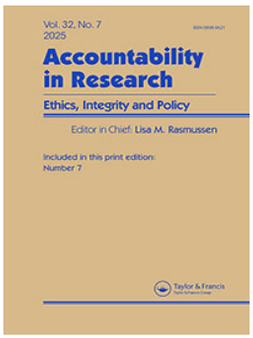

Taylor & Francis
Taylor & Francis Group

Accountability in Research
Ethics, Integrity and Policy

ISSN: 0898-9621 (Print) 1545-5815 (Online) Journal homepage: www.tandfonline.com/journals/gacr20

Replication studies in the Netherlands: Lessons learned and recommendations for funders, publishers and editors, and universities

Maarten Derksen, Stephanie Meirmans, Jonna Brenninkmeijer, Jeannette Pols, Annemarijn de Boer, Hans van Eyghen, Surya Gayet, Rolf Groenwold, Dennis Hernaus, Pim Huijnen, Nienke Jonker, Renske de Kleijn, Charlotte F. Kroll, Angelos-Miltiadis Krypotos, Nynke van der Laan, Kim Luijken, Ewout Meijer, Rachel S. A. Pear, Rik Peels, Robin Peeters, Charlotte C. S. Rulkens, Christin Scholz, Nienke Smit, Rombert Stapel & Joost de Winter

To cite this article: Maarten Derksen, Stephanie Meirmans, Jonna Brenninkmeijer, Jeannette Pols, Annemarijn de Boer, Hans van Eyghen, Surya Gayet, Rolf Groenwold, Dennis Hernaus, Pim Huijnen, Nienke Jonker, Renske de Kleijn, Charlotte F. Kroll, Angelos-Miltiadis Krypotos, Nynke van der Laan, Kim Luijken, Ewout Meijer, Rachel S. A. Pear, Rik Peels, Robin Peeters, Charlotte C. S. Rulkens, Christin Scholz, Nienke Smit, Rombert Stapel & Joost de Winter (2025) Replication studies in the Netherlands: Lessons learned and recommendations for funders, publishers and editors, and universities, Accountability in Research, 32:7, 1285-1303, DOI: 10.1080/08989621.2024.2383349

To link to this article: https://doi.org/10.1080/08989621.2024.2383349

Published online: 13 Aug 2024.

ACCOUNTABILITY IN RESEARCH
2025, VOL. 32, NO. 7, 1285–1303
HTTPS://DOI.ORG/10.1080/08989621.2024.2383349

Replication studies in the Netherlands: Lessons learned and recommendations for funders, publishers and editors, and universities

Maarten Derksen [a], Stephanie Meirmans [b], Jonna Brenninkmeijer[c], Jeannette Pols[d], Annemarijn de Boer [e], Hans van Eyghen[f], Surya Gayet [g], Rolf Groenwold [h], Dennis Hernaus[i], Pim Huijnen [j], Nienke Jonker [k], Renske de Kleijn [l], Charlotte F. Kroll [m], Angelos-Miltiadis Krypotos[n], Nynke van der Laan [o], Kim Luijken [p], Ewout Meijer [q], Rachel S. A. Pear [r], Rik Peels [s], Robin Peeters[t], Charlotte C. S. Rulkens [u], Christin Scholz [v], Nienke Smit [w], Rombert Stapel[x], and Joost de Winter [y]

[a]Theory & History of Psychology, University of Groningen, Groningen, The Netherlands; [b]Department of Ethics, Law and Humanities, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands; [c]Department of Ethics, Law & Humanities, Amsterdam Universitair Medische Centra University of Amsterdam, Amsterdam, The Netherlands; [d]Department of Ethics, Law & Humanities, Amsterdam UMC, Department of Anthropology, University of Amsterdam, Amsterdam, The Netherlands; [e]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands; [f]Tilburg School of Catholic Theology, Religion and Practice, Tilburg University, Tilburg, The Netherlands; [g]Experimental Psychology, Helmholtz Institute, Utrecht University, Utrecht, The Netherlands; [h]department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; [i]Department of Psychiatry & Neuropsychology, School for Mental Health and NeuroScience MHeNS, Maastricht University, Maastricht, The Netherlands; [j]Department of History and Art History, Utrecht University, Utrecht, The Netherlands; [k]Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, Groningen, The Netherlands; [l]Educational Center, University Medical Centre Utrecht, Utrecht, The Netherlands; [m]Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Department of Microeconomics and Public Economics (MPE), Department of Psychiatry & Neuropsychology, School for Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands; [n]Clinical Psychology, Utrecht University, Utrecht, The Netherlands; [o]Department Communication and Cognition, Tilburg University, Tilburg, The Netherlands; [p]Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands; [q]Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands; [r]Jerusalem Museum of Natural History, Jerusalem, Israel; [s]Faculty of Religion and Theology and Faculty of Humanities, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [t]Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands; [u]Department of Philosophy, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [v]Amsterdam School of Communication Research, Department of Communication Science, University of Amsterdam, Amsterdam, The Netherlands; [w]Department of Education, Utrecht University, Utrecht, The Netherlands; [x]International Institute of Social History, Amsterdam, The Netherlands; [y]Afdeling Cognitive Robotics, Faculty: Mechanical Engineering, Delft University of Technology, Delft, The Netherlands

CONTACT Maarten Derksen m.derksen@rug.nl Theory & History of Psychology, Faculty of Behavioural and Social Sciences, University of Groningen, Grote Kruisstraat 2/1, Groningen 9712 TS, The Netherlands

ABSTRACT
Drawing on our experiences conducting replications we describe the lessons we learned about replication studies and formulate recommendations for researchers, policy makers, and funders about the role of replication in science and how it should be supported and funded. We first identify a variety of benefits of doing replication studies. Next, we argue that it is often necessary to improve aspects of the original study, even if that means deviating from the original protocol. Thirdly, we argue that replication studies highlight the importance of and need for more transparency of the research process, but also make clear how difficult that is. Fourthly, we underline that it is worth trying out replication in the humanities. We finish by formulating recommendations regarding reproduction and replication research, aimed specifically at funders, editors and publishers, and universities and other research institutes.

