Introducing a replication-first rule for Ph.D. projects

doi:10.1017/S0140525X18000730, e138

Arnold R. Kochari[a] and Markus Ostarek[b]

[a]*Institute for Logic, Language and Computation, University of Amsterdam, 1090 GE Amsterdam, The Netherlands;* [b]*Max Planck Institute for Psycholinguistics, Nijmegen, 6500 AH Nijmegen, The Netherlands*

a.kochari@uva.nl markus.ostarek@mpi.nl

http://akochari.com/

http://www.mpi.nl/people/ostarek-markus

Abstract: Zwaan et al. mention that young researchers should conduct replications as a small part of their portfolio. We extend this proposal and suggest that conducting and reporting replications should become an integral part of Ph.D. projects and be taken into account in their assessment. We discuss how this would help not only scientific advancement, but also Ph.D. candidates' careers.

Commenting on the role that replications should play in a researcher's career, Zwaan et al. briefly suggest that early career researchers should conduct replications "with the goal of building on a finding or as only one small part of their portfolio" (sect. 5.5.1, para. 4). Extending this, we propose that conducting and reporting replications should become an integral part of Ph.D. projects and should be taken into account in their assessment. Specifically, we suggest adopting a *replication-first rule*, whereby Ph.D. candidates are expected to first conduct a replication when they are building on a previous finding, and only then collect data in their novel study.

One reason we consider it important to specifically address the role of replications for early career researchers is that they face enormous pressure to establish themselves in the scientific community and often fear that their careers could end before they really begin (Maher & Anfres 2016; "Many junior scientists" 2017). Currently, to secure a job in academia after obtaining a doctoral degree, one needs to build an impressive portfolio of publications (Lawrence 2003). Based on our observations of how research projects are carried out in practice, Ph.D. candidates often directly attempt innovative extensions of previous experimental work in the hope of answering a novel research question, because novelty strongly increases publishability (Nosek et al. 2012). When such extensions fail to produce the expected results, they tend to collect more data in several variations of their own experiments before turning to examine the replicability of the original effect. However, it may often turn out that they cannot reproduce the original finding, possibly because the original effect is, in fact, not robust. In these cases, replicating the original effect first would prevent what may turn out to be a substantial waste of time and resources on follow-up experiments. Moreover, the time saved as a result of replicating first can be used to further examine the robustness of the original effect, for example, by conducting an additional high-powered replication. Such replications contribute to a better estimate of effect sizes, which are currently often overestimated on account of publication bias, sampling error, or p-hacking (Fanelli 2011; Ferguson & Brannick 2012; Szucs & Ioannidis 2017a). As such, replications constitute an important scientific contribution and should be regarded as such by Ph.D. project advisors.

The above arguments demonstrate the advantages of replicating first in the case of a failed replication. Likewise, successful replications provide a great opportunity. Pressure to publish operating simultaneously with publication bias means that early career researchers are currently pressed to obtain specifically *positive* findings to publish papers. As a result, in our experience, not knowing whether an experiment will yield positive results causes anxiety in Ph.D. candidates. Incorporating replications as a first step of any new research project can help alleviate this anxiety. If an extension shows no effect or supports the null hypothesis after a successful replication of the original effect, it should be easier to interpret the theoretical significance of this outcome. For example, suppose that one replicates a previously observed priming effect but does not obtain it when the primes are masked. In this case, one can directly compare the effect in both conditions and make a convincing case about the role of visibility for the effect. These two experiments can likely be put together in a strong paper. Similarly, a successful replication and extension make for a solid package that will convince Ph.D. candidates themselves and the fellow researchers who read their work. In this way, replicating first shifts the focus from the *results* to the *underlying scientific process* (how well the work is carried out). In combination with the registered reports format (Chambers 2013), we believe a replication-first rule would minimize Ph.D. candidates' stress caused by the anticipation of negative results and increase the quality of their work.

Finally, we hope that adopting the proposed replication-first rule would result in an important shift in the necessity for early career researchers to learn and demonstrate the ability to conduct replications appropriately. Specifically, evaluating the outcome of replications often involves assessing the strength of accumulated evidence using state-of-the-art meta-analytic tools. We hope demonstration of such skills will increasingly be taken into account in quality assessment of theses and in hiring decisions. Widespread application of the replication-first rule would also generate pressure on graduate schools to organize corresponding courses and seminars.

