cortex 113 (2019) 360–362

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Discussion forum

Transcending humanness or: Doing the right thing for science

Michael Inzlicht [a,b,*]

[a] *Department of Psychology, University of Toronto, Canada*

[b] *Rotman School of Management, University of Toronto, Canada*

Science is hard. Yes, it leads us to truth, but it doesn't lead us there in a straight line; instead it meanders slowly and circuitously, sometimes taking us down wrong paths altogether. But, science eventually gets it right, it eventually corrects course, finding its slow way to truth. My home discipline of psychology is now going through just such a course correction, with allied fields such as neuroscience following suit. But why is such a correction needed in neuroscience? Certainly the now widely known errors committed by social scientists (and their attendant journals and scientific societies) are not being committed by neuroscientists. Right?

Huber, Potter, and Huszar (2019) recount a story that should cause us to doubt this convenient fiction. In brief, Huber et al. (2019) tried in earnest to build upon a finding they found noteworthy (Wimber, Alink, Charest, Kriegeskorte, & Anderson, 2015). However, after three replication attempts, years in the making, they concluded that they could not (Potter, Huszar, & Huber, 2018). When they tried to publish the results of their labour, they found that the journal that published the original paper—the venerable *Nature Neuroscience*—was not interested in publishing their replication attempt; this prestigious journal was unconcerned with correcting the record.

Huber and colleagues' story is a familiar one. Those who have been paying attention to the methodological reform movement in psychology are depressingly familiar with this sort of story (Lilienfeld & Waldman, 2017; Open Science Collaboration, 2015; Spellman, 2015). The only notable feature here is that this story is being told by hard neuroscientists and not us soft social scientists. This story reveals that neuroscience is committing some of the same errors as psychology. It reveals that important neuroscience journals appear uninterested in publishing replications. It reveals that essential neuroscience journals seem uninterested in publishing null results. It also reveals that some neuroscience papers might have been produced by what we now call questionable research practices, in this case running studies with low statistical power and making strong claims from rather modest evidence.

I won't bore you with long details about why being unconcerned with direct replications, null results, and statistical impotence is problematic for neuroscience. Other neuroscientists, far smarter than I, have already covered this terrain (Cremers, Wager, & Yarkoni, 2017; Geuter, Qi, Welsh, Wager, & Lindquist, 2018; Poldrack et al., 2017; see, also; Szucs & Ioannidis, 2017). But, please allow me to add to their refrain, just a little.

Direct replications, especially as conducted by independent labs, demonstrate repeatability of a finding or phenomenon. It is the lifeblood of science. It is how we determine if something is genuine, and without it, science loses its epistemological edge. This is science 101. And one of the lessons of the so-called replication crisis is that we psychologists have learned just how rarely direct replications were routinely conducted (notice my use of the past tense). Huber and colleagues' (2019) analysis suggest direct replications are even rarer in neuroscience.

While null results are sometimes hard to interpret and deserve scrutiny, it is essential that we enumerate them so that we can calibrate the strength of our accumulated evidence. When we choose not to publish null results we rob ourselves of the denominator telling us just how impressed we should be with our ostensible successes. After all, our estimation of the robustness of an effect that has hundreds of studies appearing to support it (Cunningham & Baumeister,

* University of Toronto, Department of Psychology, 1265 Military Trail Toronto, Ontario, M1C 1A4, Canada.
E-mail address: michael.inzlicht@utoronto.ca.
https://doi.org/10.1016/j.cortex.2018.11.032

2016) hinges radically on whether a total of hundreds versus thousands of studies were run (Friese, Loschelder, Gieseler, Frankenbach, & Inzlicht, 2018). The lack of a denominator is why meta-analyses have been rendered practically useless (Ioannidis, 2016); if a meta-analysis does not include all the studies that didn't work, all the nulls, its results are uninterpretable at best, misleading at worst (Carter & McCullough, 2014; Hagger, Wood, Stiff, & Chatzisarantis, 2010).

Then there is the controversial issue of questionable research practices, more poetically known as p-hacking (Simmons, Nelson, & Simonsohn, 2011). In brief, these are practices that mess with our ability to make correct inferences because they not only inflate the rate of false positives they inflate the size of positive effects, sometimes dramatically (Button et al., 2013). Neuroscientists can commit any number of inferential sins, but among the more problematic are running statistically underpowered studies and overfitting data by chasing significance (Gelman & Loken, 2014). Even though sample sizes in neuroimaging have increased over time, they are still far below what is required to attain even minimally acceptable statistical power (Poldrack et al., 2017). What this means is that the typical neuroimaging study is sufficiently powered to reliably capture only the very large (yet very rare) effect (Geuter et al., 2018). Exacerbating this power failure, neuroimaging researchers appear to routinely hypothesize after the results are known or, worse, specify their hypothesized regions of interest after results are known (Poldrack et al., 2017). Analysis of neuroimaging data typically involve hundreds of small decisions, with tens of thousands of possible workflows (Carp, 2012a). What this means is that there is practically limitless flexibility in data analysis, which can markedly increase the rate of false discoveries. Exploration of data is good, it is even required; but, such explorations will only produce reliable patterns if they are coupled with confirmatory replications.

