

Responsible Research is also concerned with generalizability: Recognizing efforts to reflect upon and increase generalizability in hiring and promotion decisions in psychology

Roman Stengelin^{1,2}, Manuel Bohn¹, Alejandro Sánchez-Amaro¹, Daniel B.M. Haun^{1,3,4}, Maleen Thiele¹, Matthias Allritz¹, Moritz M. Daum^{5,6}, Elisa Felsche¹, Frankie T.K. Fong^{1,7}, Anja Gampe⁸, Marta Giner Torrén⁹, Sebastian Grueneisen^{3,10}, David J.K. Hardecker^{1,3}, Lisa Horn¹¹, Karri Neldner^{1,4}, Sarah Pope-Caldwell¹, and Nils Schuhmacher⁹

¹Department of Comparative Cultural Psychology, MPI EVA, Leipzig, GER

²Department of Psychology and Social Work, University of Namibia, Windhoek, NAM

³Leipzig Research Center for Early Child Development, Leipzig University, GER

⁴LeipzigLab, Leipzig University, GER

⁵Department of Psychology, University of Zurich, SUI

⁶Jacobs Center for Productive Youth Development, University of Zurich, SUI

⁷Early Cognitive Development Centre, University of Queensland, Brisbane, AUS

⁸Institute for Socio-Economics, University of Duisburg-Essen, GER

⁹Department of Psychology, University of Muenster, GER

¹⁰Faculty of Education, Leipzig University, GER

¹¹Department of Behavioral and Cognitive Biology, University of Vienna, AUT

We concur with the authors of the two target articles that Open Science practices can help combat the ongoing reproducibility and replicability crisis in psychological science and should hence be acknowledged as responsible research practices in hiring and promotion decisions. However, we emphasize that another crisis is equally threatening the credibility of psychological science in Germany: The sampling or generalizability crisis. We suggest that scientists' efforts to contextualize their research, reflect upon, and increase its generalizability should be incentivized as responsible research practices in hiring and promotion decisions. To that end, we present concrete suggestions for how efforts to combat the additional generalizability crisis could be operationalized within Gärtner et al. (2022) evaluation scheme. Tackling the replicability and the generalizability crises in tandem will advance the credibility and quality of psychological science and teaching in Germany.

Keywords: Generalizability Crisis, Sampling Crisis, Responsible Research, Cultural Psychology

Gärtner et al. (2022) and Schönbrodt et al. (2022) advocate for a greater consideration of responsible research practices in hiring and promotion decisions in Germany. Building upon the San Francisco Declaration on Research Assessment (DORA), they propose to incentivize Open Science practices in Psychological Science by including assessments of such practices when evaluating candidates for academic positions. More generally, the authors suggest prioritizing the quality rather than quantity of publications while also considering other scientific outputs.

We agree with the authors that greater incentives for quality over quantity and encouragement of Open Sci-

ence practices are much needed to respond to the replicability crisis. However, we flag another fundamental crisis threatening the credibility and quality of psychological science they left mostly unattended: the generalizability crisis (Arnett, 2008; Henrich et al., 2010; Simons et al., 2017).

Psychological science almost exclusively relies on participants from a thin slice of humanity: Formally-educated, urban, middle to upper class communities from the wealthy Global North, such as the United States or Germany (Muthukrishna et al., 2020). These communities are rarely approached for theoretical reasons, but predominantly for convenience: Scientists tend to study participants they can recruit with rela-

tively low effort and cost. This leads to a drastic overrepresentation of psychology students from local universities across cognitive, personality, and social psychology (Arnett, 2008; i.e., “the science of the behavior of sophomores” in McNemar, 1946; Sears, 1986) or strong bias towards children from formally educated, middle-class communities in developmental psychology (Nielsen et al., 2017). All of this would be less problematic, if the authors gave explicit information about which populations their research conclusions are based on and apply to. However, this is rarely the case: data is frequently interpreted and presented as if it applies to much larger populations and, often, humans in general. Scholars rarely contextualize their research and make generalizability concerns explicit. Of course, some research is not meant to generalize beyond the population from which the sample is drawn. This is, however, the exception and not the norm and should be communicated as such.

