

Credibility and Trust in the Psychological Sciences

SOCRATES Project description (Uljana Feest)

In recent years it has become widely known that scientific psychology is in a crisis. This crisis was triggered by the recognition that many established experimental results cannot be replicated (Fidler & Wilcox 2021). There are by now a number of analyses of what can lead to scientific results not being replicable. Diagnoses range from flawed experimental designs, statistical analyses, and publication practices to lack of theoretical guidance and conceptual clarity, but also include outright fraud. In turn, several explanations have been offered for why such problems persist, some of which are pitched at the level of individual misconduct, others at more structural issues of the way science is organized. One thing that is clear, however, is that the current crisis has cast doubt on the trustworthiness of individual researchers, the adequacy of psychology's methodological frameworks, and (consequently) on the credibility of its findings.

Worries about the trustworthiness and credibility of psychology are not new, nor are they focused on reproducibility alone. Earlier debates and concerns played out (a) with respect to the value-ladenness of applied psychological research, (b) in the context of foundational methodological debates that took place (and continue to take place) within psychology, and (c) with regard to the moral and epistemic issues surrounding the relationship between psychological research and its subjects. In this research proposal, I sketch some of these issues and outline areas of concern.

Applied Psychology, Values, and Contexts

Psychology rose to prominence in early 20th century America by virtue of its promise to serve the social good and to help make diagnostic tasks in clinical, educational and occupational selection processes more objective and efficient (Fancher & Rutherford 2011). Taking a longer view, scientific interest in interindividual psychological differences and their practical significance reach well into the 19th century. Significant episodes that come to mind include Francis Galton's statistical work on the inheritability of intelligence in connection to eugenics, and Franz Gall's research program ("phrenology") of establishing a typology and diagnostic system of character traits by cataloging bumps in people's skulls. (See Fancher & Rutherford, 2011, for accounts of each)

These two historical examples illustrate clearly why a certain degree of mistrust is in order: The latter is widely regarded as pseudoscience, and the former is associated by many with scientific racism, classism, and ableism (Gould 1981). The examples raise the question of whether wider-ranging worries about questionable values and quackery are warranted when it comes to the study of interindividual differences and whether such worries generalize to other areas of psychological research. In some historical cases, the motivation for the study of interindividual differences was explicitly egalitarian, i.e., informed by the desire to make specific opportunities available on the basis of merit rather than money or influence (Carson 2006; Wijsen et al. 2022). But in those cases, too, there is a lingering concern that the a-historical and a-contextual notion of merit/intelligence is itself misguided and that psychological research in this domain might be at best promoting a particular, one-sided (e.g., white, male, and euro-centric) understanding of the psychological subject matter.

This last point is significant because it suggests that issues with the trustworthiness of psychology arise not just because its research is sometimes laden with sinister political values, but also because the very assumption of a-contextual facts about the psychological subject matter might be problematic, owing to the fact that many (perhaps all) psychological concepts are thick concepts

(e.g., Alexandrova 2018). In other words, even if the social values informing a given research program are well-intentioned, the assumption that there are “neutral” facts about psychological kinds might be mistaken.

Crisis Debates Within Psychology

Somewhat separate from issues surrounding applied psychology and mental testing, academic psychology has been plagued by recurring “crisis” debates since the 19th century (Sturm & Mülberger 2012). While the issue of trust in psychology’s applied research plays out at the interface of science and society, these earlier crisis declarations were mostly debated *within* the emerging discipline of academic psychology. They often addressed worries about internal theoretical and methodological fragmentation at a point in time when psychology was still in the process of establishing itself as a discipline. For example, Karl Bühler’s 1929 book *Die Krise der Psychologie* was mostly an attempt to argue for a vision of psychology that can integrate differing theoretical and methodological approaches. (See Volume 43, Issue 2 of *Studies in the History and Philosophy of Science* (2012) for a range of articles about past crisis discourses in psychology).

