

## A New Replication Norm for Psychology

Etienne P. LeBel

University of Western Ontario

**Draft version 2.0, March 20, 2015.** This paper is currently under review at *Collabra*. Please cite as:

LeBel, E. P. (2015). A new replication norm for psychology. *Manuscript currently under review @ Collabra*.

**Corresponding Author:** Etienne P. LeBel, Department of Psychology, University of Western Ontario, Social Science Centre, London, Ontario, Canada, N6A 5C2; **Email:** [etienne.lebel@gmail.com](mailto:etienne.lebel@gmail.com)

Word count: 4,654 (excluding abstract, references)

## Abstract

In recent years, there has been a growing concern regarding the replicability of findings in psychology, including a mounting number of prominent findings that have failed to replicate via high-powered independent replication attempts. In the face of this replicability “crisis of confidence”, several initiatives have been implemented to increase the reliability of empirical findings. In the current article, I propose a new replication norm that aims to further boost the dependability of findings in psychology. Paralleling the extant social norm that researchers should peer review about three times as many articles that they themselves publish per year, the new replication norm states that researchers should aim to independently replicate important findings in their own research areas in proportion to the number of original studies they themselves publish per year (e.g., a 4:1 original-to-replication studies ratio). I argue this simple approach could significantly advance our science by increasing the reliability and cumulativeness of our empirical knowledge base, accelerating our theoretical understanding of psychological phenomena, instilling a focus on quality rather than quantity, and by helping to facilitate our transformation toward a research culture where executing and reporting independent direct replications is seen as a completely ordinary part of the research process.

Word count: 200

### A New Replication Norm for Psychology

There is growing consensus that psychology has a replicability “crisis of confidence” (Pashler & Wagenmakers, 2012; Pashler & Harris, 2012; Simons, 2014; Open Science Collaboration 2012, 2013), stemming from the fact that a growing number of findings cannot be replicated via high-powered independent replication attempts that duplicate the original methodology as closely as possible. Across all areas of psychology, there is a growing list of (prominent) findings that have not held up to independent replication attempts, including findings from cognitive psychology (retrieval-induced forgetting, Maslany & Campbell, 2013; eye movements on recall, Matzke et al., 2014; temporal judgments, Matthews, 2012; protection effect, Wolferen, Inbar, & Zeelenberg, 2013; mental simulation, Zwaan & Pecher, 2012; Mozart effect, Steele, Bass, & Crook, 1999; Steele, Brown, & Stoecker, 1999), developmental psychology (synesthetic cross-modality correspondence, Lewkowicz & Minar, in press), neurophysiology (vestibular stimulation, Lenggenhager, Hilti, Palla, Macaуда, & Brugger, 2014), industrial/organizational psychology (utility biasing effect on selection procedures, Carson, Becker, & Henderson, 1998), positive psychology (weather effects on life satisfaction, Schmiedeberg & Schroder, 2014), political psychology (self-prophecy effect on voting, Smith, Gerber, & Orlich, 2003; status-legitimacy hypothesis, Brandt, 2013), moral psychology (“Macbeth effect”, Earp, Everett, Madva, & Hamlin, 2013), educational psychology (stereotype threat on female math performance, Ganley et al., 2013; Jensen, 2014), color influence on exam performance, Tal, Akers, & Hodge, 2008), evolutionary psychology (fertility on face preferences, Harris, 2011; ovulation on men’s testosterone, Roney & Simmons, 2012; sex differences in infidelity distress, IJzerman et al., 2014), judgment & decision making (unconscious thought advantage, Nieuwenstein et al., in press; choice-overload, Scheibehenne et al, 2010), and social cognition (e.g., “social priming”/embodiment findings, Johnson, Cheung, & Donnellan, 2014; Donnellan, Lucas, & Cesario, 2014; Lynott et al., 2014; Doyen, Klein, Pichon, & Cleeremans, 2012; Harris, Coburn, Rohrer, & Pashler, 2013; McCarthy, 2014;

Pashler, Rohrer, & Harris, 2013; Pashler, Coburn, & Harris, 2012; Steele, 2014; LeBel & Campbell, 2013; LeBel & Wilbur, 2014; Madurski & LeBel, 2014; Hone & McCullough, 2014).

