophy of Science* 67 (Proceedings): S163–S179.

Cite this article: Feest, Uljana. 2024. "What is the Replication Crisis a Crisis Of?" *Philosophy of Science*. https://doi.org/10.1017/psa.2024.2