It is worth highlighting that the most recent crisis-debate in psychology (the debate about replication failures) is somewhat different in character from previous ones: As just mentioned, previous crisis discourses often turned on disagreements about methodology and subject matter. By contrast, the replication crisis occurred before the backdrop of a well-established discipline, working within a seemingly well-established paradigm, namely one that views psychology as an experimental and quantitative science. One reason why replication failures were greeted with such alarm was that they revealed wide-spread disregard for basic principles of scientific research and statistical analyses, drawing attention to practices like p-hacking, underpowered studies, and HARKing, which seemed to be in violation of the very methodological principles psychology students are taught to respect (Neuskeptic 2012). This has not only undermined the basic trust researchers placed in each other, but also sent the message to the public that psychological findings lacked credibility. In response, some have called for a “credibility revolution” (Feldtman 2020), which not only demands heightened standards of transparency and rigor, but also mechanisms and policies to ensure that these standards are indeed followed. As Simine Vazire has put this latter point: [W]e now know that ... we need more than a “trust me” from the researcher, however sincere they are” (Vazire 2019).

Beyond Replication

We can make out several stages in, and (kinds of) responses to, the replication crisis. The most immediate response was one of shock and disbelief about scientists’ failure to apply established rules of scientific inquiry. This went hand in hand with the judgment that scientists cannot be trusted, prompting reflections about the psychological mechanisms underlying flawed reasoning (Flis 2019), the incentive structure that might be responsible for scientists engaging in questionable research practices (Heesen 2023), and the implementation of measures to prevent such practices, such as pre-registration (e.g., Scheel 2020). However, others soon argued that these responses were one-sided, suggesting (a) that replication failures were to be expected given the complexity and historicity of the subject matter (Strack 2017), (b) that replication failures can be epistemically productive (Feest 2016; Lavelle 2022), and (c) that the current debates about replication failures reveals a much deeper problem, namely a lack of theoretical, conceptual, and methodological reflection (Mutakrishna & Henrich 2019). In turn, this has given rise to work about the nature of psychological theories, exploratory experimentation, exploratory modelling, statistical methods, hypothesis testing, theory-

construction, and the underdetermination problems (among others) that present themselves to experimental psychologists. (e.g., Borsboom et al. 2021; Devezer et al 2021; Fidler et al 2018; Fried 2020; Maatman 2021; van Rooij & Baggio 2021; Rubin 2019).

Importantly, these developments take us back to some of the foundational questions that dominated crisis debates a hundred years ago, although of course before the background of a much-changed discourse in philosophy of science. One (though not the dominant) emerging theme is that questions about theory-testing are premature in the absence of sophisticated theories and that more emphasis should be paid to descriptive work (Yarkoni 2020) and to exploratory research, both experimental and formal (Rubin & Donkin 2022; Devezer & Buzbas 2023), perhaps even involving qualitative and case-study based research (Lavelle 2024). Relatedly, philosophers and sociologists have argued that replications can have epistemic functions other than theory-testing (Leonelli 2018; Guttinger 2020; Peterson & Panofsky 2021) and within psychology there are differing opinions about the function of exploratory vs. confirmatory research (Scheel et al. 2020; Rubin & Donkin 2022; Feest & Devezer forthc.). It is also becoming increasingly recognized that questions raised by the replication crisis cannot adequately be addressed without considering issues of generalizability (Yarkoni 2020; Syed 2023) and (kinds of) validity (Vazire et al. 2022). Related to questions about theoretical and conceptual work, an additional issue that has been raised recently concerns the nature of the subject matter that we take psychological theories to be about (Feest 2022; 2024).

The Entanglement of Epistemically and Ethically Questionable Research Practices

Given the debates outlined above, it is clear that we should not (blindly) “trust” psychology. It must also be noted that there is a lively discussion about how to improve the credibility of psychological research, both coming out of, and critical of, the open science movement (Nosek et al. 2015; Rubin 2023; Peterson & Panofsky 2023). Issues of trust and credibility, clearly, concern a range of questions. Can individual researchers be trusted to follow established methodological protocols? Can researchers trust other researchers’ willingness and ability to make good-faith investigative efforts? But also: Can the discipline as a whole be trusted to have picked adequate methodologies to begin with and to adequately conceptualize its subject matter?

However, there is an additional set of questions that reveal a close entanglement of ethical and epistemic issues. One such question concerns the treatment of psychological research subjects in experiments, i.e., whether subjects that participate in experiments can trust that researchers are mindful of their autonomy and well-being. Another concerns the question of whether members of various communities (especially marginalized communities) can trust that scientific results will represent them and/or be beneficial to them.

