should update their beliefs. If it turns out that researchers do not update their beliefs in this way, we will have learned something. These findings, and the comments of the researchers on why they differed from Streveens’s prescription (if they do), should illuminate how science progresses and how researchers reason.

Such a program may also help to pinpoint the disagreements that can occur between original researchers and replicating researchers. Presently, after a failed replication, a common practice is for authors of the original study to write a commentary. Frequently, the commentary highlights differences between the replication and the original study, sometimes without giving much indication of how much the authors’ beliefs have changed as a result of the failed replication. This makes it difficult to determine the degree of disagreement on the issues.

Our proposal is closely related to several proposed reforms in the literature (and already in the Registered Replication Reports now published by Advances in Methods and Practices in Psychological Science, replicating labs are routinely asked what they expect the effect size to be). The key point is the addition of a suitable quantitative framework. Zwaan et al. mention the “Constraints on Generality” proposal of Simons et al. (2017) that authors should “spend some time articulating theoretically grounded boundary conditions for particular findings” as this would mean disagreements with replicating authors “are likely to be minimized” (sect. 4, para. 11). But it may be difficult for an author to testify that a result should replicate in different conditions, as she is likely to be uncertain about various aspects. Rather than making a black-and-white statement, then, it may be better if the author communicates their uncertainty by attaching subjective probabilities to some of the auxiliary hypotheses involved. A further benefit of this system would be that authors, and the theories they espouse, would then develop a track record of making correct predictions (Rieth et al. 2013).

We recognize that in many circumstances, it may not be realistic to expect researchers to be able to quantify their confidence in the hypotheses that are part and parcel of an original experiment and potential replication experiments. Areas that are less mature, in the sense that many auxiliary hypotheses are uncertain, may be especially poor candidates. But other areas may be suitable. There are good reasons for researchers to try.

An argument for how (and why) to incentivise replication

doi:10.1017/S0140525X18000705, e135

Piers D. L. Howe and Amy Perfors
School of Psychological Sciences, University of Melbourne, VIC 3010, Australia.

pdhowe@unimelb.edu.au amy.perfors@unimelb.edu.au
https://www.findanexpert.unimelb.edu.au/display/person340666
http://psychologicalsciences.unimelb.edu.au/research/coh/csc

Abstract: Although Zwaan et al. (2018) have made a compelling case as to why direct replications should occur more frequently than they do, they do not address how such replications attempts can best be encouraged. We propose a novel method for incentivising replication attempts and discuss some issues surrounding its implementation.

Zwaan et al. (2018) convincingly argue that replication attempts should become mainstream, but they say little as to how this can best be achieved. The problem is that there are currently few mechanisms in place to encourage replication attempts. For example, a survey conducted in 2015 found that only 3% of psychology journals explicitly state that they will consider publishing replications (Martin & Clarke 2017). Although there have been some notable attempts to encourage more replications (Klein et al. 2014a; Open Science Collaboration 2015), they have been of limited scope, and replications remain scarce: A survey of the top 100 psychology journals found that only 1% of reported studies involved replication (Makel et al. 2012). Given the enormous publication pressures on academics, if replications are rarely publishable, then a mainstream culture of replication will not emerge.

Here, we propose a novel solution to this problem: Make it standard practice for journals to pre-commit to publishing adequately powered, technically competent direct replications (at least in online form) for any article they publish and link to it from the original article. This would be comparatively simple to implement and would have a relatively low cost, but would greatly change the incentive structure for researchers. It would also lead to a virtuous cycle in which the more replications are published, the more other people would be encouraged to perform replications of their own. Indeed, performing replications might become an important part of academic training. Running replications would enable early postgraduate students to gain valuable skills in research implementation and analysis while also contributing to the scientific literature.

If our proposal were to be adopted, one expectation might be that authors of the original article would discuss the extent to which they predict that their findings would replicate. For instance, authors might become more explicit in identifying when they believe that their findings are likely to apply only to a particular demographic or to occur only in particular circumstances. These discussions would enhance the interpretability of the original article and encourage authors to think more clearly about these issues during the design and analysis of their studies.