More generalizable evidence supporting a general replicability problem comes from an ambitious and unprecedented large-scale crowdsourced project, the Reproducibility Project (Open Science Collaboration 2012, 2013). In this project, researchers were unable to replicate about 60% (48 out of 82, as of February 2015) of findings quasi-randomly selected from 2008 issues of *Psychological Science*, *Journal of Personality and Social Psychology*, and *Journal of Experimental Psychology: Learning, Memory, and Cognition*. In another large-scale meta-scientific investigation, about 60% (11 out of 27) of important findings from cognitive and social psychology could also not be replicated (Nosek & Lakens, 2014). Though there are different ways to interpret successful versus unsuccessful replication results (Simonsohn, 2013), taken together, these observations strongly suggest psychology currently has a general replicability problem (as do several other areas of science including cancer cell biology and cardiovascular health literatures, Begley & Ellis, 2012; Prinz, Schlange, & Asadullah, 2011).

### **New Initiatives and Reforms**

Several new initiatives have been launched to improve research practices in order to increase the reliability of findings in psychology. For instance, higher reporting standards have recently been instituted at several prominent psychology journals (Eich, 2014; Simmons, Leif, & Simonsohn, 2011, 2012; LeBel et al., 2013). At such journals (e.g., *Psychological Science*, *Memory & Cognition*, *Learning & Behavior*, *Attention, Perception, & Psychophysics*, *Psychonomic Bulletin & Review*, *Personality and Social Psychology Bulletin*), authors submitting a manuscript must now acknowledge that they have disclosed basic methodological details critical for the accurate evaluation and interpretation of reported findings such as fully disclosing all excluded observations, all tested experimental conditions, all assessed outcome measures, and their data collection termination rule.

There is also a significant push to incentivize “open data”, the public posting of the raw data underlying studies reported in a published article (Nosek, Spies, & Motyl, 2012; Simons, 2014). For instance, at *Psychological Science*, authors who make their data publicly available now earn an open data badge that is prominently displayed alongside their published article.<sup>1</sup> Furthermore, the new Journal of Open Psychology Data now publishes data papers that feature publicly posted data sets (Wicherts, 2013). Such open data practices not only facilitate independent verification of analyses and results so crucial to identifying errors and other inaccuracies, but substantially facilitate the execution of meta-analyses and re-analyses from different theoretical perspectives, which can accelerate knowledge development.

In addition, several journals (e.g., *Cortex*, *Perspectives on Psychological Science*, *Attention, Perception, & Psychophysics*, *Comprehensive Results in Social Psychology*) now offer pre-registered publication options whereby authors submit a study proposal that pre-specifies the methodology and analytical approaches to be used to test a specific hypothesis (Chambers, 2013; Wolfe, 2013). Proposals are evaluated on the soundness of the methodology and theoretical importance of the research question. Once accepted, the proposed study is executed and the article is published regardless of the results, eliminating questionable research practices and researcher bias which can grossly mischaracterize the evidence (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012).

A final development is the growing practice of prominent journals to publish independent direct replication results, including replication results inconsistent with those originally published by the journal (e.g., *Psychological Science*, *Psychonomic Bulletin, & Review*, *Journal of Research in Personality*, *Journal of Experimental Social Psychology*). Though calls for the publication of replication results have been made for decades (e.g., Neuliep & Crandall, 1990; Rosenthal, 1997), the actual practice of prominent journals systematically publishing replication

---

<sup>1</sup>Encouragingly, there is already preliminary evidence that such open practices badges are having a significant positive impact. For instance, a rapidly growing number of authors of *Psychological Science* articles have earned open data and materials badges for publicly posting their materials (Wicherts, 2015).

results is unprecedented and has immense potential to increase the reliability of findings in psychology. Such practice directly incentivizes researchers to actually execute independent replications so crucial to ensuring a cumulative knowledge base (Feynman, 1974; Popper, 1934/1992), but also may reduce the tendency for researchers to publish unexpected, exploratory, and/or tenuous results (Schimmack, 2012).

