Meta-Psychology, 2025, vol 9, MP.2024.4155
https://doi.org/10.15626/MP.2024.4155
Article type: Commentary

Open data: Not Applicable
Open materials: Not Applicable
Open and reproducible analysis: Not Applicable
Open reviews and editorial process: Yes
Preregistration: No

Edited by: Rickard Carlsson
Reviewed by: Lukas Sotola, Leonhard Held
Analysis reproduced by: Not Applicable
All supplementary files can be accessed at OSF:
https://doi.org/10.17605/OSF.IO/Z34JK

Pre-replication: Anything Goes, Once

Veli-Matti Karhulahti[1], Marcel Martončik[1,2], and Matúš Adamkovič[1,2,3]
[1]University of Jyväskylä
[2]Slovak Academy of Sciences
[3]Charles University

In this commentary, we propose pre-replication as a necessary epistemic, analytic, and axiologic exercise to be carried out before each replication.

Keywords: axiology, epistemology, methodology

Introduction

Replication is a key part of many sciences. Because well-evidenced replications often need even more data and meticulous planning than original studies, replications are also laborious and expensive (Giles, 2006; Isager et al., 2023; Nosek & Errington, 2020). Evidently, scientists should carefully consider which findings[1] to replicate and how, and what to infer from them (e.g. Buzbas and Devezer, 2023; Isager et al., 2023). In this commentary, we propose *pre-replication* as a necessary epistemic, analytic, and axiologic exercise to be performed before each replication. In pre-replication, researchers reflect on:

1. What are the epistemic goals in replication and how is the original study related to them? (Epistemological assessment)
2. Is the chosen finding or study sufficiently high in methodological quality for replication? (Methodological assessment)
3. Why has replication been chosen to reach the desired epistemic goals? (Value assessment)

We address each of these points respectively below.

Epistemic assessment

Two central epistemic goals for scientific activities are to advance *theory* and *practice* (e.g., Dienes, 2023). For example, if a study reports a surprising finding, which is not consistent with prevailing theories—this can be a null finding too—it may be worth re-running the study (for theoretical reasons); if an intervention reports effectiveness, it may be worth re-running it to increase trust or improve predictions (for practical reasons). Effectiveness alone is not a pragmatic goal, however. To be meaningful in practice, for example, human experiments and interventions should have the potential for a minimal clinically important difference, i.e. the smallest benefit of value for humans (e.g. Jaeschke et al., 1989; McGlothlin and Lewis, 2014). Before conducting a replication, researchers should outline what specific theoretical or practical epistemic goals their work will address and how.

Alternatively, if replication does not serve theory or practice, researchers should carefully explain their alternative rationale. As Longino (2002) notes, paraphrasing Popkin, "if all one wants is knowledge or truth, why not count the number of bottle caps one can lay down between Los Angeles and San Diego?" (p. 176). In the same fashion, already three decades ago, Altman (1994) was concerned about how "useless studies pass [ethical] review even though they can reasonably be considered to be unethical" (p. 283)—and the same principle applies to replication. It can never be perfectly predicted what pieces of information turn out to be meaningful in the future, but replicating a finding without being able to justify its scientific or practical meaning is to double down on a lost gamble where funders, participants, and public trust are the ante.

Methodological assessment

Assessing methodology serves two main purposes here. First, replicating findings from poorly designed

[1]We consider replication as an attempt to reassess confidence in an earlier finding. Note that 'studies' are not the same as 'findings', and *researchers should not replicate studies, but findings*. Replication of a valuable finding should be carried out through a research design that is optimal for reevaluating the finding, irrespective of the extent to which that design is different or similar to the original study. Two points follow this stance. First, whether studies replicate or not is of little importance to science; what matters is whether *findings* replicate. Second, each replication of a finding is 'conceptual', since, by definition, the finding is the conceptual phenomenon being replicated. The directness of a replication (that is, the technical distance between the original and replication designs) is not as relevant as a careful assessment of the replication's ability to reassess confidence in an identified finding (for a relevant discussion, see Buzbas and Devezer, 2024).

studies may not even contribute to the null literature Scheel (2022) but might further contaminate science by generating misleading knowledge (e.g., Rotello et al., 2015). For a finding to be considered for replication, it should come from a design that was sufficiently strong for its findings to be worth attention. Second, if the methodology of a study is sufficient for its findings to be worth attention and in need of replication, the replication should be designed with the priority given to reassess the *finding*, not the *study* (Footnote 1). Methodological weakness should be improved to prevent them from multiplying. As Zarin et al. (2019) assert, in most cases "preventable uninformativeness is a serious breach of trust and a violation of research ethics" (for a review of informativeness also involving epistemic matters, see their eTable 1).

