

are typically quite prestigious and could therefore afford a slight drop in submissions—have no stated penalties for researchers who go against guidelines and refuse to share data. One option is to make it standard journal policy that papers are retracted when authors refuse to share data from recently published papers unless there are compelling mitigating circumstances that prevent sharing. Any inability to share data with interested psychologists should be disclosed to the editor at the time of submission (Kashy, Donnellan, Ackerman, & Russell, 2009).

What are other ways that data sharing can be encouraged? One possibility is simply to make data sharing more normative. If you are interested in someone's data, you should request it and make sure you can replicate their results. In fact, it is probably not a bad idea to ask our close colleagues for their data, just to make the process more commonplace and less adversarial. As anyone who has been asked to share data knows, it only takes 1- or 2-day-long scrambles to compile and annotate existing messy data before you develop better procedures to prevent future occurrences.

In addition to targeting recommendations to those who have leverage, it is also worthwhile considering which recommendations have the largest 'bang for the buck'. It should be clear that many (if not most) studies in psychology are underpowered. The small sample sizes that plague our field have serious consequences in terms of imprecise parameter estimates and reduced credibility. Fortunately, this problem is easy to fix by demanding larger sample sizes. Editors and reviewers should simply require that authors start with the assumption that their effects will be no larger than what is typical for the field unless there is solid evidence that the specific effect under investigation will be larger. Thus, we suggest that power and precision be used as explicit grounds for a desk rejection.

Similarly, replication studies are easy to conduct and will have great benefit for the field. It is less important whether such replications are conducted by students or senior researchers or whether they are published in online repositories or special sections of existing journals. The real issue is making sure that the results are made available and that those who conduct independent replications are given credit for their efforts. Any reader who agrees with the recommendations provided in the target article can make an immediate contribution to the field by committing to conduct regular replications of their own and others' work and to make sure that the results are made accessible. In addition, concerned researchers should consider refusing to support journals that do not publish replications as a matter of policy.

The fact that so much has been written about methodological reform in the last 2 years is both encouraging and depressing. It is encouraging because these articles could be a harbinger of major changes in how psychological science is conducted. Such articles can also be depressing because the current discussions have an eerie similarity to those from the past decades. As it stands, many of the discussions about methodological reform operate on the assumption that there is basic agreement about the ultimate point of psychological research, which is to gain a clearer understanding of reality. However, it might be worth questioning this basic assumption. What if some researchers believe that the point of psychological science is simply to amass evidence for a particular theoretical proposition? Those with such a worldview might find the recommendations provided by the target article to be unnecessary roadblocks that limit their productivity. If so, then methodological reform needs to confront the reality that improving psychological research must involve changing hearts and minds as well as encouraging more concrete changes in behaviours.

Put Your Money Where Your Mouth Is: Incentivizing the Truth by Making Nonreplicability Costly

CORY A. RIETH¹, STEVEN T. PIANTADOSI², KEVIN A. SMITH¹, EDWARD VUL¹

¹University of California, San Diego

²University of Rochester

edwardvul@gmail.com

Abstract: We argue that every published result should be backed by an author-issued 'nonreplication bounty': an amount of money the author is willing to pay if their result fails to replicate. We contrast the virtuous incentives and signals that arise in such a system with the confluence of factors that provide incentives to streamline publication of the low-confidence results that have precipitated the current replicability crisis in psychology. © 2013 The Authors. *European Journal of Personality*

A major part of the replicability 'crisis' in psychology is that commonly reported statistics often do not reflect the authors' confidence in their findings. Moreover, there is little incentive to attempt direct replications, as they are difficult, if not impossible, to publish. We propose a solution to both problems: For each result, authors must name a one-time non-replication 'bounty' specifying the amount they would be willing to pay if the result did not replicate (e.g. $t(30) = 2.40$, $p < .05$, nonreplication bounty: \$1000). Thus, when you report a finding, you are effectively making a one-sided bet: if it

replicates, you gain nothing, but if it fails to replicate, you pay the bounty using personal income. The bounty should be proportional to your confidence—if you are unsure, it could be \$1; if you know the results replicate, it could be a huge sum. This bounty measures the authors' subjective confidence on a scale that is universally interpretable, penalizes authors for overconfidence, and provides direct incentives for replication. Tabling the implementation details, consider the benefits:

(1) *Author confidence is clearly reported*

Ultimately, only the authors know exactly how their study was conducted and how they analysed their results. Their confidence is the best available signal of the robustness of their results, and a nonreplication bounty offers a clear signal of this confidence. This clear signal offers naïve readers an effortless assessment of the soundness of a result, as well as a quantitative metric to evaluate authors and journals. Thus, instead of rewarding raw publication and citation counts and encouraging the frequent publication of surprising, low-confidence results—one systemic problem contributing to the replicability crisis (Ledgerwood & Sherman, 2012)—sound results could be rewarded for both authors and journals.