So, just like us soft psychologists, hard neuroscientists appears to be making mistakes that can affect replicability. A number of neuroscientists have sounded the alarm in relation to this issue (Barch & Yarkoni, 2013; Carp, 2012b; Cremers et al., 2017; Geuter et al., 2018; Poldrack et al., 2017; Vul & Pashler, 2012), and have taken steps to remedy it (e.g., Botvinick-Nezer et al., 2018). What the neuroscience establishment does next, however, is critical. If we learn any lessons from psychology's replication woes, it is that admitting to our past errors is hard, really hard. It is human to want to defend one's record. It is human to want to deflect blame onto others, to look for reasons that one's initial findings were fine and good and robust—maybe the replication team was incompetent, maybe they didn't conduct a close-enough replication, maybe participants in Amherst are truly different than participants in Cambridge, etc. My point here is that it is natural to want to protect oneself, one's reputation, one's record. However, just because it is natural, human even, it doesn't mean that we should do it.

Admitting one's errors is painful to fathom, scarier to do. It is especially painful when the errors look so systematic and widespread. Science, however, demands no less. It is only after we see our errors that we can correct them. And the more of us admit that we have strayed from our scientific ideals—and by "we" I am not only referring to individual scientists, but also to journals, editors, reviewers, scientific societies, and granting agencies—the sooner we can find our way again.

Science is self-correcting, we are told. But, it can only be self-correcting if we transcend our human need to defend and instead acknowledge there is something to correct.

Acknowledgments

This research was supported by grant RGPIN-2014-03744 from the Natural Sciences and Engineering Research Council, Canada and by grant 435-2014-0556 from the Social Sciences and Humanities Research Council, Canada.

REFERENCES

Barch, D. M., & Yarkoni, T. (2013). Introduction to the special issue on reliability and replication in cognitive and affective neuroscience research. *Cognitive, Affective, & Behavioral Neuroscience, 13*(4), 687–689. https://doi.org/10.3758/s13415-013-0201-7.

Botvinick-Nezer, R., Camerer, C., Gorgolewski, C., Dreber, A., Holzmeister, F., Huber, J., et al. (2018). *Neuroimaging analysis replication and prediction study*. Retrieved November 28, 2018, from https://www.narps.info/.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376. https://doi.org/10.1038/nrn3475.

Carp, J. (2012a). On the plurality of (methodological) worlds: Estimating the analytic flexibility of fMRI experiments. *The Florida Nurse, 6*, 149. https://doi.org/10.3389/fnins.2012.00149.

Carp, J. (2012b). The secret lives of experiments: Methods reporting in the fMRI literature. *NeuroImage, 63*(1), 289–300. https://doi.org/10.1016/j.neuroimage.2012.07.004.

Carter, E. C., & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: Has the evidence for ego depletion been overestimated? *Frontiers in Psychology*, 5(July), 1–11. https://doi.org/10.3389/fpsyg.2014.00823.

Cremers, H. R., Wager, T. D., & Yarkoni, T. (2017). The relation between statistical power and inference in fMRI. *Plos One, 12*(11), e0184923. https://doi.org/10.1371/journal.pone.0184923.

Cunningham, M. R., & Baumeister, R. F. (2016). How to make nothing out of something: Analyses of the impact of study sampling and statistical interpretation in misleading meta-analytic conclusions. *Frontiers in Psychology, 7* (OCT) https://doi.org/10.3389/fpsyg.2016.01639.

Friese, M., Loschelder, D. D., Gieseler, K., Frankenbach, J., & Inzlicht, M. (2018). Is ego depletion real? An analysis of arguments. *Personality and social psychology review*. https://doi.org/10.1177/1088868318762183.

Gelman, A., & Loken, E. (2014). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Psychological Bulletin, 140*(5), 1272–1280. https://doi.org/dx.doi.org/10.1037/a0037714.

Geuter, S., Qi, G., Welsh, R. C., Wager, T. D., & Lindquist, M. A. (2018). *Effect size and power in fMRI group Analysis*. BioRxiv (abstract) http://biorxiv.org/content/early/2018/04/06/295048.

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin, 136*(4), 495–525. https://doi.org/10.1037/a0019486.

Huber, D. E., Potter, K. W., & Huszar, L. D. (2019). Less “Story” and more “Reliability” in cognitive neuroscience. *Cortex, 113*, 347–349.
Ioannidis, J. P. A. (2016). The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *Milbank Quarterly, 94*(3), 485–514. https://doi.org/10.1111/1468-0009.12210.
Lilienfeld, S. O., & Waldman, I. (Eds.). (2017). *Psychological science under scrutiny. Psychological science under scrutiny: Recent challenges and proposed solutions*. New York, NY: Wiley/Blackwell. https://doi.org/10.1002/9781119095910.
Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science (New York, N.Y.), 349*(6251). aac4716 https://doi.org/10.1126/science.aac4716.
Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., et al. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience, 18*(2), 115–126. https://doi.org/10.1038/nrn.2016.167.
Potter, K. W., Huszar, L. D., & Huber, D. E. (2018). Does inhibition cause forgetting after selective retrieval? A reanalysis and failure to replicate. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 104*, 26–45. https://doi.org/10.1016/j.cortex.2018.03.026.
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, 22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632.
Spellman, B. A. (2015). A short (personal) future history of revolution 2.0. *Perspectives on Psychological Science, 10*(6), 886–899. https://doi.org/10.1177/1745691615609918.
Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology, 15*(3), e2000797. https://doi.org/10.1371/journal.pbio.2000797.
Vul, E., & Pashler, H. (2012). Voodoo and circularity errors. *NeuroImage, 62*(2), 945–948. https://doi.org/10.1016/j.neuroimage.2012.01.027.
Wimber, M., Alink, A., Charest, I., Kriegeskorte, N., & Anderson, M. C. (2015). Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nature Neuroscience, 18*(4), 582–589. https://doi.org/10.1038/nn.3973.

Received 23 November 2018
Reviewed 28 November 2018
Revised 28 November 2018
Accepted 29 November 2018
Published online 24 December 2018