The methodological assessment above should be carried out with appropriate expertise on the finding (White & Barnett, 2025). In the field of medicine, Freedman (1987) adds that assessment of validity also involves "background knowledge of the clinical and bench science required to judge whether the assumptions [are] reasonable" (p. 9). Indeed, methodology can unlikely be comprehensively assessed, for instance, by statistical elements alone but the design also depends on topic expertise. Whether a design can contribute to relevant epistemic goals with appropriate methodology should be evaluated by researchers with expertise in the appropriate topics *and* methods.

Value assessment

Replicating a finding may not be the most optimal means of achieving epistemic goals. Whatever theoretical, pragmatic, or other goal is being pursued, replicators should always justify *why* replication has been chosen over alternative means to pursue those goals. For example, researchers with epistemic goals related to the advancement of a specific theory might benefit more from addressing the unexplored aspects of that theory. It is beyond pre-replication to provide a system of axiological quantification (cf. Weinberg, 1970); rather, its function is to encourage replicators to reflect on their values. Choosing to replicate findings of a study implies that those specific findings have been considered most valuable in terms of one's epistemic goals, given the pragmatic context.

Although the value of replication cannot be fully detached from the larger discourse on values in science (Merton, 1973), researchers must essentially evaluate each replication case-by-case: what value is expected to be gained from re-running the study, and why doing that is superior to a second or third alternative? If a finding is sufficiently costly and strong, would it make more sense to design new studies for application and generalisation? An axiologic statement of justification for choosing a replication can show that researchers have *at least* reflected on the values in question.

Pre-replication

Pre-replication is a multidimensional assessment of a finding that is planned to be replicated. Authors may include a pre-replication report in their study by explicitly addressing epistemic, analytic, and axiologic assessments in their replication. However, the primary purpose of pre-replication should be to make researchers critically reflect on their work *before* having started it. The more time and resources one invests on replicating a finding, the more difficult it may become to abandon or change the plan. If pre-replication reveals a finding to be scientifically meaningless or low in meaning, a successful outcome is to conclude it *not worth replication*—before replication. A justification of lacking meaning can falsify or question findings without replication, as can reflections on related methodology and contextual value.

Optimally, in-text pre-replications thus serve as compact, well-argued justifications of the meaningfulness of a specific replication project in conversation with epistemic goals, methodology, and values. Naturally, pre-replication does not protect researchers from lacking reflection and biases—as do not other related statements (Karhulahti, 2024)—but makes these issues more transparent. Moreover, pre-replication is not a tool for evaluating which studies to replicate in general, but it merely helps researchers to assess whether a particular study meets the minimum criteria for replication, either for specific researchers or the scientific community more widely.

If the pre-replication of a study reveals its finding to be meaningless for general epistemic, methodological, or value reasons, replicators could publish their findings as *negative pre-prelication reports*, similar in length to 'corrections' and 'expressions of concern' to briefly communicate such issues. Examples of negative pre-replication reports are provided in the supplement (Appendix 1). Following the pottery barn rule "you break it, you buy it," it has been suggested that journals should invite replications of findings that they have previously published (Srivastava in Chambers, 2017). Likewise, journals could also welcome negative pre-replication reports that allow replicators to reassess findings not only empirically but also by critical investigation. The discovery of lacking scientific meaning can be a meaningful contribution to science.