(2) *Authors have incentive to provide an accurate signal*

The nonreplication bounty is not only a clear signal of confidence but also costly to fake. A low-confidence result offers authors two choices: overestimate their own confidence and suffer a considerable risk, or publish a result with low confidence, which readers will know should be ignored. Neither of these will be appealing, so authors will be altogether less eager to publish low-confidence results. If authors systematically overstate their own confidence, intentionally or not, they will face high costs and will either calibrate or leave the field.

(3) *Replications are directly encouraged*

Replication attempts receive direct incentives: Nonreplications pay a bounty. Moreover, replication attempts would be targeted towards the same results that naïve readers of the literature would have most confidence in: The higher the bounty, the more seriously the result will be taken, and the greater is the incentive for replications. Furthermore, such a system necessitates publication of replication successes and failures, adding further replication incentives.

We believe that many of the other proposed solutions to the replicability crisis ultimately will not work because they fail to provide appropriate incentive to authors (Nosek, Spies, & Motyl, 2012). For instance, the literature has suggested a number of metrics offering more reliable objective signals of result soundness: use of confidence intervals (Cumming & Finch, 2005), effect sizes (Cohen, 1994), Bayesian posterior intervals (Burton, Gurrin, & Campbell, 1998; Kruschke, Aguinis, & Joo, 2012), Bayes factors (Wagenmakers, Wetzels, Borsboom, & Van der Maas, 2011), and various disclaimers pertaining to the analysis procedures (Simmons, Nelson, & Simonsohn, 2012). Although these are useful statistical tools and policies, none is so sound as to avoid the possibility of being gamed, as

they do not make errors costly to the authors. Running many low-powered studies, *post hoc* selection of independent or dependent variables, and other p-hackery (Simmons et al., 2011) would all yield nice results under these metrics. We believe that a remedy to these ailments must provide incentives to authors to offer clear, unbiased estimates of the soundness of their results, in place of the current incentives for authors to directly or indirectly overstate their confidence in and the reliability of their data.

Similarly, many proposals for remedying the replicability crisis (such as the target article) have focused on rules that publication gatekeepers (reviewers and editors) should enforce so as to increase the soundness of results. In contrast, nonreplication bounties would provide a clear and reliable signal that would alleviate some of the burden on volunteer reviewers and editors, rather than increase it. Authors would no longer receive incentives to sneak low-confidence results past reviewers, and reviewers could take on more thoughtful roles in trying to assess the validity of the measures and manipulations: Does the empirical result really have the theoretical and practical implications that the authors claim? Furthermore, as long as we have a reliable confidence signal associated with each result, there need not be an argument about whether type I or type II errors are more worrisome (Fiedler et al., 2012): Journal editors can choose to publish exciting, but speculative, findings or to publish only high-confidence results.

As proposed (Asendorpf et al., this issue; Koole & Lakens, 2012), encouraging replication attempts and the publicity of their outcomes is certainly beneficial. However, without quantitative metrics of result soundness, there is little incentive for journals to publish replications as impact factor only rewards short-term citations, which largely reflect the novelty and newsworthiness of a result.

The status quo indirectly provides incentives for rapid publication of low-confidence outcomes and their misrepresentation as high-confidence results: a practice that appears to be undermining the legitimacy of our science. We believe that local changes that do not restructure authors' incentives are only stop-gaps for a deep-seated problem. Under our scheme, authors would have incentives to offer the most calibrated, precise estimates of the soundness of their available results.

Our position is best summarized by Alex Tabarrok (2012): 'I am for betting because I am against bullshit. Bullshit is polluting our discourse and drowning the facts. A bet costs the bullshitter more than the non-bullshitter so the willingness to bet signals honest belief. A bet is a tax on bullshit; and it is a just tax, tribute paid by the bullshitters to those with genuine knowledge'.

Increasing Replicability Requires Reallocating Research Resources

ULRICH SCHIMMACK AND GIUSEPPINA DINOLFO

University of Toronto Mississauga
uli.schimmack@utoronto.ca

Abstract: We strongly support the recommendation to increase sample sizes. We recommend that researchers, editors, and granting agencies take statistical power more seriously. Researchers need to realize that multiple studies, including exact replication studies, increase the chances of type II errors and reduce total power. As a result, they have to either publish inconclusive null results or use questionable research methods to report false-positive results. Given limited resources, researchers should use their resources to conduct fewer original studies with high power rather than use precious resources for exact replication studies. Copyright © 2013 John Wiley & Sons, Ltd.