The habitual reliance on convenience sampling and the widespread tendency to assume generalizability from such data have drastic consequences: Psychological science is built upon participants who are outliers on many cultural metrics known to guide human behavior and experience (Henrich et al., 2010). They come from predominantly White (Remedios, 2022; Roberts et al., 2020), ethnically homogenous (Drazanova, 2019), individualistic (Schulz et al., 2018), western, educated, industrialized, rich, and democratic (“WEIRD”) communities (Henrich et al., 2010). Ad-hoc generalizations from such peculiar participants are inadequate. Moreover, impact-related publication incentives towards broad and universal claims lead scholars to portray their research as robust and generalizable, discounting effects of cultural background (Castro Torres & Alburez-Gutierrez, 2022; Roberts et al., 2020).

In contrast, samples outside these focal convenience communities often require justification. This double standard encourages biased research participation, evaluation, and impact (Castro Torres & Alburez-Gutierrez, 2022; Kahalon et al., 2022) and feeds into deficit or non-normative models of communities outside “standard” convenience samples (Forbes et al., 2022; Scheidecker et al., 2022). As of today, the field’s reluctance to situate and reflect upon its participants perpetuates global disparities in scientific knowledge production and representation (Draper et al., 2022). As we outline below, we propose that appropriate contextualization of psychological research, attempts to test and increase generalizability, and discussions of limitations to generalizability are responsible research practices that help address this crisis. The replicability and generalizability crisis share some key features: Both be-

came relevant to a broader audience around the same time (Arnett, 2008; Henrich et al., 2010; Schmidt, 2009; Simmons et al., 2011; Syed, 2022), and – in both cases – effective countermeasures have been put forward. On the downside, both movements have, until today, received some skepticism, ignorance, and even resistance. To overcome the status quo, changes need to be made on a science-policy level (Doebel & Frank, 2022; Nielsen et al., 2017; Schönbrodt et al., 2022). Generalizability issues need to be accounted for when assessing the replicability of psychological research, and vice versa (Fischer & Poortinga, 2018; Milfont & Klein, 2018; Syed & Kathawalla, 2021). It is thus surprising that the replicability crisis and the generalizability crisis have hitherto barely engaged with one another (Syed & Kathawalla, 2021). Cultural perspectives and adequate generalizations are foundational to psychological science (Fahrenberg, 2016; Wundt, 1906). Contrastingly, generalizability issues are often treated as relevant only for specific subfields of psychology (e.g., (cross-)cultural psychology, comparative psychology) with associated journals, conferences, and scientific societies. In result, there has long been a drastic underrepresentation, or avoidance, of cultural perspectives in psychological science (Haun et al., 2020; Helfrich, 2021). In Germany, dedicated professorships or junior groups researching culture or generalizability are almost absent, as are synergies with closely related disciplines, such as anthropology or ethnography (see also Wissenschaftsrat, 2018). In consequence, cultural and generalizability issues are underrepresented in research and teaching in Germany.

A final parallel between the replicability and generalizability crises are the additional efforts researchers face when attempting to mitigate them. For both crises, some measures can easily be undertaken by all, such as by contextualizing research in scientific publications and teaching or adopting Open Science practices. Other measures require substantial devotion: for example, building and maintaining scientific infrastructure to increase the replicability or generalizability of psychological science. Today, efforts to contextualize research and improve generalizability are barely incentivized in funding schemes, hiring decisions or publication processes. Concerning the replicability crisis, Schönbrodt et al. (2022) show why such efforts are important and provide practical recommendations for how they should be recognized. We advocate that similar steps be undertaken to reflect upon and promote generalizability in psychological science in Germany. Next, we provide concrete recommendations on how this could be achieved during hiring and promotion decisions. Our recommendations could be incorporated into the eval-

uation scheme proposed by Gärtner et al. (2022). We outline three primary practices relating to generalizability that can be implemented by all psychological scientists, but also flag how more effortful and structural investments could be considered as scientific contributions beyond the proposed publication formats.