Author Contact

Corresponding author: Veli-Matti Karhulahti, vmkarhwu@jyu.fi ORCID - Veli-Matti Karhulahti: 0000-0003-3709-5341 ORCID - Marcel Martončik: 0000-0003-4869-6900 ORCID - Matúš Adamkovič: 0000-0002-9648-9108

Conflict of Interest and Funding

MM and MA have been involved in several replication studies. VMK has never carried out a replication study. MM and MA are specifically trained in psychology, whereas VMK is an interdisciplinary social scientist with a background in philosophy. These lineages limit our experiences in science, and while we intend pre-replication as a concept for various scientific fields where replications occur, we cannot know how possible caveats might interfere with pre-replication in all fields. All authors are part of their local reproducibility networks.

Funded by the European Union (ERC, ORE, 101042052). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. The study was supported by project PRIMUS/24/SSH/017 and by the Slovak Research and Development Agency under contract no. APVV-22-0458 and APVV-23-0421.

Author Contributions

VMK: Conceptualization, Funding Acquisition, Investigation, Methodology, Writing – Original Draft; MM: Validation, Writing – Review & Editing; MA: Validation, Writing – Review & Editing.

Open Science Practices

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

References

Altman, D. G. (1994). The scandal of poor medical research. *BMJ*, *308*(6924), 283–284. https://doi.org/10.1136/bmj.308.6924.283

Buzbas, E. O., & Devezer, B. (2023). Tension between theory and practice of replication. *Journal of Trial & Error*, *4*(1). https://doi.org/10.36850/mr9

Buzbas, E. O., & Devezer, B. (2024). Statistics in service of metascience: Measuring replication distance with reproducibility rate. *Entropy*, *26*(10), 842.

Chambers, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice*. Princeton University Press.

Dienes, Z. (2023). The pragmatics of statistical inference. In D. Trafimow (Ed.), *Research handbook on the replication crisis*. Edward Elgar Publishing, Inc.

Freedman, B. (1987). Scientific value and validity as ethical requirements for research: A proposed explication. *IRB: Ethics & Human Research*, *9*(6), 7–10. https://doi.org/10.2307/3563623

Giles, J. (2006). The trouble with replication. *Nature*, *442*(7101), 344–348. https://doi.org/10.1038/442344a

Isager, P. M., Van Aert, R., Bahník, Š., Brandt, M. J., DeSoto, K. A., Giner-Sorolla, R., Lakens, D., et al. (2023). Deciding what to replicate: A decision model for replication study selection under resource and knowledge constraints. *Psychological Methods*, *28*(2), 438. https://doi.org/10.1037/met0000438

Jaeschke, R., Singer, J., & Guyatt, G. H. (1989). Measurement of health status: Ascertaining the minimal clinically important difference. *Controlled clinical trials*, *10*(4), 407–415. https://doi.org/10.1016/0197-2456(89)90005-6

Karhulahti, V. M. (2024). Positionality statements in science. *Open Research Europe*, *4*, 62. https://doi.org/10.12688/openreseurope.17058.2

Longino, H. E. (2002). *The fate of knowledge*. Princeton University Press.

Martončik, M., Babjáková, J., Čupková, L., Köverová, N., & Kačmárová, M. (2020). Simulation of vocational activities: Experimental evidence of (no) changes in vocational interests. *Empirical Research in Vocational Education and Training*, *12*(1). https://doi.org/10.1186/s40461-020-00093-w

McGlothlin, A. E., & Lewis, R. J. (2014). Minimal clinically important difference: Defining what really matters to patients. *JAMA*, *312*(13), 1342–1343. https://doi.org/10.1001/jama.2014.13128

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. The University of Chicago Press.

Ng, K., Kaskinen, A. P., Katila, R., Koski, P., & Karhulahti, V. M. (2022). Associations between sports videogames and physical activity in children. *Physical Culture and Sport. Studies and Re-*

search, *95*(1), 68–75. https://doi.org/10.2478/pcssr-2022-0012

Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLoS biology*, *18*(3), e3000691. https://doi.org/10.1371/journal.pbio.3000691

Rajčáni, J., Vargová, L., Adamkovič, M., & Kačmár, P. (2023). Statistical misconceptions, awareness, and attitudes towards open science practices in slovak psychology researchers. *Studia Psychologica*, *65*(1), 56–70. https://doi.org/10.31577/sp.2023.01.866

Rotello, C. M., Heit, E., & Dubé, C. (2015). When more data steer us wrong: Replications with the wrong dependent measure perpetuate erroneous conclusions. *Psychonomic Bulletin & Review*, *22*, 944–954. https://doi.org/10.3758/s13423-014-0759-2

Scheel, A. M. (2022). Why most psychological research findings are not even wrong. *Infant and Child Development*, *31*(1), e2295. https://doi.org/10.1002/icd.2295

Weinberg, A. M. (1970). The axiology of science: The urgent question of scientific priorities has helped to promote a growing concern with value in science. *American Scientist*, *58*(6), 612–617.