ARTICLE HISTORY
Received 20 April 2024
Accepted 18 July 2024.

KEYWORDS
Replication; reproduction; funding; transparency

Background

In the last decade, researchers in several scientific fields have raised concerns about the replicability of research findings. In psychology, for example, a large-scale effort to replicate 100 experiments drawn at random from three psychological journals could only replicate around 40% of the original results, depending on how replication success was defined (Open Science Collaboration 2015; see also Camerer et al. 2018). A systematic review of the medical literature showed that the replicability of pre-clinical research findings was even lower (Begley and Ioannidis 2015). Researchers declared a replication crisis in psychology and in medicine, and questions were increasingly raised in other fields as well (e.g., Mueller-Langer et al. 2019). In a survey, 52% of researchers thought there was a significant crisis (Baker 2016b). In response a number of initiatives was set up to encourage and facilitate replication studies. For example, since 2018 *Royal Society Open Science* guarantees publication for replications of studies that were originally published in this journal (Chambers 2018).

Replications are considered important for two reasons. Firstly, it is a well-established idea among scientists that at a fundamental level science is based on repeatable observations (Zwaan et al. 2018). "Reproducibility is a defining feature of science" (Open Science Collaboration 2015). A failure to replicate the results of an earlier study suggests that they are possibly false. The low replicability rate in science is then taken to be an indication of a high prevalence of scientific errors, including questionable research practices and fraud (Edlund et al. 2021; Van Ravenzwaaij et al. 2023).

Secondly, replication studies are seen as necessary to investigate the generalizability of the original results to other populations and settings (Klein et al. 2018) and to explore the possible role of moderators, mediators, and

boundary conditions (Edlund et al. 2021). An experimental manipulation that works in one sample, in one location, may not work elsewhere with a different sample. Moderating and mediating variables may be involved in the experimental effect, and the manipulation may only work under certain conditions.

While the importance of replication studies is widely recognized, concerns have also been raised. Individual replication studies, including some of the studies of the landmark Open Science Collaboration reproducibility project, have been criticized for being poorly designed or conducted, or overstating their conclusions (e.g., Dijksterhuis 2018; Gilbert et al. 2016). More generally the emphasis on "direct replications," i.e., replications that follow the protocol of the original study, has been criticized. Replicability is not a guarantee that the original finding is correct, nor does a failure to replicate it necessarily mean that it is false (Devezer et al. 2021; Munafò and Davey Smith 2018; Rotello, Heit, and Dubé 2015; Van Ravenzwaaij et al. 2023). A replication study should be considered in the context of a series of studies in which variations on the original protocol may play a role next to direct replications, jointly providing information about a phenomenon or finding (Edlund et al. 2021; Hüffmeier, Mazei, and Schultze 2016).

Another concern was raised by some social psychologists, who reject the interpretation of failed direct replications as suggestive of scientific error. They consider direct replication to be an inappropriate method in their field of study, due to the contextual variability of social behavior. Since the same experimental manipulation may elicit different behavioral responses depending on the historical, cultural and social context, a failed direct replication of an earlier study is said to be "uninformative" without a theory that explains the deviating results (Strack and Stroebe 2018, 40). In their view, social psychology should instead focus on "conceptual replications," that test the same theory with a different experiment (Crandall and Sherman 2016; Iso-Ahola 2017; Stroebe and Strack 2014).

Finally, several authors have argued that replication has different meanings in different disciplines, and has no role in some fields of research, including parts of the humanities (Penders, Holbrook, and de Rijcke 2019). For example, in research where there is very little control of the study environment, one does not expect a replication to yield the exact same results, but skilled observers should detect similar patterns (Leonelli 2018). Qualitative interpretative research typically aims at trustworthiness but not replicability (Penders, Holbrook, and de Rijcke 2019).

In the Netherlands the replication crisis was an important reason for the Dutch funding agency NWO to initiate a funding scheme for replication studies in 2017 (https://www.nwo.nl/en/researchprogrammes/replication-studies). This dedicated replication funding was the very first of its kind and received international attention (Baker 2016a). NWO funded

a total of 24 projects from medicine, the social sciences, and the humanities to conduct replication studies of highly influential studies in their field. In 2021, NWO provided additional funding to conduct an ethnographic study of "Replication in Action" primarily based on these 24 projects. In March 2023, this ethnographic project team hosted a workshop that brought together many of the PI's, PhD students and postdocs of the NWO replication projects, as well as a few researchers involved in other replication studies, to share their experiences and insights, and discuss the role of replication in academic scholarship. With this paper we aim to contribute to the discussion about replication in science by describing the lessons we collectively learned about replication studies, and by formulating recommendations for researchers, policy makers, and funders about the role of replication in science and how it should be supported and funded.

This paper is internationally the first collaborative effort highlighting cross-disciplinary experiences with studies supported by funds dedicated to replication research. We are a group of researchers from various disciplines – medicine, social sciences, and humanities – and the insights and recommendations that we describe here reflect the commonalities of our experiences as well as some of the differences. Other papers looking at replication across disciplines focus on replication rates (Cobey et al. 2023), or they are of a more conceptual-philosophical nature (e.g., Leonelli 2018; Penders, Holbrook, and de Rijcke 2019). We believe it is important to supplement the philosophical and methodological discussions about replication in science with a conversation about the actual practice of doing replication studies. Much has been said about the role of replication in science, about different kinds of replication, and about the rate of replicability in different fields, but comparatively little has been written about researchers' experiences conducting replication studies. The paper by Errington et al. (2021) also describes practical experiences with replication studies, but they focus on a single discipline. Our contribution also stands apart as a reflection on a unique funding program dedicated to supporting only replication studies. We discuss the merits and challenges of separate funding for replication studies in our recommendations.