With regard to the first issue, well-known examples of unethical treatment of research subjects come to mind, such as the Milgram experiment and the Stanford prison experiment. Even though such egregious violations of trust are unlikely to occur in current psychological research, there are still concerns about the wide-spread practice of deceiving subjects about the true purpose of a study. As Kelman (1967) has pointed out already some decades ago, this ethical problem is intertwined with an epistemic one: Subjects expect to be lied to, which prompts them to engage in guesswork about the purpose of the study, thus potentially undermining the purpose of the study. Addressing this set of issues, Kelman suggests that an entirely different mode of psychological research may be needed, where researchers collaborate with their subjects. (See also Bortolotti & Mamelì 2006; Pascual-Leone et al. 2010, for other takes on the problem of deception in research).

The second question concerns the issue of whether experimental subjects are representative of the populations that researchers make inferences about. This problem was memorably formulated in terms of the worry that psychological subjects tend to be white, educated, come from industrialized countries, tend to be financially stable (rich) and tend to live in democratic societies (WEIRD) (Henrich et. al 2010), raising the worry that psychologists might be radically overgeneralizing their findings. This possibility has ethical and practical implications insofar as (a) overgeneralization in research can lead to applications (e.g., specific treatments or education programs) that are harmful or ineffective for some populations, and (b) the lack of representation of specific groups (both as researchers and as subjects) threatens the robustness of psychological research (Roberts & Mortensen 2022; Roberts 2022). This latter worry feeds into concerns that lack of plurality hampers epistemic progress and objectivity (Longino 1990). Moreover, lack of representation can also give rise to a feeling of alienation and thereby feed public distrust.

The WEIRD-problem obviously ties in with the worry about generalizability mentioned above. But it also touches on the question of how much context-sensitivity we should expect with respect to specific psychological traits and how to individuate relevant contexts, populations and traits in the first place. Within the WEIRDness-literature there seems to be an emerging consensus that it is not at all obvious what are relevant categories to consider when sampling (Ghai 2021) and the problem goes beyond the question of which populations of subjects to sample from, extending, for example, also to the choice of questions, researchers and background assumptions (Sanches de Oliveira & Baggs 2023). As such, the WEIRDness problematic also touches on foundational issues regarding the very nature of the psychological subject matter.

Conclusion

Throughout the history of psychology, and leading up to the current moment, there have been worries and debates that either directly addressed psychology's lack of credibility or at least revealed a significant degree of unease about the proper methodological foundations and scope, value-neutrality, and practical applicability of psychological research and findings. In the course of the replication crisis, these worries were extended to a range of questions concerning both the trustworthiness of individual researchers and the scientific endeavor as a whole.

The underlying contention of this research proposal is that the replication crisis has revealed deeper tensions within the discipline of scientific psychology, bringing to the fore issues of trust and credibility that converge with those raised by other debates, concerning the hidden (and not so hidden) values and assumptions that inform both foundational and applied research.

The group will engage with these debates from a variety of disciplinary perspectives (psychology, philosophy, sociology/STS). Research fellows will work on their own projects, but it is expected that there will also be synergy effects. Projects can focus on technical methodological and conceptual issues as they arise within scientific research, but can also take a broader historical, sociological, or STS view on the ways in which psychology has evolved and is currently being practiced and applied.

My working hypothesis is that the ways in which psychology has – over the years – been both invested with trust and deemed to lack credibility and trustworthiness reveals both high hopes with regard to its potential and a lot of uncertainty about a variety conceptual, methodological and practical/ethical, and practical issues. Debating and clarifying some of the foundations of this uncertainty will contribute to the (still) emerging field of philosophy of psychology and provide novel impulses to current debates within psychology.