Though this final development is particularly exciting, many researchers are currently afraid or unsure about possible social and career-related risks involved in executing and publishing independent replication results given several recent high-profile cases where the publication of replication results lead to nasty threats, retaliation, and personal attacks of incompetence by original authors (Funder, 2012; e.g., Johnson, Cheung, & Donnellan, 2014; Schnall, 2014; Gilbert, 2014; Bargh, 2012a, 2012b; Yong, 2012). This situation represents a serious hurdle that substantially interferes with the development of a research culture where the execution and publication of independent direct replications is seen as a routine part of the research process rather than something done mostly by selfless “open science” psychologists. To overcome this important hurdle, I propose a new replication norm that has the potential to substantially increase the execution and publication of independent direct replications so important to ensuring a self-correcting cumulative knowledge base. As Cohen (1994) propounded “...we must finally rely, as have the older sciences, on replication” (p. 1002). Similarly, as Sir Ronald Fisher (1926) stated: “A scientific fact should be regarded as experimentally established only if a properly designed (independent) experiment rarely fails to give this level of significance [referring to  $p < .05$ ]” (p. 504).

### **Extant Peer Review Norm**

The new replication norm is inspired directly from the extant peer review social normal that currently exists in psychology and other areas of science. This informal (and implicitly adopted) social norm states that psychologists should aim to review other peers’ papers at a rate approximately three times the number of first-author papers they themselves publish per

year.<sup>2</sup> The threefold rate is based on the logic that most papers are typically reviewed by three other peers. For example, if a researcher publishes 4 first-author publications in a year, they should aim to review (at least) approximately 12 papers submitted by other researchers. Given that each accepted paper collectively costs the field a certain amount of work, the social norm aims to guide researchers to contribute *to the system* as much work reviewing papers as they themselves *cost* the system. Of course, because such a norm is informal – and in no way enforceable – inevitably some individuals may end up “free riding”, that is intentionally or unintentionally drawing more from the collective resource than they themselves contribute to it (Forsythe, 2009; e.g., by publishing 5 papers per year, but only reviewing 3 papers per year, a net deficit of 12 “units” of work.) Notwithstanding such suboptimality, the informal social norm ends up benefiting everyone in terms of clearer and strengthened manuscripts that make more important theoretical contributions to the field.

### **New Replication Norm**

Following directly from the extant peer review social norm, I propose a new replication social norm whereby researchers strive to execute and publish independent replications of other findings in their research area in proportion – in some ratio – to the number of (first-author) studies they themselves publish per year. For example, the norm could be that researchers strive to execute and publish 1 independent direct replication (of another researcher’s finding) for every 4 (first-author) original studies they publish per year. This would mean that if a researcher publishes 3 articles reporting a total number of 8 studies in a year, they would be expected to execute and publish independent replications of (at least) 2 important findings in their own area of research in that same year (in other words, 20%  $[2/2+8]$  of one’s published studies per year should involve independent *direct* or *systematic* replications). Paralleling the

---

<sup>2</sup>Strictly speaking, the social norm should be three times the number of first-author *submissions* rather than *accepted* publications because many papers are rejected at several journals before finding a home. However, given that revised versions of rejected manuscripts submitted to a different journal are often reviewed by some of the same reviewers (and involve less work to review subsequently), for the sake of simplicity, it would appear researchers calibrate their peer review behavior in reference to number of *published* rather than *submitted* articles.

peer review norm, the logic is that each original finding published by a researcher *costs* the collective field a certain amount of work to independently corroborate, hence the replication norm aims to guide researchers to contribute *to the system* a roughly commensurate amount of replication work (of other researchers' findings) as they themselves *cost* the system. The more findings you publish, the more independent replications you need to execute and publish for everyone to benefit from a self-correcting cumulative knowledge base.