In this paper we do not report on the outcomes of the individual replication studies or offer a meta-analysis of their results, nor does this paper amount to an ethnography by the Replication in Action team. It is the product of our collective reflection on the preliminary outcomes of the Replication in Action project, and on our experiences doing replication studies. The results of the replication studies and the Replication in Action project are or will be published in separate articles.

NWO replication studies funding program

In 2017, the Dutch science funding organization NWO designated three million euros for replication projects in the social sciences, medical sciences and, in the last round, also in the humanities. The aim of the three replication calls (2017–2019) was to encourage researchers to carry out replication research and increase insight into the replicability of the results of the original studies (Replication Studies | NWO). In addition, the program aimed to gain insight into whether and how replication studies could effectively be included more broadly in NWO research programs. Finally, NWO hoped its program would contribute to an understanding of the importance of making research more transparent, and how to achieve this in practice.

NWO distinguished between three different types of replication studies:

(1) Reproduction: a study that repeats the analysis of the datasets of the original study.
(2) Replication with new data: a study with the same research question and the same protocol as the original study, but with newly-collected data (what is usually called a direct replication).
(3) Replication with the same research question: a study with the same research question as the original study, but with a different protocol and newly collected data.

The NWO pilot program only funded replications of the first two types. In both cases the maximum duration for projects was two years. All projects had to replicate "cornerstone research: research that has had substantial consequences with respect to theory or policy and for which it is therefore important to assess whether the results on which these consequences are based are reproducible" (Netherlands Organization for Scientific Research 2016, p. 4). In total, NWO funded two reproduction studies (both in medical science) and 22 direct replication studies: 18 social science studies, three medical studies and one study in the humanities. Of these, the "Replication in Action" project follows 21 studies, and an additional 7 funded by other programs.

The process that led to this paper

The Replication in Action team has conducted 68 interviews with researchers who are involved with the replication studies it follows. Audio recordings and notes were made of all first interviews and many follow up interviews. In some cases, only notes were taken after informal follow up interviews. Team members have conducted extensive observations of experiments, desk work and research meetings of replication researchers. Furthermore, team

members have been participant observers at seminars, symposia, conferences and workshops devoted to replication and scientific reforms, and have conducted interviews with staff members of NWO. At the time of the workshop, transcription and coding of interviews was ongoing. The transcripts and notes of interviews and observations, as well as other relevant material were shared among the team members. All fieldwork experiences were shared and discussed in weekly team meetings. Our analyses and draft papers were discussed in regular in-depth analysis sessions.

To share and discuss its preliminary observations with the researchers it follows, the Replication in Action team hosted a workshop on 16 March 2023 in Amsterdam, the Netherlands bringing together researchers of the NWO replication projects, as well as several researchers involved in other replication studies. Based on the material it had gathered until that date, the team had identified a number of key findings with regard to replication studies, as well as some key questions to explore in the workshop. The team first presented the findings along the lines of four stages of doing a replication project: (1) motives and aims to do a replication study; (2) considerations regarding the design and materials of the replication study; (3) problems and surprises when doing a replication study; and (4) the impact of a replication study on the original research, the researcher and the field more broadly.

This presentation was followed by a plenary discussion regarding a few of the central questions the team had identified prior to the workshop, and which were also inspired by the aims of their own project, including giving policy advice. These questions were: what unforeseen problems have replication researchers encountered during the execution of the replication projects, and what went better than expected? What did researchers learn, what would they do different next time? In what way does replication contribute to good science? And finally, what conclusions can we draw for research policy and future replication practices? The plenary discussion was recorded and the Replication in Action team took notes while the participating replication researchers shared their experiences. Using the notes and the recordings, the Replication in Action team subsequently summarized the main experiences and recommendations discussed in the plenary discussion in a short document. It appeared that there was quite some agreement in our experiences with replication and our ideas on how to continue, and this encouraged us to coauthor a paper. All replication experts who are followed in the "Replication in Action" project (including those who had not been able to attend the workshop) were invited to co-write (extend and edit) the manuscript, resulting in the present paper.

From the outset, the workshop was part of the project plan of the "Replication in Action" project, with a collective white paper as an ideal outcome. The data management plan of the "Replication in Action" project has been drawn up by members of the Department of Ethics, Law and

Humanities at the Amsterdam UMC (JP and SM) and has been checked and approved by NWO. Although this project is exempt from ethical approval at the institution where it is based (Amsterdam UMC), it follows the ethical guidelines of the Dutch Anthropological Association ABv (described and discussed in De Koning et al. 2019). The emphasis in these guidelines on ethnographic knowledge as co-produced in interaction between observers and participants is reflected in the process that led to this paper. Rather than the ethnographers independently drawing conclusions from their observations and interviews, reflecting on replication in practice was turned into a participatory project, in which consensus was an important goal. Thus, the lessons and recommendations provided below are the result of our collective effort.

Lessons

The workshop and co-writing of this article brought to light four key takeaways:

1. Conducting a replication study can have a variety of benefits

We collectively identified a variety of benefits that conducting replication studies can have. Some of these were the reason we took on a replication project, others were by-products, unexpected advantages.

a) Corroborating the original findings

One main category of reasons which we identified revolves around corroboration. Replications can be used to corroborate the original finding and provide further evidence of its trustworthiness. Do its claims generalize to other populations, and do they hold when the experiment is conducted in a different lab environment with different experimenters? Sometimes there is conflicting evidence from earlier replications, and one might thus seek to gather further evidence about the replicability of the effect. The wish to corroborate the original findings can also stem from concerns regarding methodological aspects of the original study. For example, a direct replication may be deemed necessary because the original sample size was too small to rigorously support the conclusion. In humanities research, this can translate to extending the number of sources. Sometimes, a divergence from the original protocol may be considered necessary to evaluate the same research question using an approach that is deemed more suitable or state of the art by replicators. One may also want to redo the original research as a pre-registered replication. See also section 2, below, on other ways to improve the original study.