REFERENCES

- Alexandrova, Anna (2018), Can the Science of Well-Being Be Objective? *Brit. J. Phil. Sci.* 69, 421–445
- Borsboom, Denny; van der Maas, Han L. J.; Dalege, Jonas; Kievit, Rogier A.; Haig, Brian D. (2021), Theory Construction Methodology: A Practical Framework for Building Theories in Psychology. *Perspectives on Psychological Science* 1–11. DOI: 10.1177/1745691620969647
- Bortolotti, Lisa & Mameli (2006), Deception in Psychology: Moral Costs and Benefits of Unsought Self-Knowledge. *Policies and Quality Assurance* 13(3), 1–20
- Bühler, Karl (1929), *Die Krise der Psychologie*. Fischer
- Carson (2006), *The Measure of Merit. Talents, Intelligence, and Inequality in the French and American Republics, 1750-1940*. Princeton University Press
- Devezer, Berna & Buzbas, Erkan (2023), Rigorous Exploration in a Model-Centric Science via Epistemic Iteration. *Journal of Applied Research in Memory and Cognition*. Vol. 12, No. 2, 189–194
- Devezer, Berna; Navarro, Danielle, Vandekerckhove, Joachim, Ozge Buzbas, Erkan (2021), The case for formal methodology in scientific reform. *R. Soc. Open Sci.* 8: 200805.
<https://doi.org/10.1098/rsos.200805>
- Fancher, Raymond & Rutherford, Alexandra (2011), *Pioneers of Psychology*. W. W. Norton & Company; Fourth edition
- Feest, Uljana (2016), The experimenters' regress reconsidered: Replication, tacit knowledge, and the dynamics of knowledge generation. *Studies in History and Philosophy of Science*.
<http://dx.doi.org/10.1016/j.shpsa.2016.04.003>
- Feest, Uljana, (2022), Progress in Psychology. In Shan, Yafeng (ed.): *New Philosophical Perspectives on Scientific Progress*. Routledge, 184-203
- Feest, Uljana (2024), What is the Replication Crisis a Crisis of? *Philosophy of Science* 91, 1361–1371.
doi:10.1017/psa.2024.2
- Feest, Uljana & Devezer, Berna (forthc.), Toward a More Accurate Notion of Exploratory Research (And Why it Matters). In Hutmacher, F. & Wendt, A. N. (Eds.). *Theory- and model-building in psychology*. Palgrave Macmillan.
- Feldtman, Gilad (2020), Promoting "Credibility Revolution": Journey to Open Science Collaboration, transparency, and mass mobilizing to improve current practices. *Psychology Today*.
<https://www.psychologytoday.com/intl/blog/psychologists-credibility-revolution/202005/promoting-credibility-revolution-journey-open-science>
- Fidler Fiona, Singleton Thorn F, Barnett A, Kambouris S, Kruger A. (2018), The Epistemic Importance of Establishing the Absence of an Effect. *Advances in Methods and Practices in Psychological Science*. 1(2):237-244. doi:10.1177/2515245918770407
- Fidler, Fiona & Wilcox, John (2018), Reproducibility of Scientific Results, *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2021/entries/scientific-reproducibility/>>.
- Flis, Ivan (2019), Psychologists psychologizing scientific psychology: An epistemological reading of the replication crisis. *Theory & Psychology* 29(2) 158–181

- Fried, Eiko I. (2020), Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry* 31(4) 271-288. <https://psyarxiv.com/zg84s>
- Ghai, Sakshi (2021). It's time to reimagine sample diversity and retire the WEIRD dichotomy. *Nature Human Behaviour* 5, 971–972. <https://www.nature.com/articles/s41562-021-01175-9>
- Gould, Stephen Jay (1981). *The Mismeasure of Man*. New York: W Norton and Co
- Guttinger, Stephan (2020), The limits of replicability. *European Journal for Philosophy of Science*. <https://doi.org/10.1007/s13194-019-0269-1>
- Heesen, Remco (2024), Cumulative Advantage and the Incentive to Commit Fraud in Science. *British Journal for the Philosophy of Science* 75 (3), 561-586
- Henrich, Joseph; Heine, Steven; Norenzayan, Ara (2010), The weirdest people in the world? *Behavioral and Brain Science* 33, 61–135. doi:10.1017/S0140525X0999152X
- Kelman, Herbert (1967), Human Use of Human Subjects: The Problem of Deception in Social Psychological Experiments. *Psychological Bulletin* 67(1), 1-11.
- Lavelle, J. S. (2022). When a crisis becomes an opportunity: The role of replications in making better theories. *British Journal for the Philosophy of Science*. <https://doi.org/10.1086/714812>
- Lavelle, Suilin (forthc.), Less theory, more observation: A response to psychology's 'theory crisis', *Philosophy*
- Leonelli, Sabina (2018), Rethinking Reproducibility as a Criterion for Research Quality. In *Research in the History of Economic Thought and Methodology*, Vol. 36B, 129-146. <https://doi.org/10.1108/S0743-41542018000036B009>
- Longino, Helen (1990), *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.
- Maatman, Freek Oude (2021). Psychology's Theory Crisis, and Why Formal Modelling Cannot Solve It. <https://psyarxiv.com/puqvs/>
- Muthukrishna, Michael & Henrich, Joseph (2019), A problem in theory. *Nature Human Behavior* 3, 221–229. <https://doi.org/10.1038/s41562-018-0522-1>
- Neuroskeptic (2012), Nine Circles of Scientific Hell. *Perspectives on Psychological Science* 7(6). 643-644. <https://journals.sagepub.com/doi/10.1177/1745691612459519>
- Nosek et al. (2015), Promoting an open research culture. Author guidelines for journals could help to promote transparency, openness, and reproducibility. *Science* Vol 348, Issue 6242, pp. 1422-1425. DOI: 10.1126/science.aab2374
- Pascual-Leone, A., Singh, T., & Scoboria, A. (2010), Using deception ethically: Practical research guidelines for researchers and reviewers. *Canadian Psychology / Psychologie canadienne*, 51(4), 241–248. <https://doi.org/10.1037/a0021119>
- Peterson, David & Panofsky, Aaron (2021), Self-correction in science: The diagnostic and integrative motives for replication. *Social Studies of Science* 51(4), 583-605.
- Peterson, David & Panofsky, Aaron (2023), Metascience as a scientific social movement. *Minerva*.