**Original-to-replication-studies ratio to strive for?** Though it will be difficult in practice to identify and defend one particular optimal original-to-replication-study ratio, the act of proposing a reasonable ratio is more important than the actual ratio researchers end up adopting and achieving.<sup>3</sup> Nonetheless, I will now present logical and empirical considerations that support the idea that a 4:1 original-to-replication-studies ratio may be a reasonable ratio the modal researcher should strive for.

On logical grounds, it is straightforward that small original-to-replication-studies ratios (e.g., 1:1 or 2:1) are suboptimal given that (1) the primary goal of science is to adduce new facts rather than verify old findings and (2) many findings are never cited nor important, hence valuable resources should not be spent attempting to replicate all findings. Consequently, such small ratios would likely be seen as an unwise use of resources and hence be very unlikely to be adopted by researchers. It would seem, then, that a more optimal ratio would involve a much larger proportion of original compared to replication studies. But how much larger? A large-scale survey asking psychologists' attitudes toward newly proposed research and editorial practices provides empirical evidence as guidance in answering this question (Fuchs, Jenny, & Fiedler, 2012). In that survey, psychologists indicated – on average – that 23% of journal space should be dedicated to direct replications. If psychologists want about 23% of journal space dedicated to direct replications, then one can make the case that a good starting point for the new

---

<sup>3</sup> Also, like the peer review norm, researchers may naturally calibrate their required contribution to the system depending on their faculty position and available resources (e.g., research vs. teaching position, access to large vs. small subject pools, etc.).



replication norm original-to-replication-studies ratio should be about 4:1 (given that 1 divided by 5 is 20%). Assuming that journals actually abide by the democratic voice of the community to dedicate about 20% of their pages to direct replications, then it would make sense for researchers to strive for a 4:1 ratio given that it ensures their replication work will be rewarded. As it currently stands, most replication results are relegated to lower status journals, though some improvements have recently occurred on this front (as mentioned above; see also Funder, 2014). Of course, in practice, not all researchers will be able to accomplish a 4:1 ratio, and a minority of researchers may even disagree on principle in the value of independent direct replications (Mitchell, 2014). Nonetheless, the 4:1 ratio can act as an upper-bound to strive for, and the field can benefit immensely even if researchers' actual modal ratio is much higher. For instance, even if only half of researchers aim for an 8:1 ratio, this would dramatically increase the number of independent replications in the published literature relative to the current state of affairs whereby independent direct replications represent less than 0.2% of the published literature (Makel et al., 2012).

It is important to emphasize that the independent replications executed as part of the new norm need to be *direct* replications (same methodology) rather than *conceptual* replications (different methodology). A direct replication is the attempt to duplicate the conditions and procedure that existing theory and evidence anticipate as necessary for obtaining the effect (Nosek & Lakens, 2014; Open Science Collaboration, 2012). This is crucial because if a study is repeated with a different measure or manipulation and discrepant results emerge, then it is completely ambiguous as to whether the discrepant results are due to the different methodology or because the original finding was false (LeBel & Peters, 2011; Pashler & Harris, 2012). Of course, for some psychological phenomena (e.g., complex culturally-bound or historically-sensitive social psychological phenomena), it may be difficult in practice to know if all essential conditions were duplicated (e.g., identifying a counter-attitudinal essay topic in a different culture). These challenges speak even more loudly to the crucial importance of direct

replications given the potentially large number of confounded variables that may inadvertently arise across original and replication studies (e.g., different language, culture, historical time period, attention span due to technological advancements, etc.)

It many instances, however, it will be more efficient (if possible) to execute *systematic replications* whereby a direct replication of an original result is tested within particular cells of one's design (the "anchor cell"; Hendrick, 1991) with a conceptual replication tested in separate independent cells (e.g., using a different measure or experimental condition; see also Lykken, 1968, who used the term *constructive replications*). The direct replication anchor cell ensures that one can observe the original result in a new independent sample whereas the conceptual replication cells test whether the result generalizes to other measures, manipulations, conditions, or contexts. Table 1 demonstrates a hypothetical example of a systematic replication design of Schnall, Benton, & Harvey's (2008) cleanliness priming on moral judgments finding.