b) Learning more about a study or a field of research

For many workshop participants, this had been an unexpected benefit of doing a replication. Whereas replication is usually discussed in terms of the corroboration of outcomes, we noticed that conducting a replication also provides many insights into the methodology of the original research. It can thus help to learn or test a new research technique, or explore new sources (in a replication in the humanities). But it can also help to identify and address weaknesses in a research field more broadly. More specifically, we noticed that when studies cannot be reproduced or their results are not replicated, the replication study can provide insight into the reasons for disparities between studies. The original report may turn out to be an insufficient description of the procedure, the influence of certain decisions on the results can become clear, flaws of the original study are exposed, and/or a different method or improvements for the study design suggest themselves. That knowledge can be used to interpret the results of the original study under a new light, or to design a new study. Although this was often not the original motivation to perform a replication study, it emerged as a very important benefit and adds to the value of performing replication studies. Replication work can then in turn also provide important ideas regarding future research efforts, and provide more clarity about available methodological choices.

c) Conducting replications as an educational tool

For many of us, this too was an unexpected advantage of replication studies. In line with the benefits outlined in b), we also realized that reproductions and direct replications can effectively be used as an educational tool. Even highly experienced senior researchers can learn a lot from replication work. It provides deep insights into methodologies, making their intricacies and the assumptions behind them more visible, which we see as crucial in any academic education. Such education targets could be tailored to various academic stages: in an MSc research project, for example, an individual student could set out to replicate part of an original study, while a PhD student could replicate the study in total. Some of us have assigned replication studies to BSc students, and our impression is that the students found it somewhat unusual to engage deeply with a highly bounded research question and framework right away, but that, perhaps precisely because of this, a lot was learned regarding the details of research techniques and methods. Reproductions and/or relatively easy and straightforward direct replications could also become part of course work, where students work in groups to conduct a reproduction, design a replication attempt, or make a start with conducting an actual replication. Some of us already have started using reproductions or direct replications in PhD training, and students experience it as insightful.

2. It is often crucial to improve aspects of the original study

Many of us felt constrained by the emphasis of NWO on closely following the same research protocol as the original study (see details on NWO funding call above). In some cases, the original researchers who shared their protocols likewise demanded that the original protocol be followed to the letter. However, even when the original study was well designed, there were often legitimate reasons to improve on it (see also 1a). For example, for replications carried out years or even decades after the original study, improved instruments were available or the old instruments were no longer available. Many of us experienced that faithfully following the original protocol was not possible or even not sensible.

We do acknowledge that not following the original protocol to the letter may cause others, including the original researchers, to reject the replication as insufficiently similar to the original to be informative regarding the original hypothesis and results. However, we think it is important to realize that two studies are *never* entirely identical, and that discussion is *always* possible about the relevance of the differences between them. Much has been written about this in the philosophy and sociology of science (see e.g., Earp and Trafimow 2015 for an overview) and, indeed, many of the controversies about replication studies of the last ten years revolve around this issue. Thus, striving for a perfect replica of the original study is bound to fail. Instead the replication study should be seen as the first step in examining the relevance of the procedural choices that were made in the original study (see also lesson 1b) and the possible role of moderator variables. In regard to this, an important and novel idea that several of us explored was to perform replication studies that allowed for both direct *and* extended replications of the original findings. One of our projects first used a replica of the original equipment (a slide projector) for the presentation of stimuli, but when that was found to create artifacts it was decided to use a computer monitor instead (De Winter et al. 2021). Responses were measured using modern, more precise methods. Thus, this replication was not direct, but close to the original. In this study, several of the original findings replicated, some did not. In another of our projects the authors first directly replicated the original analysis pipeline (nowadays considered flawed) and supplemented this with a conceptual replication by applying an updated, alternative analysis approach following up on unexpected findings, and then testing the sensitivity of results to key analytical choices. This approach allowed a direct comparison between results obtained from direct versus conceptual replication, highlighting the impact (or lack thereof) of experimenter choices that are deemed to be important. The original finding could only be replicated with the original analysis pipeline, not with the updated one (Scholz et al. 2022). A dual approach was also applied for the same reasons in the

"Replicating a Rembrandt Study," which revolved around the question whether two paintings should be attributed to Rembrandt or not. This was investigated by first reproducing an earlier attribution study (determining the attribution based on the methods and data of that earlier study), followed by a conceptual replication that used additional, more modern technical research methods and another team of experts (Rulkens et al. 2022, 2023). The study allowed for triangulation by approaching the same question with different methods. (The results of this study have not been published yet.) We do acknowledge that such dual replication studies will typically be more costly and time-intensive.

To summarize this lesson we draw from our experiences: the boundaries of direct replication should not be a hindrance to doing good research. Researchers should be allowed, and should allow themselves, to deviate from the protocol of the original study and thus make their replication more "conceptual" when that seems necessary. If time and resources allow, a dual, direct *and* conceptual replication can also be considered.

3. Replication studies highlight the importance of and need for more transparency of the research process, but also make clear how difficult that is

Direct replications, by definition, require us to follow all steps of the original protocol in the exact same manner. Yet, even experienced, highly conscientious researchers often find it difficult to document their protocols in enough detail to support direct replication. Performing a replication study made many of us reflect on the process of doing research in general. Replicating the work of other researchers confronted us with the many degrees of freedom inherent in designing a study and analyzing data (see also Silberzahn et al. 2018). As researchers we know that many decisions (e.g., regarding design, analysis, etc.) need to be made to conduct a study, because we have conducted or are currently conducting (or doing an ethnographic study of) original studies that required us to make such decisions. But if there is already a study in place that prescribes the steps that one should take to replicate it, there is a new and often unexpected occasion to reflect on each step and whether or not they make sense in light of the overall study aim. Many of us noticed that the design and analysis decisions are often insufficiently described in the original report or are not mentioned at all. This, we noticed, is especially the case with older studies, published at a time when journals had strict word limits and no option to add supplementary material.