Even though adopting the replication-first rule may be difficult in cases where data collection is costly for the budget or resources available for a Ph.D. project, this should not be seen as a sufficient reason to omit replications, as also pointed out by Zwaan et al. Because such studies often have smaller sample sizes and more room for arbitrary data analysis choices, replicability is an even larger issue for them (see Poldrack et al. [2017] for a discussion of this for fMRI findings). The growing awareness of this state of affairs in the field will likely lead to greater appreciation and higher rewards for replication in these cases. Ph.D. candidates are thus well advised to go the extra mile and replicate first. If two separate experiments are not feasible, incorporating a replication into the novel study design would be an option.

In sum, we believe that adopting the replication-first rule for Ph.D. projects would not only contribute to scientific progress in the way Zwaan et al. lay out, but also would be beneficial for the Ph.D. candidates themselves. We predict that this will result in a larger number of solid findings and publishable papers, as well as incentivize Ph.D.'s to master the necessary meta-analytic statistical tools for assessing evidence in cumulative science. In this way, we believe conducting replications could be a great boost for early researchers' careers rather than only a "service to the field." That said, we of course do not suggest obliterating the value of creativity and original thinking in doctoral theses and their assessment. The replication-first rule is intended as a constant reminder that a balance between the two is needed to ensure solid science.

Selecting target papers for replication

doi:10.1017/S0140525X18000742, e139

Anton Kuehberger[a] and Michael Schulte-Mecklenbeck[b,c]

[a]*Centre for Cognitive Neuroscience, University of Salzburg, Salzburg 5020, Austria;* [b]*Department of Business Administration, University of Bern, Bern 3012, Switzerland;* [c]*Max Planck Institute for Human Development, Berlin 14195, Germany.*

anton.kuehberger@sbg.ac.at michael@schulte-mecklenbeck.com

https://ccns.sbg.ac.at/people/kuehberger/

http://www.schulte-mecklenbeck.com

Abstract: Randomness in the selection process of to-be-replicated target papers is critical for replication success or failure. If target papers are

chosen because of the ease of doing a replication, or because replicators doubt the reported findings, replications are likely to fail. To date, the selection of replication targets is biased.

Altough running a replication study is difficult to impossible in some domains, it is quite easy in others. Zwaan et al. (2017) state, in their concern III (sect 5.3), that direct replications are not feasible in some domains, for example, large-scale observational studies, or even not possible, for example, for studies capitalizing on rare events. This argument pertains to domains, but more directly to journal articles, that is, experimental studies. We argue that, beyond the larger obstacles described above, studies that are easier to replicate are indeed more frequently replicated, which introduces selection bias into the replication enterprise.

The selection of a to-be-replicated experiment often depends on how easy it is to do a direct replication. An instructive example is the multilab preregistered replication of the ego-depletion effect (Hagger et al. 2016). The authors selected, as the target of their replication, a procedure introduced by Sripada et al. (2014) and not by the original authors of ego depletion (Baumeister et al. 1998; note that Baumeister recommended the alternative procedure!). The reason for selecting this procedure was described as follows: "tasks used in the original experiments were deemed too elaborate or complex to be appropriate for a multilab replication" (Sripada et al. 2014, p. 548). Why is selection bias at work here? Ease of application is frequently related to the quality of manipulating the independent variable, such that the strength of the manipulation is often limited in easy-to-administer operationalization. A vicious circle is generated: Ease of application breeds a multitude of primary studies (e.g., using simple procedures like questionnaires or vignettes). Many of these studies lead to significant results and therefore publication, but often they are false positives and effect sizes are overestimated (e.g., because of publication bias). If such studies become predominantly targets for direct replication, these replications have little power and are doomed to fail. We end up with many failed replications that are also published, even using the new gold standards of preregistration and multilab collaborations. This circle artificially increases the number of replication failures. The choice of the to-be-replicated target study thus is crucial.

Another selection criterion could be even more harmful: doubt. Many papers become replication targets, not because they are theoretically interesting or important, but because other researchers doubt their results. If there is something to researchers' intuitions of whether a result is likely to be true or not, less likely results have a lower base rate to be true. Even after a significant result, the posterior probability of the hypotheses tested in studies with a small prior is low. Selecting "doubted studies" as targets for direct replication also is doomed to result in failure under most definitions of successful replication. Again, if people select replication targets because they doubt the original findings, and if their doubt is reasonable, the literature will be filled with many failed replications.