Why should journals adopt our proposal? We suggest that a simple modification to the calculation of impact would encourage journals to publish replications of original articles, regardless of how those replications turn out. Currently, the Thomas Reuters journal’s impact factor is determined by the number of citations of that journal within a designated period, divided by the number of citable documents published overall during that period. Importantly, the denominator does not include documents considered to be “Editorial Material”—a term covering a wide range of document types from true editorials to commentaries such as this one (even when the commentaries report original data). It should be comparatively simple to agree that non-peer-reviewed, online-only, direct replication attempts should also not count toward the denominator. If so, then hosting direct replication attempts on journals’ websites would never hurt. Indeed, if these replication attempts could still be cited (just like editorials can be cited), they would only increase the journal’s impact factor. This creates an incentive for journals to publish replications, which is a necessity for replications to become mainstream.

What about funding agencies? Like journals, grant agencies greatly value novelty, but they even more greatly value reliable science; a novel finding can have a long-term impact only if it is true. It should, therefore, be in a funding body’s interest either to offer grants that are focused solely on replication or to mandate that a certain percentage of each grant be devoted to replicating previous research.

In one sense, our suggestion is a minor alteration in how science is traditionally done but, in another sense, it is a paradigm shift in how to evaluate scientific work. Although novelty and originality are clearly vital, replicability is no less important. Our failure to systematically replicate our findings results in biased estimates of effect sizes, hampers future work, and makes it hard to obtain a realistic evaluation of what we know (Anderson et al. 2017). Because the best way to obtain accurate estimates of a finding’s effect size and robustness is to combine multiple independent replication attempts, we need to actively encourage replications. Within our paradigm, the initial publication of an article is just the starting point in an extended conversation that will conclude with a multitude of replication attempts, an increasingly accurate estimate of the effect size, and a much greater understanding of the circumstances for which the findings hold.
How might we appropriately acknowledge replication attempts for the purposes of career advancement? One obvious possibility would be to adopt a convention on curricula vitae in which replication attempts are classified as distinct from other types of publications — much as books, journal articles, and conference proceedings are classified separately now. It would then be up to the individual’s university, grant review panel or promotion committee to decide how much to value replication attempts relative to other forms of publication.

Our proposal represents a “win” for academics, journals, and the progress of science as a whole. The ability to easily publish replications would mean that academics would be incentivised to perform replications. Indeed, doing so may become a routine and accepted part of academic training. Within a culture of replicability, the impact of any single replication failure would diminish, making replications less personally threatening and simply part of the process (much as reviews are part of science now). Journals would increase in prestige and citation rates by publishing replications. Fundamentally, incentivising replication attempts is the only way to achieve a mainstream culture of replicability. It is vital for our future that science is built on truth rather than sand.

How to make replications mainstream

doi:10.1017/S0140525X18000717, e136

Hans IJzerman,a Jon Grahe,b and Mark J. BrandtC

*Department of Psychology (LIP/P2S), Université Grenoble Alpes, BP 47-38040, Grenoble Cedex 9, France; †Department of Psychology, Pacific Lutheran University, Tacoma, WA 98447; ‡Department of Social Psychology, Tilburg University, Tilburg 5000 LE, The Netherlands.

h.ijzerman@gmail.com graheje@plu.edu m.j.brandt@tilburguniversity.edu www.hansijzerman.org www.tlsaboratory.com

Abstract: Zwaan et al. integrated previous articles to promote making replications mainstream. We wholeheartedly agree. We extend their discussion by highlighting several existing initiatives – the Replication Recipe and the Collaborative Education and Research Project (CREP) - which aim to make replications mainstream. We hope this exchange further stimulates making replications mainstream.