Conducting a replication study can thus generally help to highlight the need for more details and specificity in the reporting of studies. It can also more specifically influence how one reports and shares one's own studies. Replicating a study made many of us think about the information that would

be necessary to replicate one's own studies. In connection with the educational efforts described in 1d, replication studies could thus present an effective boost for reproducible science practices.

A lack of transparency can have various reasons. In the very worst, but probably infrequent cases, researchers might intentionally "hide" questionable choices to increase their chance of publication. We experienced that more often, a lack of transparency is due to (incorrect) assumptions about what is common knowledge or practice, and therefore certain specifics are not deemed necessary to be reported. In other instances, the number of decisions that must be taken when setting up a study is so great that it is almost inevitable that some go unreported. Although online supplements may allow more space for reporting the details of a study, for the majority of readers of our future papers it likely is not useful to write 40-page descriptions of every minuscule decision we made. Still, the necessary information should somehow be available to the readers that do need to know (e.g., by publishing all materials and methodological details on an online open repository). Anything that provides some guidance and structure to this process is helpful. Reporting guidelines already exist in several domains of study (see for example Nichols et al. 2016; Poldrack et al. 2008; Smeets et al. 2019). Another example is the practice of making neuroimaging datasets broadly available using a standardized folder structure and naming convention, called BIDS: https://bids.neuroimaging.io/ We did notice that sometimes methods can appear sufficiently standardized and thus compatible, but may not be so in reality. For example, in medical studies, countries/institutes often apply the same coding systems (e.g., ICD-codes for disease diagnosis), but in one of our studies (unpublished) it became clear that registration practices may differ between countries/institutes to such degree that they cannot easily be transferred. Assuming standardization in this case resulted in erroneous interpretation of results, so caution is advised even when dealing with supposedly standardized data.

4. Replication in the humanities is an idea worth exploring

As noted in the introduction, it has been argued that replication studies are not appropriate for many fields in the humanities. Although research in the humanities sometimes deals with matters of fact, for which replication would increase credibility (Penders, Rijcke, and Holbrook 2020), replication would not make sense for studies that rely on interpretation and/or deal with unique events (Penders, Holbrook, and de Rijcke 2019). The scholars from the humanities who were present at the workshop certainly found doing a replication study a rewarding experience, thus confirming what some of those involved had previously suggested for those humanities studies that employ empirical methods (Peels and Bouter 2018). For example, one of our

studies is a replication of a historical essay on the relation between puritanism and support for science (Van Eyghen, Van den Brink, and Peels in press). Although the original text did not include a clear explanation of the way the conclusions were arrived at, a study protocol was reconstructed and discussed with the original author. The original sources were then re-analyzed, as well as several new ones that had not been available to the original author. Thus, this replication arguably also crossed the strict limits of a direct replication. The replicating authors came to somewhat different conclusions than the original author, with several of their interpretations of sources departing from his. Inclusion of the new sources made the greatest difference to the overall conclusion.

Even when it is not immediately clear what "conducting a replication" means with respect to a particular type of study, exploring this question can lead to new insights into the research practices in one's field. As in the example above, replicating earlier work shows that it is not always clear how interpretations have been arrived at. This implies that greater transparency about the analytical process is required. Revisiting sources that were analyzed by other scholars can be seen as a form of replication that is already central to the humanities – even if they are not acknowledged as such. The goal of such "replications" is usually a better or different interpretation of the material. In the field of history, this is known as the "historiographical debate" and at the core of its self-image as "a discussion without end." It is likely, however, that the increasing digital availability of widely-used archives will also lead to an increase in replications with the goal of quality control and transparency (Huijnen and Huistra 2022). At the same time we would like to emphasize that this does not mean that replication is always a sensible practice in all parts of the humanities.

Recommendations

The lessons we learned from conducting reproduction and replication studies are not only relevant for other researchers, but also for funders, publishers, editors, and universities and other research institutes. We have come to appreciate replication and reproduction studies as a valuable and rewarding kind of research. In general, we recommend that reproduction and replication are acknowledged as a standard part of the production of knowledge in most disciplines, not a separate category of research. We have also seen, however, that replication studies present unique challenges that may require adjustments to the research support system. Moreover, there are many differences between fields, and we also recommend being careful with making replication studies a requirement. It is crucial to keep this process flexible in order to accommodate the need of individual scientific endeavors rather than standardize procedures across entire fields. There is variation in when

a replication is useful and what type of replication should be done – and this varies not only across fields but also between specific studies. Based on our collective experience with conducting replication studies, we believe that it is especially useful to encourage replication in those fields where the reproduction of knowledge is theoretically valued but not self-evident in practice. With regard to those fields, we think the following recommendations are worth considering.

1. Funders: Appreciate replication and reproduction studies

We hope that other funding organizations will follow NWO's example and increase their support of replication and reproduction studies. The importance of replication and reproduction in corroborating earlier findings has of course been highlighted many times, but our experiences, detailed above, show that they also have a key role in exploring and developing the methodology in a field of research (lesson 1b). Dual (direct and conceptual) replication studies (lesson 2) can offer further insight into a methodology and how it can be improved. That leaves the question how funding organizations should support replication studies. Many of us believe that replication efforts should be integrated into normal funding streams. Typically funding streams emphasize novelty and innovation in combination with extending earlier work. While at first glance replication research does not meet these criteria, in practice replications are not mere duplicates of an earlier study: they often improve on the original study, for example with a larger sample size, better instrumentation, and a more transparent research process (see lessons 1b, 2). We emphasize the importance of giving replicators the opportunity and freedom to improve the original research protocol where needed, and detail the arguments for and implications of doing so in their papers. In this sense, replications might in practice already be more "innovative" than generally appreciated (Apart from this, we believe that both novelty *and* corroboration should be important in, for example, excellence funding schemes. See also Brembs 2019.). However, there are also situations where funding in a separate funding program is more advisable. One argument for dedicated replication funding is that certain areas of research may have a backlog of studies that have never been replicated, although they have attracted citations.