The process for choosing the to-be-replicated target study thus is crucial. Ease of application and doubt may contribute to the selection of target papers, leading to an overestimation of replication failures. The best way to avoid this is random selection of replication targets. We pick the Replication Project: Psychology (Open Science Collaboration 2015) as an example of such a procedure. However, inspection of this selection process reveals a variety of judgments, deviating from a purely random choice. The decision tree in Figure 1 illustrates the selection that cuts down an overall 488 articles in the 2008 issues of three journals (*Psychological Science; Journal of Experimental Psychology: Learning, Memory and Cognition; Journal of Personality and Social Psychology*) to an ultimate 100 completed replications.

We identify the following nonrandom selections in the Replication Project: Psychology: (a) publication (only published papers are included); (b) year (papers published in 2008); (c) journal (*Psychological Science, Journal of Experimental Psychology: Learning, Memory and Cognition, Journal of Personality and Social Psychology*); (d) type (488 original research papers); (e) eligibility (158/488, i.e., 32.4%); (f) claim (113 of 158 claimed by replicators, i.e., 71.5%); (g) completion (100 of 113 papers completed and data uploaded to the Open Science Framework within given time frame, i.e., 88.5%). Eventually, a fifth (100/488, i.e., 20.5%) of all possible replications were run and ultimately published. Bias caused by the difficulty of doing a replication surely exists for eligibility (step e, see description in Open Science Collaboration 2015, Methods appendix) and is likely for claiming (step f) and completion (step g). Bias caused by doubt influences claiming (step f). In

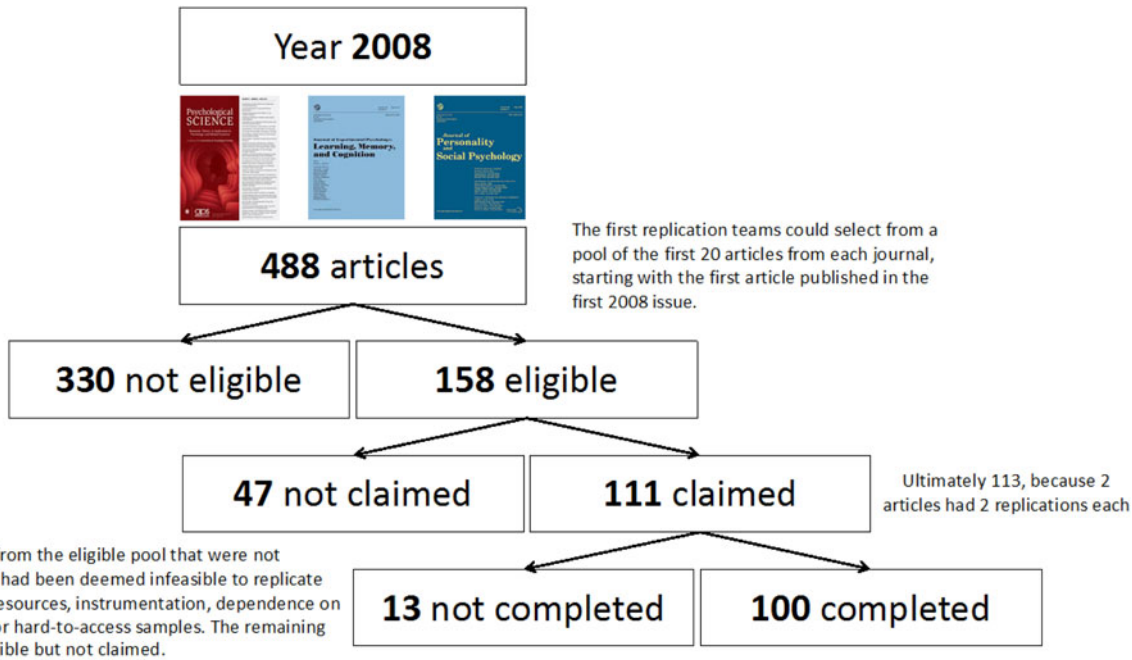

Figure 1. (Kuehberger & Michael Schulte-Mecklenbeck). Selection process for to-be-replicated papers in the Replication Project: Psychology (RPP). Texts in small print are citations from the Open Science Collaboration (2015).