A first criterion would be whether researchers contextualize their research by providing relevant details about the participants and describing how the tested sample relates to the research question and methodology. For any research involving human participants, scholars can provide cultural metrics and ethnographic details that may affect participants' performance in the research. Which information is required depends on the research and should hence be informed by theory. A second criterion would be to include dedicated constraints on generality statements discussing the scope of research explicitly (Simons et al., 2017). Note that both these steps can help assess and increase the replicability of psychological science by making the target populations explicit. A third criterion would be to invest efforts into collecting data that tests or fosters generalizability (Doebel & Frank, 2022). The efforts invested here may vary depending on the research approach (Lakens et al., 2022) and be graded correspondingly. Some findings may already benefit from adding another convenience sample including different language speakers, or participants from more diverse socio-economic backgrounds. Others may involve testing individuals from multiple, culturally diverse small-scale societies (e.g., Blake et al., 2015; House et al., 2020; van Leeuwen et al., 2018). Other research may benefit from data analytic approaches promoting generalizability (e.g., Deffner et al., 2022).

These three criteria aim to incentivize responsible research practices by contextualizing research as well as discussing and fostering its generalizability. Such efforts can be undertaken by all psychological scientists and could thus be added as evaluation criteria for publications in the scheme proposed by Gärtner et al. (2022). Other contributions are difficult to assess on the level of single publications, particularly for researchers contributing to sustainable infrastructures dedicated to improving the generalizability of psychological science more generally. Examples for this are collaborative networks like the Psychological Science Accelerator (Moshontz et al., 2018), ManyLabs (Klein et al., 2014), or ManyPrimates (Primates et al., 2019). Others may build and maintain research infrastructure with underrepresented communities and invite external scientists to collaborate and increase the generalizability of their work. Such contributions exceed the scope of single publications, but are central to the problem at

hand. We suggest adding efforts and documentation related to such infrastructures as alternative research outputs to those proposed by Gärtner et al. (2022). This would ensure that hiring and promotion committees in psychology could account for the diversity with which scholars contribute to pressing issues in psychological science.

The fundamental importance of culture in enabling and constraining human behavior and cognition is deeply rooted in the history of psychological science in Germany (Wundt, 1906). However, current practice in the field rarely incentivizes but even discourages scholars from grappling with their participants and the resulting generalizability of their research. To combat the status quo, action needs to be taken on a science-policy level. This includes hiring and promotion decisions in Germany and other countries. We hope this comment serves as a starting point to think about the two fundamental crises of psychological science as one: Responsible research in psychology is concerned with replicability and generalizability.

Author Contact

roman_stengelin@eva.mpg.de

RS: <https://orcid.org/0000-0003-2212-4613>

MB: <https://orcid.org/0000-0001-6006-1348>

ASA: <https://orcid.org/0000-0003-4036-2455>

DH: <https://orcid.org/0000-0002-3262-645X>

MT: <https://orcid.org/0000-0002-1695-1850>

MA: <https://orcid.org/0000-0003-2694-3261>

MD: <https://orcid.org/0000-0002-4032-4574>

EF: <https://orcid.org/0000-0001-9996-8158>

FF: <https://orcid.org/0000-0002-6135-1379>

AG: <https://orcid.org/0000-0001-9812-9694>

MGT: <https://orcid.org/0000-0003-4945-5551>

SG: <https://orcid.org/0000-0002-0888-9102>

DJKH: <https://orcid.org/0000-0001-7897-2967>

LH: <https://orcid.org/0000-0002-9586-915X>

KN: <https://orcid.org/0000-0002-8237-5679>

NS: <https://orcid.org/0000-0002-3927-2447>

Conflict of Interest and Funding

The authors declare no conflict of interest. Writing this commentary was supported by the Max Planck Society for the Advancement of Science.

Author Contributions

RS: Conceptualization, Writing - Original Draft, Project Administration

MB: Conceptualization, Writing - Original Draft

ASA: Conceptualization, Writing - Original Draft

DH: Conceptualization, Writing - Review & Editing

MT: Conceptualization, Writing - Review & Editing
 MA: Writing - Review & Editing
 MD: Writing - Review & Editing
 EF: Writing - Review & Editing
 FF: Writing - Review & Editing
 AG: Writing - Review & Editing
 MGT: Writing - Review & Editing
 SG: Writing - Review & Editing
 DJKH: Writing - Review & Editing
 LH: Writing - Review & Editing
 KN: Writing - Review & Editing
 SPC: Writing - Review & Editing
 NS: Writing - Review & Editing

Open Science Practices

This article is theoretical and not eligible for any Open Science badges. The entire editorial process, including the open reviews, is published in the online supplement.

