



The power is in collaboration: Developing international networks to increase the reproducibility of science

XENIA SCHMALZ¹

1. Università degli Studi di Padova

READ REVIEWS

WRITE A REVIEW

CORRESPONDENCE:

xenia.schmalz@gmail.com

DATE RECEIVED:

April 27, 2016

DOI:

10.15200/winn.146178.82672

ARCHIVED:

April 27, 2016

KEYWORDS:

#LJAFreproducibility ; Power;
Social media

CITATION:

Xenia Schmalz, The power is in collaboration: Developing international networks to increase the reproducibility of science, *The Winnower* 3:e146178.82672, 2016, DOI: 10.15200/winn.146178.82672

© Schmalz This article is distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author and source are credited.



The field of psychological science is in a pandemonium. With failures to replicate well-established effects, evidence for a skewed picture of science in the published literature, and a media hype about the replication crisis – what is left for us to believe in these days?

Luckily, researchers have done what they do best – research – to try to establish the causes, and possible solutions to this replication crisis. A coherent picture has emerged. There are three key factors that seem to have led to the replication crisis: (1) Underpowered studies, (2) publication bias, and (3) questionable research practices. Studies in psychology often test a small number of participants. As effects tend to be small and measures noisy, larger samples are required to reliably detect an effect. An *underpowered study*, trying to find a small effect with a small sample sizes, runs a high probability of not finding an effect, even if it is real (Button et al., 2013; Cohen, 1962; Gelman & Weakliem, 2009)

By itself, this would not be a problem, because a series of underpowered studies can be, in principle, combined in a meta-analysis to provide a more precise effect size estimate. However, there is also *publication bias*, as journals tend to prefer publishing articles which show positive results. Authors often do not even bother trying to submit papers with non-significant results, leading to a file-drawer problem (Rosenthal, 1979). As the majority of research papers are underpowered, the studies that do show a significant effect capture the outliers of a normal distribution around a true effect size (Ioannidis, 2005; Schmidt, 1992, 1996). This creates a biased literature: even if an effect is small or non-existent, a number of published studies can provide apparently consistent evidence for a large effect size.

The problems of low power and publication bias are further exacerbated by questionable research practices, where researchers – often unaware that they are doing something wrong – use little tricks to get their effects above a significance threshold, such as removing outliers until the threshold is reached, or including *post-hoc* covariates (John, Loewenstein, & Prelec, 2012; Simmons, Nelson, & Simonsohn, 2011).

As a lot of research and discussion exists on how to fix the problem of publication bias and questionable research practices – which mostly require a top-down change of the incentive structure. Here, I focus on the issue of underpowered studies, as this can be addressed by individual researchers. Increasing power is in everyone's best interests: It strengthens science, but it also gives the researcher a better chance to provide a meaningful answer to their question of interest.

On the surface, the solution to the problem of underpowered studies is very simple: we just have to run

bigger studies. The simplicity is probably why this issue is not discussed very much. However, the solution is only simple if you have the resources to increase your sample sizes. Running participants takes time and money. Therefore, this simple solution poses another problem: the possibility of creating a Matthew effect, where the rich get richer by producing large quantities of high-quality research, while researchers with fewer resources can produce either very few good studies, or numerous underpowered experiments for which they will get little recognition*.

On the surface, the key to avoiding the Matthew effect is also simple: if the rich collaborate with the poor, even researchers with few resources can produce high-powered studies. However, in practice, there are few perceived incentives for the rich to reach out to the poor. There are also practical obstacles for the poor in approaching the rich. These issues can be addressed, and it takes very little effort from an average researcher to do so. Below, I describe why it is important to promote collaborations in order to improve replicability in social sciences, and how this could be achieved.

WHY?

In order to ensure the feasibility of creating a large-scale collaboration network, it would be necessary to promote the incentives for reaching out to the poor. Collecting data for someone with fewer resources may seem like charity. However, I argue that it is a win-win situation. Receivers are likely to reciprocate. If they cannot collect a large amount of data for you, perhaps they can help you in other ways. For example, they could provide advice on a project with which you got stuck and which you had abandoned years ago; they could score that data that you never got around to having a look at; or simply discuss new ideas, which could give you a fresh insight into your topic. If they collect even a small amount of data, this could improve a dataset. In the case of international collaborations, you would be able to recruit a culturally diverse sample. This would ensure that our view of psychological processes is generalisable beyond a specific population (Henrich, Heine, & Norenzayan, 2010).

HOW?

There are numerous ways in which researchers can reach out to each other. Perhaps one could create an online platform for this purpose. Here, anyone can write an entry for their study, which can be at any stage: it could be just an idea, or a quasi-finished project which just needs some additional analyses or tweaks before publication.

Anyone can browse a list of proposed projects by topic, and contact the author if they find something interesting. The two researchers can then discuss further arrangements: whether together, they can execute this project, whether the input of the latter will be sufficient for co-authorship, or whether the former will be able to reciprocate by helping out with another project.

Similarly, if someone is conducting a large-scale study, and if they have time to spare in the experimental sessions, they could announce this in a complementary forum. They would provide a brief description of their participants, and offer to attach another task or two for anyone interested in studying this population.

To reach a wider audience, we could rely on social media. Perhaps a hashtag could be used on twitter. Perhaps #LOOC ("LOOKing for Collaborator")**? One could tweet: "Testing 100 children, 6-10 yo. Could include another task up to 15 minutes. #LOOC". Or: "Need more participants for a study on statistical learning and dyslexia. #LOOC", and attach a screen shot or link with more information.

In summary, increasing sample sizes would break one of the three pillars of the replication crisis: large studies are more informative than underpowered studies, as they lead to less noisy and more precise effect size estimates. This can be achieved through collaboration, though only if researchers with resources are prepared to take on some amount of additional work by offering to help others out. While this may be perceived as a sacrifice, in the long run it should be beneficial for all parties. It will both become easier to diversify one's sample, and help researchers who study small, specific populations (e.g., a rare disorder), to collaborate with others to recruit enough participants to draw meaningful

conclusions. It will provide a possibility to connect with researchers from all over the world with similar interests and possibly complementary expertise. And in addition, it will lead to an average increase in sample sizes, and reported effects which can be replicated across labs.

REFERENCES

- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Confidence and precision increase with high statistical power. *Nature Reviews Neuroscience*, 14(8). doi:10.1038/nrn3475-c4
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, 65(3), 145-153.
- Gelman, A., & Weakliem, D. (2009). Of beauty, sex and power: Too little attention has been paid to the statistical challenges in estimating small effects. *American Scientist*, 97(4), 310-316.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3), 61-83.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *Plos Medicine*, 2(8), 696-701. doi:10.1371/journal.pmed.0020124
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 0956797611430953.
- Rosenthal, R. (1979). The "File Drawer Problem" and Tolerance for Null Results. *Psychological Bulletin*, 86(3), 638-641.
- Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, 47(10), 1173.
- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, 1(2), 115-129. doi:10.1037//1082-989x.1.2.115
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 0956797611417632.

FOOTNOTES

* One may or may not consider this a problem – after all, the issue of the replicability crisis is solved.

** Urban dictionary tells me that "looc" means "Lame. Stupid. Wack. The opposite of cool. (Pronounced the same as Luke.)"