- Roberts, Steven Othello & Mortenson, E. M. (2022), Challenging the white = neutral framework in psychology. *Perspectives on Psychological Science*. Advanced online publication. doi: 10.1177/17456916221077117
- Roberts, Steven Othello (2022), Dealing with Diversity in Psychology: Science and Ideology. <https://osf.io/preprints/psyarxiv/xk4yu/>
- Rubin, Mark (2021), What type of Type I error? Contrasting the Neyman–Pearson and Fisherian approaches in the context of exact and direct replications." *Synthese* 198, 5809–5834
- Rubin, Mark (2023), Questionable Metascience Practices. *Journal of Trial and Error* <https://doi.org/10.36850/mr4>
- Rubin, Mark & Donkin, Chris (2022), Exploratory hypothesis tests can be more compelling than confirmatory hypothesis tests. *Philosophical Psychology*. <https://doi.org/10.1080/09515089.2022.2113771>
- Sanches de Oliveira, Guilherme & Baggs, Edward (2023), *Psychology's Weird Problems*. Cambridge University Press.
- Scheel, Anne (2020), Registered Reports: a process to safeguard high-quality evidence. *Quality of Life Research* 29(12), 3181-3182. <https://doi.org/10.1007/s11136-020-02698-w>
- Scheel, Anne; Tiokhin, Leonid; Isager, Peter; & Lakens, Daniël (2021), Why Hypothesis Testers Should Spend Less Time Testing Hypotheses. *Perspectives on Psychological Science* 16(4), 744-755.
- Strack, Fritz (2017), From Data to Truth in Psychological Science. A Personal Perspective. *Frontiers in Psychology*, Volume 8 (article 702)
- Sturm, Thomas & Mülberger, Annette (2012). Crisis discussions in psychology—New historical and philosophical perspectives. *Studies in History and Philosophy of Science* <http://dx.doi.org/10.1016/j.shpsc.2011.11.001>
- Syed, Moin (2023), Replication or Generalizability? How Flexible Inferences Uphold Unfounded Universal Claims. Paper presented at Metascience 2023 Conference (<https://osf.io/6xsmc/>)
- van Rooij, Iris & Baggio, G. (2021), Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science*. <https://journals.sagepub.com/doi/full/10.1177/1745691620970604>
- Vazire, Simine (2019), Do We Want to Be Credible or Incredible? <https://www.psychologicalscience.org/observer/do-we-want-to-be-credible-or-incredible>
- Vazire, Simine; Sarah R. Schiavone, Sarah; Bottesini, Julia (2022), Credibility Beyond Replicability: Improving the Four Validities in Psychological Science. *Current Directions in Psychological Science* 31(2), 162-168. <https://doi.org/10.1177/09637214211067779>
- Wijzen, L. D., Borsboom, D., & Alexandrova, A. (2022), Values in Psychometrics. *Perspectives on Psychological Science* 17(3), 788-804. <https://doi.org/10.1177/17456916211014183>
- Wijzen, Lisa D.; Borsboom, Denny & Alexandrova, Anna (2022). Values in Psychometrics. *Perspectives on Psychological Science* 17(3), 788-804. <https://doi.org/10.1177/17456916211014183>
- Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences*, 1–37. <https://doi.org/10.1017/S0140525X20001685>