**Table 1: Example systematic replication by adding a new measure.**

		Measures	
		Original moral vignettes	New & improved vignettes
Priming condition	Cleanliness Priming	Direct replication	Conceptual replication
	Control condition		

In this case, participants in the anchor cell would judge the original series of moral actions after having been either primed with cleanliness-related or neutral words, to ensure the original result can be replicated. Completely independent participants would be randomly assigned to either the cleanliness-related versus neutral priming condition and then judge a series of new --

perhaps more rigorously standardized and/or more ecologically-valid moral actions -- to see if the original finding generalizes to these arguably more valid measures (see Table 2 for a systematic replication example where a new independent variable is tested).

**Table 2: Example systematic replication by adding a new independent variable.**

		Cognitive load	
		Control condition	High cognitive load
Priming condition	Elderly priming	Direct replication	Conceptual replication
	Control condition		

This systematic use of direct and conceptual replications is crucial for cumulative knowledge development because as scientists we need to make sure we can replicate past results in our own labs to make sure our instruments, measures, and participants/rats are behaving appropriately. This is arguably all the more important in psychology whereby most psychological phenomena are multiply-determined by large sets of variables that may exert different influences in different contexts for different individuals. Indeed, Feynman (1974) was highly skeptical of psychological studies specifically because he repeatedly observed psychologists not executing such systematic replications.

### **Direct and Indirect Benefits of New Replication Norm**

The new replication norm will have several direct and indirect benefits. First, the norm will significantly increase the overall number of independent direct replications in the literature (currently < 0.2%, Makel et al., 2012). This will be true whether the replication results are formally published in a prominent journal, published in newer open-access journals, or simply

posted to PsychFileDrawer.org. Ultimately, independent replications need to occur more often and be disseminated as widely as possible so that the broader community of scientists can calibrate their confidence in empirical findings accordingly. It is most optimal for the journal that originally published an original finding to publish (unsuccessful) independent direct replication results – a standard known as the “Pottery barn rule” (Srivastava, 2012) – because this most effectively alerts readers of that journal that a finding may not be as robust as initially thought (e.g., LeBel & Campbell, 2013 independent replications of Vess, 2012; both original and replication work published at *Psychological Science*). This is to be contrasted with other situations where prominent journals that published an original finding were unwilling to publish sound high-powered unsuccessful independent replications (e.g., Madurksi & LeBel, 2014; LeBel & Wilbur, 2014; see also LeBel, 2014)

Following directly from the first benefit, a second direct benefit of the new replication norm is that it will facilitate cumulative knowledge development and hence accelerate theory development. That is, accelerate the rate at which we can deepen our theoretical understanding of psychological phenomena. It is straightforward that in the case of successive (published) successful independent corroborations of earlier findings, we amass compelling evidence that our literature does in fact cumulatively build upon itself over time (compared to present reality where this is simply publicly unknown or privately and unsystematically known). In the case of unsuccessful independent direct replications, publicly knowing about a much larger number of such “failed” replication studies will significantly help protect against false, flawed, (or fraudulent) findings from going unquestioned for an extended period of time (Makel et al., 2012). This not only prevents other researchers from wasting precious resources and time following up on blind alleys, more importantly, it forces the field to seek alternative theoretical models that have a higher probability of reflecting actual psychological realities (i.e., publishing failed independent direct replications reduces theoretical false negatives). In the words of Nobel-laureate physicist Richard Feynman, “we are trying to prove ourselves wrong as quickly as possible, because only

in that way can we find (theoretical) progress.” In other words, being proven wrong via independent direct replications is a good thing; it is *not knowing one is wrong* that is catastrophic for theory development.