Funders simultaneously need to be aware that replication studies tend to cost more and take more time than the original study. They often need a larger sample size, because standards regarding statistical power have been raised. Moreover, having to follow someone else's protocols can slow the process down, for example because the original authors cannot remember all the details of their protocol or do not have the materials any more. Or, as happened in one of our projects, because one has to run additional studies to

show that the stimuli perform just like the original stimulus set. If a replication study adds additional experimental or analysis branches, which can yield novel insights (see lesson 2), this will also bring extra costs and it is more time-consuming. Replication studies may therefore require more, rather than less, funding compared to the original studies. In our experience, a replication study requires a budget and time frame that is different from, and typically larger than, the original research.

Finally, replication research should be appreciated in the assessment of researchers (see also recommendation to universities below), and funders should keep in mind that replication research can be time-consuming and may result in fewer publications and citations than "novel" research.

2. Publishers and editors: Appreciate the value of replication studies

Although a replication study adds to the evidence that is relevant to the original research question, this new evidence is not visible if one only reads the original research report. As a result replication studies often have little impact on whether and why the original study is cited. Studies have shown that the publication of a replication study, even if its results go counter to those of the original study, often has little impact on whether and why the original study is cited (Hardwicke et al. 2021; Schafmeister 2021). Some of us had similar experiences: Scholz et al. (2022), a replication that was published over a year ago, has not been cited yet, while the original study has amassed over a hundred citations since January 2023. We urge publishers and editors to consider several strategies that could effectively solve this problem. First, journals can provide links within the original studies to any published replication studies. Scientific aggregators such as Pure, Google Scholar, ORCID, etc. could be used in such efforts. Second, journal editors can invite the original researchers – if they are still active – to write a reply to the replication study, which can then be published together with the replication. One of our humanities replication projects successfully convinced a journal editor to use this format, and we believe this is an interesting new way to present replications in publications. Third, journals can devote special issues to replication of cornerstone studies in their fields. Fourth, journals could allocate space for replication studies in special article types, or mention explicitly within their scope or author guidelines that replication studies are welcome in their journal (as is already the case in many journals). This could include explicit guidelines for the submission of replication articles. Finally, we would like to emphasize that high impact journals should also accept (more) replication studies – especially when they have also published the original study. We need to move away from an exclusive emphasis on novel results and consider reproduction

and replication studies as an integral part of making inferences regarding a particular topic.

3. Universities: Appreciate replication efforts in researcher assessments and university policies

We urge universities to appreciate and stimulate replication and reproduction efforts in a way that they deserve. One effective way to increase the number of replication studies performed would be to encourage PhD students to include at least one replication study in their PhD thesis, in those fields and studies where this is relevant. Another way to better incorporate replication in research practice is by encouraging and facilitating replication in teaching practices (see lesson 1b). For some fields, gathering (theoretical and practical) experience with replication studies could be a valuable part of the bachelor or master curriculum.

In general, we see a welcome attitude change regarding replication research, which is increasingly seen as an important, integral part of the scientific process, including by crucial academic stakeholders such as NWO. Although our experiences vary somewhat, most of us have received appreciation for our replication research from colleagues. There is an important role for employers to join this development by emphasizing the importance and value of replication, crediting it in their hiring policies and rewarding it in their evaluation of researchers.

Conclusion

Conducting replications can be exciting, and is much more than merely repeating the work of other researchers. We found that there are many benefits to doing replications, and many of us noted the deep insights replication work can provide with regards to methodologies, procedures and transparency. Several of us also started to explore replications as a teaching tool. Along with our insights and experiences, we have formulated a list of recommendations for policymakers. Replications should ideally become part of normal research, rather than being set apart as a special category. But because replications are currently not treated as such, one might need special policies to foster their uptake, such as raising awareness in hiring and funding committees, and making the publication of replication studies easier and more effective.

Acknowledgments

We would like to thank the researchers who participated in the workshop, but chose not to take part in writing this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by: NWO grant [401.18.035] (Surya Gayet); NWO grant [401.16.001/3873] (Meijer); Templeton World Charity Foundation grant [TWCF0163] (Rik Peels, Charlotte Rulkens, Hans van Eyghen, and Rachel S.A. Pear; the opinions expressed in this publication are those of the authors and do not necessarily reflect the views of TWCF); NWO grant [401.18.056] (Angelos-Miltiadis Krypotos); NWO grant [401.18.048] (Nienke Jonker); NWO grant [401.19.038] (Rombert Stapel); Maastricht University Centre for Integrative Neuroscience (CIN) Interfaculty grant, and NWO grant [401.19.006] (Charlotte F. Kroll and Dennis Hernaus); Fostering Open Science Fund Utrecht University 2020 (Pim Huijnen); NWO grant [10259] (Annemarijn de Boer); NWO grant [401.16.023] (Nynke van der Laan); NWO grant [406.20.FR.007] (Derksen, Meirmans, Brenninkmeier and Pols); NWO grant [401.19.015] (Nienke Smit); NWO grant [VI.Veni.191.G034] (Christin Scholz); NWO grant [401.16.083] (Joost de Winter).

ORCID

Maarten Derksen http://orcid.org/0000-0003-1572-4709
Stephanie Meirmans http://orcid.org/0000-0001-6509-1212
Annemarijn de Boer http://orcid.org/0000-0003-4200-0917
Surya Gayet http://orcid.org/0000-0001-9728-1272
Rolf Groenwold http://orcid.org/0000-0001-9238-6999
Pim Huijnen http://orcid.org/0000-0002-5477-6352
Nienke Jonker http://orcid.org/0000-0002-2567-3776
Renske de Kleijn http://orcid.org/0000-0001-9206-4199
Charlotte F. Kroll http://orcid.org/0000-0003-2539-5814
Nynke van der Laan http://orcid.org/0000-0003-4307-0888
Kim Luijken http://orcid.org/0000-0001-5192-8368
Ewout Meijer http://orcid.org/0000-0001-9590-3699
Rachel S. A. Pear http://orcid.org/0009-0002-1793-7767
Rik Peels http://orcid.org/0000-0001-8107-5992
Charlotte C. S. Rulkens http://orcid.org/0000-0002-4617-9507
Christin Scholz http://orcid.org/0000-0001-6567-7504
Nienke Smit http://orcid.org/0000-0002-2672-9794
Joost de Winter http://orcid.org/0000-0002-1281-8200

References

Baker, M. 2016a. "Dutch Agency Launches First Grants Programme Dedicated to Replication." *Nature*. https://doi.org/10.1038/nature.2016.20287.