References

- Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, 63(7), 602–614. <https://doi.org/10.1037/0003-066X.63.7.602>
- Blake, P. R., McAuliffe, K., Corbit, J., Callaghan, T. C., Barry, O., Bowie, A., Kleutsch, L., Kramer, K. L., Ross, E., Vongsachang, H., Wrangham, R., & Warneken, F. (2015). The ontogeny of fairness in seven societies. *Nature*, 528(7581), 258–261. <https://doi.org/10.1038/nature15703>
- Castro Torres, A. F., & Albrez-Gutierrez, D. (2022). North and south: Naming practices and the hidden dimension of global disparities in knowledge production. *Proceedings of the National Academy of Sciences*, 119(10), e2119373119. <https://doi.org/10.1073/pnas.2119373119>
- Deffner, D., Rohrer, J. M., & McElreath, R. (2022). A causal framework for cross-cultural generalizability. *Advances in Methods and Practices in Psychological Science*, 5(3), 251524592211063. <https://doi.org/10.1177/25152459221106366>
- Doebel, S., & Frank, M. C. (2022). Broadening convenience samples to advance theoretical progress and reduce bias in developmental science [preprint]. *PsyArXiv*. <https://doi.org/10.31234/osf.io/s8zxm>
- Draper, C. E., Barnett, L. M., Cook, C. J., Cuartas, J. A., Howard, S. J., McCoy, D. C., Merkley, R., Molano, A., Maldonado-Carreño, C., Obradović, J., et al. (2022). Publishing child development research from around the world: An unfair playing field resulting in most of the world's child population under-represented in research. *Infant and Child Development*. <https://doi.org/10.1002/icd.2375>
- Drazanova, L. (2019). *Historical index of ethnic fractionalization dataset (hief) [data set]*. <https://doi.org/10.7910/DVN/4JQRCL>
- Fahrenberg, J. (2016). *Wilhelm wundts kulturpsychologie (völkerpsychologie): Eine psychologische entwicklungstheorie des geistes*. <https://doi.org/10.23668/PSYCHARCHIVES.10919>
- Fischer, R., & Poortinga, Y. H. (2018). Addressing methodological challenges in culture-comparative research. *Journal of Cross-Cultural Psychology*, 49(5), 691–712. <https://doi.org/10.1177/0022022117738086>
- Forbes, S. H., Aneja, P., & Guest, O. (2022). The myth of normative development. *Infant and Child Development*. <https://doi.org/10.1002/icd.2393>
- Gärtner, A., Leising, D., & Schönbrodt, F. D. (2022). Responsible research assessment ii: A specific proposal for hiring and promotion in psychology [preprint]. *PsyArXiv*. <https://doi.org/10.31234/osf.io/5yexm>
- Haun, D. B. M., Liebal, K., Amici, F., et al. (2020). Ein plädoyer für die relevanz der vergleichenden psychologie für das verständnis menschlicher entwicklung. *Psychologische Rundschau*, 71(1), 40–41. <https://doi.org/10.1026/0033-3042/a000466>
- Helfrich, H. (2021). Cross-cultural psychology in germany. In H. Grad, A. Blanco, & J. Georgas (Eds.), *Key issues in cross-cultural psychology* (1st, pp. 52–67). Garland Science. <https://doi.org/10.1201/9781003077442-6>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- House, B. R., Kanngiesser, P., Barrett, H. C., Broesch, T., Cebiglu, S., Crittenden, A. N., Erut, A., Lew-Levy, S., Sebastian-Enesco, C., Smith, A. M., et al. (2020). Universal norm psychology leads to societal diversity in prosocial behaviour and development. *Nature Human Behaviour*, 4(1), 36–44. <https://doi.org/10.1038/s41562-019-0734-z>