The new replication norm will also have indirect benefits, meaning that it will benefit the field in ways not directly tied to increasing the execution and publication of independent direct replications. First, it will indirectly incentivize researchers to (1) publish fewer overall papers (per year) and (2) to publish papers with fewer studies. Executing fewer studies per year means there will be more resources available per study executed, hence increasing the quality of the studies executed (e.g., larger sample size, higher power, more precise effect size estimates, more time to more openly document the research process and perhaps even pre-register confirmatory studies, Campbell, LeBel & Loving, 2014). Second, executing and publishing fewer studies per year will noticeably reduce the overall peer-review workload because there will be fewer and shorter papers to independently evaluate. This again means more time to design sounder, costlier, and/or more advanced experimental designs (e.g., highly-repeated within-person designs, experience sampling designs, eye-tracking studies). This should again drive up the quality of the researcher and hence increase the informational value and potential impact of published findings.

A final indirect benefit of the new replication norm is that it will help facilitate our transformation toward a new research culture where executing and reporting independent direct replications is seen as completely routine and mundane part of the research process. This is to be contrasted with the current atmosphere where executing and publish replications can be perilous, given that it is often seen as confrontational or antagonistic (e.g., stepping on others’ toes; Funder, 2012), disrespectful (a senior collaborator used the expression “you’re throwing them under the bus”), or highly risky in terms of one’s career advancement (e.g., replicators may be perceived as having a “difficult personality” that may negatively affect a department’s atmosphere). This is no better demonstrated by several recent high-profile cases where the

publication of replication results lead to nasty personal threats and attacks of incompetence by original authors (Funder, 2012; e.g., Johnson, Cheung, & Donnellan, 2014; Schnall, 2014; Gilbert, 2014; Bargh, 2012a, 2012b; Yong, 2012). The new replication norm will help eliminate misconceptions by the minority of researchers (who unfortunately may influence views of the majority) who see independent replications as a sign of mistrust aimed to discredit the reputations of other researchers. Eliminating such misconceptions is crucial to establishing a new research culture where independent replications are seen as a normal part of the research process.

### **Promoting the Norm**

I contend the new replication norm can be promulgated via different informal and formal channels. First, a natural venue for the dissemination of the new replication norm could take place at student mentoring events at annual society conferences. The norm could also be discussed during professional development sessions often given to graduate students at the beginning of their graduate career. This seems reasonable given this is the context where the peer review norm is typically introduced, including among other things peer reviewing strategies more generally, professional networking at conferences, managing one's website and online presence, etc. Indeed, this approach resonates well with a recent article arguing for the pedagogical benefits of having students execute independent direct replication as part of their undergraduate and graduate experimental methods training (Frank & Saxe, 2012; Janz, 2015). Frank and Saxe have found such an approach to be highly effective in getting students more engaged in learning the intricacies and importance of independent replications given that such classroom projects have the potential of making a real scientific contribution if done well.

Finally, we can look to the social psychology literature to identify factors that may help in increasing adoption of the new replication norm. First, social psychology theorists (Forsythe, 2009; Latané, Williams, & Harkins, 1979) have argued that clearly specifying collective goals can promote pro-social behaviors and minimize free riding (a.k.a., social loafing) in relation to

collectively beneficial social norms, which in part is what I have attempted to achieve in this paper (i.e., strive for a 4:1 original-to-replication studies ratio). Furthermore, social norms theory (Perkins & Berkowitz, 1986) posits that behavior can be influenced by incorrect perceptions regarding how other group members think or act. As applied to alcohol use, for example, undergraduates tend to over-estimate the frequency and quantity of alcohol use by their peers (Perkins, Meilman et al., 1999) and this can increase problem drinking among the misperceiving majority (*pluralistic ignorance*: incorrect perception that majority belief is different from their own belief due to memorable exemplars) and can reinforce problem drinking among the minority of individuals actually exhibiting problematic drinking (*false consensus*: problem drinkers misperceive their behavior as normative). Social norms theory hence implies that correcting misperceptions about the target belief is crucial to maximize norm adoption. In our case, what is the target belief? Though the replication norm situation is arguably more complicated than the alcohol example, a relevant target belief in our context is that a majority of researchers may incorrectly believe (or overestimate the extent to which) independent direct replications will be negatively received and/or create antipathy due to memorable recent exemplars (e.g., the “Repligate” scandal, Johnson et al., 2014; Schnall, 2014; Gilbert, 2014). In this way, social norms theory would predict that it is crucial to raise awareness regarding what psychologists actually believe with respect to direct replications (i.e., approximately 22% of journal space should be dedicated to direct replications, Fuchs et al., 2012) so as to correct misperceptions regarding the extent to which publishing replications will be a negative experience. This in turn should increase the number of researchers who feel comfortable executing and publishing independent replications. Indeed, one could argue that the publication of this new replication norm and/or official endorsement by professional societies (e.g., Association for Psychological Science) may further reduce misperceptions regarding negative experiences surrounding replications.