Baker, M. 2016b. "Is There a Reproducibility Crisis?" *Nature* 533 (7604): 452–454. https://doi.org/10.1038/533452a.

Begley, C. G., and J. P. A. Ioannidis. 2015. "Reproducibility in Science." *Circulation Research* 116 (1): 116–126. https://doi.org/10.1161/CIRCRESAHA.114.303819.

Brembs, B. 2019. "Reliable Novelty: New Should Not Trump True." *PLOS Biology* 17 (2): e3000117. https://doi.org/10.1371/journal.pbio.3000117.

Camerer, C. F., A. Dreber, F. Holzmeister, T.-H. Ho, J. Huber, M. Johannesson, M. Kirchler, G. Nave, B. A. Nosek, T. Pfeiffer, et al. 2018. "Evaluating the Replicability of Social Science Experiments in Nature and Science Between 2010 and 2015." *Nature Human Behaviour* 2 (9): 637–644. https://doi.org/10.1038/s41562-018-0399-z.

Chambers, C. 2018. "Reproducibility Meets Accountability: Introducing the Replications Initiative at Royal Society Open Science | Royal Society." https://royalsociety.org/blog/2018/10/reproducibility-meets-accountability/ .

Cobey, K. D., C. A. Fehlmann, M. C. Franco, A. P. Ayala, L. Sikora, D. B. Rice, C. Xu, J. P. A. Ioannidis, M. M. Lalu, A. Ménard, et al. 2023. "Epidemiological Characteristics and Prevalence Rates of Research Reproducibility Across Disciplines: A Scoping Review of Articles Published in 2018-2019." *eLife* 12: e78518. https://doi.org/10.7554/eLife.78518.

Crandall, C. S., and J. W. Sherman. 2016. "On the Scientific Superiority of Conceptual Replications for Scientific Progress." *Journal of Experimental Social Psychology* 66: 93–99. https://doi.org/10.1016/j.jesp.2015.10.002.

De Koning, M., B. Meyer, A. Moors, and P. Pels. 2019. "Guidelines for Anthropological Research: Data Management, Ethics, and Integrity." *Ethnography* 20 (2): 170–174. https://doi.org/10.1177/1466138119843312.

Devezer, B., D. J. Navarro, J. Vandekerckhove, and E. 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.

De Winter, J. C. F., S. M. Petermeijer, L. Kooijman, and D. Dodou. 2021. "Replicating Five Pupillometry Studies of Eckhard Hess." *International Journal of Psychophysiology* 165: 145–205. https://doi.org/10.1016/j.ijpsycho.2021.03.003.

Dijksterhuis, A. 2018. "Reflection on the Professor-Priming Replication Report." *Perspectives on Psychological Science* 13 (2): 295–296. https://doi.org/10.1177/1745691618755705.

Earp, B., and D. Trafimow. 2015. "Replication, Falsification, and the Crisis of Confidence in Social Psychology." *Frontiers in Psychology* 6:6. https://doi.org/10.3389/fpsyg.2015.00621.

Edlund, J. E., K. Cuccolo, M. S. Irgens, J. R. Wagge, and M. S. Zlokovich. 2021. "Saving Science Through Replication Studies." *Perspectives on Psychological Science* 17 (1): 1–10. https://doi.org/10.1177/1745691620984385.

Errington, T. M., A. Denis, N. Perfito, E. Iorns, and B. A. Nosek. 2021. "Reproducibility in Cancer Biology: Challenges for Assessing Replicability in Preclinical Cancer Biology." *eLife* 10:e67995. https://doi.org/10.7554/eLife.67995.

Gilbert, D. T., G. King, S. Pettigrew, and T. D. Wilson. 2016. "Comment on 'Estimating the Reproducibility of Psychological Science'." *Science* 351 (6277): 1037–1037. https://doi.org/10.1126/science.aad7243.

Hardwicke, T. E., D. Szűcs, R. T. Thibault, S. Crüwell, O. R. van den Akker, M. B. Nuijten, and J. P. A. Ioannidis. 2021. "Citation Patterns Following a Strongly Contradictory Replication Result: Four Case Studies from Psychology." *Advances in Methods and Practices in Psychological Science* 4 (3): 25152459211040837. https://doi.org/10.1177/25152459211040837.

Hüffmeier, J., J. Mazei, and T. Schultze. 2016. "Reconceptualizing Replication as a Sequence of Different Studies: A Replication Typology." *Journal of Experimental Social Psychology* 66:81–92. https://doi.org/10.1016/j.jesp.2015.09.009.

Huijnen, P., and P. Huistra. 2022. "On the Use of Replications in History." A white paper, Utrecht. https://zenodo.org/record/7037401 .

Iso-Ahola, S. E. 2017. "Reproducibility in Psychological Science: When Do Psychological Phenomena Exist?" *Frontiers in Psychology* 8. https://doi.org/10.3389/fpsyg.2017.00879.

Klein, R. A., M. Vianello, F. Hasselman, B. G. Adams, R. B. Adams, S. Alper, M. Aveyard, et al. 2018. "Many Labs 2: Investigating Variation in Replicability Across Samples and Settings." *Advances in Methods and Practices in Psychological Science* 1 (4): 443–490. https://doi.org/10.1177/2515245918810225.