Zwaan et al. integrated previous articles to promote making replications mainstream. We wholeheartedly agree. We extend their discussion by highlighting several existing initiatives that aim to make replications mainstream and that have already helped resolve several of the concerns discussed by Zwaan et al. Specifically, we discuss how to Increase Replication Quality and how to Make Replications Habitual. These facets should facilitate addressing the concerns of not having a standard method and that expertise of the original and replication authors may not be sufficiently relevant.

Increasing replication quality. Zwaan et al. discussed criticisms of the limited theoretical value of replication and the role of contextual variables in replications. This criticism stems from a well-known discussion in psychology whether quality of research should be results- or theory-centered (e.g., Greenberg et al. 1988; Greenwald et al. 1986). One strategy to resolve the conflict between theoretical value and obtained results is to follow the guidelines outlined in the Replication Recipe (RR; Brandt, IJzerman et al. 2014). The RR suggests that replications include 36 “ingredients” for high-quality replications (including, but not limited to, choosing a finding with high replication value, sufficient power, exclusion criteria that are defined a priori, identified differences between original and replication studies, and pre-registration). Following the RR helps replication researchers identify the central parameters of a study and thus the key components of the replication, so that the replication is as convincing as possible. This not only facilitates communication between original and replication researchers, but also between readers of both the replication and the original research. The RR, for example, suggests that replication researchers list contextual features that likely differ between the original and replication research (e.g., different cultural setting? Different population?). This helps communicate to the original authors and readers what the differences in the studies are and the degree the study is a direct or more of a conceptual replication. There may not always be agreement on these designations, but at least the information is clearly available for the reader to make up their own mind. The results from the RR can also be used by future scholars to identify (and then test with pre-registered studies) potential moderators of the effect across both original studies and replication studies, increasing the theoretical value of replications.

Interestingly, Zwaan et al. misinterpreted the RR as something that should be included in original articles. Our original paper was focused on replications and so we did not discuss original articles, but this misinterpretation highlights the important point that many, if not all, of the qualities of a convincing and high-quality replication are exactly the same as the qualities of a convincing and high-quality original study. Therefore, authors can specify the conditions they consider necessary and relevant for their finding and any limits on generalizability (Simons et al. 2017), resulting in increasingly specified psychological theories.

Making replications habitual. Another key facet to making replication mainstream is making replications habitual. One way of doing so is by developing an appreciation for replication early in the academic career. We created the Collaborative Education and Research Project (CREP; Grahe et al. 2015) with the goal of training undergraduate researchers to conduct high quality replication research through standardized procedures as part of research methods courses. The CREP board selects - through a rigid selection process - impactful studies that are feasible to conduct by bachelor students. Prior to data collection, the CREP board communicates with original authors that we selected their study and invite them to provide any original materials and to comment about any conditions that would facilitate successful replication. Students – often in groups and always under the supervision of a faculty supervisor – create a project page on the Open Science Framework, submit their proposed protocol (including video, methods, and evidence of international review board approval) for review by a CREP review team (three advanced researchers and a student administrative advisor). This review process is at least as stringent (and perhaps sometimes more so) than the journal review process. After receiving approval, they complete a general registration of their study, and then collect data. Upon project completion, they go through a second review where the CREP review team reviews their presentation of their data and findings.

CREP projects directly contribute to the research literature by reporting high-quality replications (with one manuscript published [Leighton et al. 2018] and two more in progress [Ghel et al. in preparation; Wagge et al. in preparation]). Additionally, and more importantly, the CREP educates students about modern psychological research methods, training them to be the researchers with the relevant expertise we need. These skills transfer to original research. Students must understand the hypothesis and theory from the original study as they identify which materials are necessary in an original study. They learn to properly document a study (including, but not limited to, obtaining informed consent, collecting and analyzing data, and reporting findings requires the same resources as original research). Further, by interacting with the CREP team, these students experience a review process with faculty at different institutions than their own. As a bonus, instructors are not challenged with reading and supervising poorly conceptualized or poorly planned research that is developed quickly, without adequate preparation that can understandably be typical of students’ first research project.