### **Concerns and Challenges**

At first glance, it may seem impossible or unfeasible for researchers to adopt the new replication norm for expensive and/or time-consuming studies (e.g., fMRI or longitudinal studies). Upon closer scrutiny, however, this may simply not be the case. For such studies, researchers can simply include independent replications as *systematic replications* where both a direct and conceptual replication are built into the study design (Lykken, 1968). Though achieving this may increase a study's sample size or number of observations (in the case of between- and within-subjects designs, respectively), it should nonetheless be possible in most situations, especially if such considerations are taken into account ahead of time when planning research. For example, when writing grant proposals, a researcher could decide ahead of time to include a systematic replication for some of the proposed studies. This will ensure that sufficient funds are requested that takes into account the larger sample sizes required to have high-power to (1) replicate the original finding and (2) potentially discover new boundary conditions or underlying mechanisms.

Another concern is that the new replication norm may be difficult or impossible to enforce. Though this is strictly true, I contend that proposing and promoting the *idea* of a new replication norm has a lot value even if such a norm is not enforceable. Indeed, the extant peer reviewer norm is also not strictly enforceable, but we nonetheless all collectively benefit from the existence of such unenforceable informal norm. In a similar sense, there is immense value in promoting a new research culture whereby p-hacking (Simmons et al., 2011) is no longer condoned even though such normative behavior also cannot be enforced. As previously mentioned, the new replication norm idea could benefit our field in several respects even if only a minority of researchers adopts it using a higher original-to-replication studies ratio (e.g., 8:1 or even 10:1).

## Conclusion

In the current article, I propose the idea of a new replication norm whereby researchers should aim to independently replicate important findings in their own research areas in



proportion to the number of original studies they themselves publish per year. As a starting point, I have proposed that researchers should strive for a 4:1 original-to-replication studies ratio based on logical and empirical considerations. However, such ratio can be calibrated depending on a researcher's faculty position and resources. Though the norm may not be enforceable (just like the peer reviewer norm), I contend our field could benefit in several respects even if only a minority of researchers adopts it using an original-to-replication studies ratio higher than the 4:1 ratio suggested. I argue this simple new norm could significantly advance our field by increasing the reliability and cumulativeness of our empirical knowledge base, accelerating our theoretical understanding of psychological phenomena, increasing the quality (rather than quantity) of studies, and by helping to facilitate our transformation toward a research culture where publishing independent direct replications is seen as a completely ordinary part of the research process.