White, N., & Barnett, A. (2025). Appropriate statistical methods are necessary for ethical medical research. *BMJ*, *388*, r3. https://doi.org/10.1136/bmj.r3

Zarin, D. A., Goodman, S. N., & Kimmelman, J. (2019). Harms from uninformative clinical trials. *JAMA*, *322*(9), 813–814. https://doi.org/10.1001/jama.2019.9892

Appendix

We provide three examples of negative pre-replication reports using our own earlier published studies (one from each of us) to avoid involving external researchers to invest labor in reviewing/replying. Note that positive pre-replications should be integrated into replication studies when assessment supports carrying out a replication, and for this reason we do not provide examples of positive pre-replications.

Pre-replication of: Martončik, M., Babjáková, J., Čupková, L., Köverová, N., & Kačmárová, M. (2020). Simulation of vocational activities: Experimental evidence of (no) changes in vocational interests. *Empirical Research in Vocational Education and Training*, *12*(1). https://doi.org/10.1186/s40461-020-00093-w

The aim of the study was to test if exposing adolescents to simulated vocational activities would trigger changes in their vocational interests. This builds upon the assumptions of several career change theories, including that of Holland's (1973) and the *four-phase model of interest development* (Hidi and Renninger, 2006). The results were null. Replicating the findings via this design would not be informative for two reasons. First, the four-phase model of interest development would be better explored by capturing triggered situational interest immediately after the intervention rather than after several days. Second, the applied paper-based simulation did not resemble vocational activities well and was thus a weak method for the chosen epistemic goals. For instance, current 3D and VR technologies enable more truthful simulation of professional activities—researchers interested in the mentioned models and theories will find more value in such interventions instead.

Pre-replication of: Ng, K., Kaskinen, A. P., Katila, R., Koski, P., & Karhulahti, V. M. (2022). Associations between sports videogames and physical activity in children. *Physical Culture and Sport. Studies and Research*, *95*(1), 68–75. https://doi.org/10.2478/pcssr-2022-0012

The study's core finding is that children's *sports gaming* (FIFA, NHL, etc) appears to be associated with physical activity, unlike other gaming. This aligns with the theory of social worlds: children who are active in team sports (football, hockey etc.) belong to their respective social worlds, and sports gaming is part of those worlds (Koski 2008). Although the data are tentative, a replication is likely to be meaningless, as the finding has already been remade in two large datasets from the same country (Ng et al. 2023; Ng et al. 2024). To better understand the degree to which social words theory explains physical activity and technology use, it would be more valuable to investigate whether the phenomenon generalises to adults or other countries with different social worlds. It would also be more valuable to model how diverse sports and corresponding gaming products differ in this regard.

Pre-replication of: Rajčáni, J., Vargová, L., Adamkovič, M., & Kačmár, P. (2023). Statistical misconceptions, awareness, and attitudes towards open science practices in slovak psychology researchers. *Studia Psychologica*, *65*(1), 56–70. https://doi.org/10.31577/sp.2023.01.866

The study examined statistical misconceptions among Slovak psychology researchers. A key finding is that researchers frequently misinterpret statistics—no participant achieved a perfect score of 14/14 on a quiz-based survey, with the average number of correct responses falling below 40%. Similar rates have been documented in other populations (Hoekstra et al., 2014; Lytsy et al., 2022). The design led to an overrepresentation of academics who teach methods and statistics courses, for which replicating the design as such would be uninformative regarding more general rates. Due to the recent popularisation of large language models, replicating the study as an online-survey would be increasingly unreliable. Future efforts to map out statistical expertise could be combined with interventions, which can simultaneously obtain data on baseline levels and potential improvements.