Leonelli, S. 2018. "Rethinking Reproducibility as a Criterion for Research Quality." In *A Symposium on Mary Morgan: Curiosity, Imagination, and Surprise*. 129–146. Vol. 36B. Emerald Publishing Limited. https://doi.org/10.1108/S0743-41542018000036B009.

Mueller-Langer, F., B. Fecher, D. Harhoff, and G. G. Wagner. 2019. "Replication Studies in Economics—How Many and which Papers are Chosen for Replication, and Why?" *Research Policy* 48 (1): 62–83. https://doi.org/10.1016/j.respol.2018.07.019.

Munafò, M. R., and G. Davey Smith. 2018. "Repeating Experiments is Not Enough." *Nature* 553 (7689): 399–401. https://doi.org/10.1038/d41586-018-01023-3.

Netherlands Organisation for Scientific Research. 2016. *Call for Proposals. Replication Studies*.

Nichols, T. E., S. Das, S. B. Eickhoff, A. C. Evans, T. Glatard, M. Hanke, N. Kriegeskorte, et al. 2016. *Best Practices in Data Analysis and Sharing in Neuroimaging Using MRI* bioRxiv. 054262. https://doi.org/10.1101/054262.

Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349 (6251): aac4716. https://doi.org/10.1126/science.aac4716.

Peels, R., and L. Bouter. 2018. "The Possibility and Desirability of Replication in the Humanities." *Palgrave Communications* 4 (95). https://doi.org/10.1057/s41599-018-0149-x.

Penders, B., J. B. Holbrook, and S. de Rijcke. 2019. "Rinse and Repeat: Understanding the Value of Replication Across Different Ways of Knowing." *Publications* 7 (3): 52. https://doi.org/10.3390/publications7030052.

Penders, B., S. D. Rijcke, and J. B. Holbrook. 2020. "Science's Moral Economy of Repair: Replication and the Circulation of Reference." *Accountability in Research* 27 (2): 107–113. https://doi.org/10.1080/08989621.2020.1720659.

Poldrack, R. A., P. C. Fletcher, R. N. Henson, K. J. Worsley, M. Brett, and T. E. Nichols. 2008. "Guidelines for Reporting an fMRI Study." *Neuroimage: Reports* 40 (2): 409–414. https://doi.org/10.1016/j.neuroimage.2007.11.048.

Rotello, C. M., E. Heit, and C. Dubé. 2015. "When More Data Steer Us Wrong: Replications with the Wrong Dependent Measure Perpetuate Erroneous Conclusions." *Psychonomic Bulletin & Review* 22 (4): 944–954. https://doi.org/10.3758/s13423-014-0759-2.

Rulkens, C. C. S., R. Peels, L. Bouter, M. Stols-Witlox, S. Meloni, E. Buijsen, and R. van Woudenberg. 2023. "Replicating a Rembrandt Study." (Preregistration) October. 10. https://doi.org/10.17605/OSF.IO/VDKAX.

Rulkens, C. C. S., H. Van Eyghen, R. Pear, R. Peels, L. Bouter, M. Stols-Witlox, G. van den Brink, E. Buijsen, S. Meloni, and R. van Woudenberg. 2022. *Exploring the Strengths and Limitations of Replication in the Humanities: Two Case Studies. Web Publication or Website*. Center for Open Sciences. https://www.cos.io/blog/exploring-the-strengths-and-limitations-of-replication-in-the-humanities .

Schafmeister, F. 2021. "The Effect of Replications on Citation Patterns: Evidence from a Large-Scale Reproducibility Project." *Psychological Science* 32 (10): 1537–1548. https://doi.org/10.1177/09567976211005767.

Scholz, C., H.-Y. Chan, R. A. Poldrack, D. T. D. de Ridder, A. Smidts, and L. N. van der Laan. 2022. "Can We Have a Second Helping? A Preregistered Direct Replication Study on the Neurobiological Mechanisms Underlying Self-Control." *Human Brain Mapping* 43 (16): 4995–5016. Advance online publication. https://doi.org/10.1002/hbm.26065.

Silberzahn, R., E. L. Uhlmann, D. P. Martin, P. Anselmi, F. Aust, E. Awtree, Bahnik, S., et al. 2018. "Many Analysts, One Data Set: Making Transparent How Variations in Analytic

Choices Affect Results." *Advances in Methods and Practices in Psychological Science* 1 (3): 337–356. https://doi.org/10.1177/2515245917747646.

Smeets, P. A. M., A. Dagher, T. A. Hare, S. Kullmann, L. N. van der Laan, R. A. Poldrack, H. Preissl, D. Small, E. Stice, and M. G. Veldhuizen. 2019. "Good Practice in Food-Related Neuroimaging." *The American Journal of Clinical Nutrition* 109 (3): 491–503. https://doi.org/10.1093/ajcn/nqy344.

Strack, F., and W. Stroebe. 2018. "What Have We Learned? What Can We Learn?" *Behavioral and Brain Sciences* 41:39–40. https://doi.org/10.1017/S0140525X18000870.

Stroebe, W., and F. Strack. 2014. "The Alleged Crisis and the Illusion of Exact Replication." *Perspectives on Psychological Science* 9 (1): 59–71. https://doi.org/10.1177/1745691613514450.

Van Eyghen, H., G. Van den Brink, and R. Peels. in press. "Brooke on the Merton Thesis: A Direct Replication of John Hedley Brooke's Chapter on Scientific and Religious Reform." *Zygon: Journal of Religion and Science*.

Van Ravenzwaaij, D., M. Bakker, R. Heesen, F. Romero, N. Van Dongen, S. Crüwell, S. M. Field, et al. 2023. "Perspectives on Scientific Error." *Royal Society Open Science* 10 (7): 230448. https://doi.org/10.1098/rsos.230448.

Zwaan, R. A., A. Etz, R. E. Lucas, and M. B. Donnellan. 2018. "Making Replication Mainstream." *Behavioral and Brain Sciences* 41: 1–13. https://doi.org/10.1017/S0140525X17001972.