

The red thread in the maze

D. Samuel Schwarzkopf

School of Optometry & Vision Science, University of Auckland, New Zealand & Experimental Psychology, University College London, United Kingdom

Doing science can feel like running in a maze. If we base our research on false assumptions, we can easily go astray – just ask the phrenologists. Scientists must therefore validate their claims through direct replication of previous findings. But certainly, this mustn't come at the expense of actually moving the field forward. Nobody wants to err aimlessly through a maze, but you also won't find your way out, if you're too scared to walk at all. Replication should be the red thread we unroll behind us as we traverse this maze, lest those who follow us waste endless time and resources getting stuck in dead-ends.

In this issue of *Cortex*, Huber et al. argue that the balance in cognitive neuroscience is still tipped too far towards novelty and narrative at the expense of scientific reliability. To encourage replication, and thus improve the reliability of published research, they suggest we must raise the bar for accepting novel results by mandating internal replication and preregistered designs. While I entirely agree with promoting replication in general, I nevertheless feel their proposals are somewhat unclear.

Publication should not depend on results

For novel, "high-impact" results – often serendipitous findings that involved at least some degree of exploration – it makes perfect sense to ask for an internal replication before publication to ensure robustness. This replication should ideally be a *Registered Report*, where the hypotheses and methods are peer-reviewed and preregistered before data collection commences (<https://cos.io/rr>; Chambers, 2013). This ensures that authors have maximal freedom to explore their data in the initial discovery, but also guarantees adequate statistical power and minimises methodological flexibility for the confirmation.

But successful replication should not be a *condition* of publication. If we reject the study if the original finding fails to replicate, this actually violates the core philosophy behind *Registered Reports*: acceptance for publication must not be contingent on the results. This contingency is the very driving force behind publication bias and the file-drawer problem. It certainly would *not* reduce the number of researchers being stuck in dead-ends that nobody knows about, and may in fact put authors under pressure to successfully replicate their initial discoveries.

A better solution in my view is to continue the publication of exploratory findings but explicitly label them as such. Where available, this could take the form of an

Exploratory Report (McIntosh, 2017)); more generally, it simply entails that the initial result is labelled as "Pending replication". During the peer-review of such a finding authors are given the option to flesh out the design of an internal replication. This second stage then constitutes a *Registered Report* that is preregistered, conducted, and published independently of the initial result – but upon completion, the two results form a unit and could eventually even receive the same DOI. This approach would give the authors an incentive to conduct their replication without running the risk of wasting all that effort on nothing: both the exploratory finding and its replication attempt are publications in their own right.

End our addiction to "high impact" journals

Huber et al. seem to argue that the same journal that published an original finding must also publish attempts to replicate it. However, by focussing on the journals, with the unstated implication that this particularly concerns "high impact" publications, in my mind this idea merely strengthens our field's obsession with impact and narrative over robust science.

To be clear, I agree that journals should publish replication attempts of studies they published. But rather than slowly convincing individual editorial boards and publishers in gilded corner offices to adopt this "Pottery Barn Rule" (Srivastava, 2012), it seems far more important to me that replication efforts are widely *visible*. We do not need high impact journals for that but means to publicise replication attempts more effectively. Indeed, Huber et al. also suggest something along these lines: each study should come with links to any replication attempts so that readers, especially those from outside the immediate subfield of research, can assess the strength and reliability of a finding.

No more significance counting

Huber et al. propose that the links to curate the reliability of a finding are classified into replication failures and successes. But this is far from trivial. How do you define a failure to replicate? While intuitively appealing, basing this purely on significance tests is insufficient. Even for some of the major replication efforts (Open Science Collaboration, 2015), both the original results as well as the replications actually provided only inconclusive evidence (Etz & Vandekerckhove, 2016). Moreover, the critical test of a replication is whether the replication is reliable and consistent with the original effect (Boekel et al., 2015; Nieuwenhuis, Forstmann, & Wagenmakers, 2011), not whether or not the replication is significant.

Therefore, the list of replication links should come with effect sizes and confidence intervals. Additionally, there could even be a web applet that allows a quick meta-analysis. The list should also flag which of the replications were Registered Reports. Taken together, this would allow readers to make up their own mind about the strength of the available evidence – including singular findings that are clearly already highly reliable even though nobody has yet attempted to replicate them.

Be the change you want to see in the world

The thing is, we have heard all this before. It is easy to say there must be more replication – but nothing will change unless we change. Everybody knows that the incentive structures in science are a problem (Yarkoni, 2018). As long as we use high-impact publications as a metric for hiring, funding, and tenure decisions, and as long as funding agencies expect grant proposals to be all about ground-breaking impact, we are actively discouraging replications.

The change must start with us. We must take steps to reward reliable and careful science when we sit on grant panels, tenure committees, or on editorial boards. Furthermore, we must lobby funders and use our influence on grant panels to place greater value on replication, so that a healthy proportion of research funding supports that purpose. I don't mean single grants dedicated to replication efforts (<http://doi.org/10.1038/nature.2016.20287>). While well intended, such initiatives cannot possibly sustain all necessary replications. Rather, every research grant proposal should clearly state how it seeks to confirm the past research on which it is based.

We must argue that this is best for everyone. By encouraging that every funded project also involves preregistered replications, we can actually guarantee that the grant yields publications. Surely, it must be also more economical to spend a proportion of research funds on ensuring reliable discoveries than focusing on novelty and impact alone. Because when planning to run into a maze, it pays to spend a little extra on that little ball of thread to help you find your way back out.

References

- Boekel, W., Wagenmakers, E.-J., Belay, L., Verhagen, J., Brown, S., & Forstmann, B. U. (2015). A purely confirmatory replication study of structural brain-behavior correlations. *Cortex*. <https://doi.org/10.1016/j.cortex.2014.11.019>
- Chambers, C. D. (2013). Registered reports: a new publishing initiative at *Cortex*. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 49(3), 609–610. <https://doi.org/10.1016/j.cortex.2012.12.016>
- Etz, A., & Vandekerckhove, J. (2016). A Bayesian Perspective on the Reproducibility Project: Psychology. *PloS One*, 11(2), e0149794. <https://doi.org/10.1371/journal.pone.0149794>
- McIntosh, R. D. (2017). Exploratory reports: A new article type for *Cortex*. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 96, A1–A4. <https://doi.org/10.1016/j.cortex.2017.07.014>
- Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience*, 14(9), 1105–1107. <https://doi.org/10.1038/nn.2886>
- Open Science Collaboration. (2015). PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science (New York, N.Y.)*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>

Srivastava, S. (2012, September 27). A Pottery Barn rule for scientific journals. Retrieved October 28, 2018, from <https://thehardestscience.com/2012/09/27/a-pottery-barn-rule-for-scientific-journals/>

Yarkoni, A. T. (2018, October 2). No, it's not The Incentives—it's you. Retrieved October 3, 2018, from <http://www.talyarkoni.org/blog/2018/10/02/no-its-not-the-incentives-its-you/>