- Kahalon, R., Klein, V., Ksenofontov, I., Ullrich, J., & Wright, S. C. (2022). Mentioning the sample's country in the article's title leads to bias in research evaluation. *Social Psychological and Personality Science*, 13(2), 352–361. <https://doi.org/10.1177/19485506211024036>
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., et al. (2014). Investigating variation in replicability. *Social Psychology*.
- Lakens, D., Uygun Tunç, D., & Necip Tunç, M. (2022). There is no generalizability crisis. *Behavioral and Brain Sciences*, 45, e25. <https://doi.org/10.1017/S0140525X21000340>
- McNemar, Q. (1946). Opinion-attitude methodology. *Psychological Bulletin*, 43, 289–374.
- Milfont, T. L., & Klein, R. A. (2018). Replication and reproducibility in cross-cultural psychology. *Journal of Cross-Cultural Psychology*, 49(5), 735–750. <https://doi.org/10.1177/0022022117744892>
- Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J., Musser, E. D., Antfolk, J., et al. (2018). The psychological science accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science*, 1(4), 501–515.
- Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond western, educated, industrial, rich, and democratic (weird) psychology: Measuring and mapping scales of cultural and psychological distance. *Psychological Science*, 31(6), 678–701. <https://doi.org/10.1177/0956797620916782>
- Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, 162, 31–38. <https://doi.org/10.1016/j.jecp.2017.04.017>
- Primates, M., Altschul, D. M., Beran, M. J., Bohn, M., Call, J., DeTroy, S., Duguid, S. J., Egelkamp, C. L., Fichtel, C., Fischer, J., Flessert, M., et al. (2019). Establishing an infrastructure for collaboration in primate cognition research. *PLOS ONE*, 14(10), e0223675. <https://doi.org/10.1371/journal.pone.0223675>
- Remedios, J. D. (2022). Psychology must grapple with whiteness. *Nature Reviews Psychology*, 1(3), 125–126. <https://doi.org/10.1038/s44159-022-00024-4>
- Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D., & Mortenson, E. (2020). Racial inequality in psychological research: Trends of the past and recommendations for the future. *Perspectives on Psychological Science*, 15(6), 1295–1309. <https://doi.org/10.1177/1745691620927709>
- Scheidecker, G., Chaudhary, N., Oppong, S., Röttger-Rössler, B., & Keller, H. (2022). Different is not deficient: Respecting diversity in early childhood development. *The Lancet Child & Adolescent Health*, 6(12), e24–e25. [https://doi.org/10.1016/S2352-4642\(22\)00277-2](https://doi.org/10.1016/S2352-4642(22)00277-2)
- Schmidt, S. (2009). Shall we really do it again? the powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(2), 90–100. <https://doi.org/10.1037/a0015108>
- Schönbrodt, F. D., Gärtner, A., Frank, M., Gollwitzer, M., Ihle, M., Mischkowski, D., Phan, L. V., Schmitt, M., Scheel, A. M., Schubert, A.-L., et al. (2022). Responsible research assessment i: Implementing dora for hiring and promotion in psychology [preprint]. *PsyArXiv*. <https://doi.org/10.31234/osf.io/rgh5b>
- Schulz, J., Bahrami-Rad, D., Beauchamp, J., & Henrich, J. (2018). The origins of weird psychology. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3201031>
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, 51(3), 515–530. <https://doi.org/10.1037/0022-3514.51.3.515>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (cog): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, 12(6), 1123–1128. <https://doi.org/10.1177/1745691617708630>
- Syed, M. (2022). Three myths about open science that just won't die [preprint]. *PsyArXiv*. <https://doi.org/10.31234/osf.io/w8xs2>
- Syed, M., & Kathawalla, U.-K. (2021). Cultural psychology, diversity, and representation in open sci-

- ence. In K. C. McLean (Ed.), *Cultural methods in psychology* (pp. 427–454). Oxford University Press. <https://doi.org/10.1093/oso/9780190095949.003.0015>
- van Leeuwen, E. J., Cohen, E., Collier-Baker, E., Rapold, C. J., Schäfer, M., Schütte, S., & Haun, D. B. (2018). The development of human social learning across seven societies. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-04468-2>
- Wissenschaftsrat. (2018). *Perspektiven der psychologie in deutschland*.
- Wundt, W. M. (1906). *Völkerpsychologie: Bd., 1-3. t. mythus und religion. 1905-09*. W. Englemann.