## References

- Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A, Fiedler, K, ..., Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, 27, 108-119.
- Bakker, M., van Dijk, A, & Wicherts, J. M., (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, 7, 543-554.
- Bargh, J. A. (March 5, 2012a). Nothing in their heads. Debunking the Doyen et al. claims regarding the elderly-priming study. Retrieved from <http://web.archive.org/web/20120309203900/http://psychologytoday.com/blog/the-natural-unconscious/201203/nothing-in-their-heads> March 13, 2015.
- Bargh, J. A. (May 11, 2012b). Priming effects replicate just fine, thanks. Retrieved from <https://www.psychologytoday.com/blog/the-natural-unconscious/201205/priming-effects-replicate-just-fine-thanks> March 13, 2015.
- Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, 483(7391), 531-533.
- Brandt, M. J. (2013). Do the disadvantaged legitimize the social system? A large-scale test of the status–legitimacy hypothesis. *Journal of Personality and Social Psychology*, 104(5), 765.
- Campbell, L., Loving, T. J., & LeBel, E. P. (2014). Enhancing transparency of the research process to increase accuracy of findings: A guide for relationship researchers. *Personal Relationships*, 21(4), 531-545.
- Carson, K. P., Becker, J. S., & Henderson, J. A. (1998). Is utility really futile? A failure to replicate and an extension. *Journal of Applied Psychology*, 83(1), 84.
- Chambers, C. D. (2013). Registered Reports: A new publishing initiative at Cortex. *Cortex*, 49, 609–610.
- Cohen, J. (1994). The earth is round ( $p < .05$ ). *American Psychologist*, 49, 997-1003.

- Cumming, G. (2013). The new statistics why and how. *Psychological science*, 25, 7-29.
- Donnellan, M. B., Lucas, R. E., & Cesario, J. (2014). On the association between loneliness and bathing habits: Nine replications of bargh and shalev (2012) study 1.
- Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: it's all in the mind, but whose mind?. *PloS one*, 7(1), e29081.
- Earp, B. D., Everett, J. A., Madva, E. N., & Hamlin, J. K. (2014). Out, damned spot: can the "Macbeth Effect" be replicated?. *Basic and Applied Social Psychology*, 36(1), 91-98.
- Eich, E. (2014). Business not as usual. *Psychological Science*, 25(1), 3-6.
- Feynman, R. P. (1974). Cargo cult science. *Engineering & Science*, 37, 10-13.
- Fisher, R. (1926). The Arrangement of Field Experiments, *The Journal of the Ministry of Agriculture*, 1926, 33, 504.
- Forsyth, D. R. (2009). Group dynamics: New York: Wadsworth. [Chapter 10].
- Frank, M. C., & Saxe, R. (2012). Teaching replication. *Perspectives on Psychological Science*, 7(6), 600-604.
- Fuchs, H. M., Jenny, M., & Fiedler, S. (2012). Psychologists are open to change, yet wary of rules. *Perspectives on Psychological Science*, 7(6), 639-642.
- Funder, D. (Oct 31, 2012). The perilous plight of the (non)- replication. Retrieved from <https://funderstorms.wordpress.com/2012/10/31/the-perilous-plight-of-the-non-replicator/> March 13, 2015.
- Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype threat effects on girls' mathematics performance. *Developmental psychology*, 49(10), 1886.
- Gilbert, D. (May 28, 2014). Some thoughts on shameless bullies. Retrieved from <http://www.wjh.harvard.edu/~dtg/Bullies.pdf> March 13, 2015.
- Harris, C. R. (2011). Menstrual cycle and facial preferences reconsidered. *Sex Roles*, 64(9-10), 669-681.

- Harris, C. R., Coburn, N., Rohrer, D., & Pashler, H. (2013). Two failures to replicate high-performance-goal priming effects. *PloS one*, 8(8), e72467.
- Hendrick, C. (1991). Replication, strict replications, and conceptual replications: Are they important? In J. W. Neuliep (Ed.), *Replication research in the social sciences* (pp. 41–49). Newbury Park: Sage.
- Hone, L. S., & McCullough, M. E. (2014). Does religious cognition really down-regulate hand grip endurance in men? A failure to replicate. *Evolution and Human Behavior*.
- IJzerman, H., Blanken, I., Brandt, M. J., Oerlemans, J. M., Van den Hoogenhof, M. M., Franken, S. J., & Oerlemans, M. W. (2014). Sex differences in distress from infidelity in early adulthood and in later life: A replication and meta-analysis of Shackelford et al.(2004). *Social Psychology*, 45(3), 202.
- Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives in Psychological Science*, 7, 645–654.
- Janz, N. (2015). Bringing the Gold Standard into the Classroom: Replication in University Teaching. *International Studies Perspectives*.
- Jensen, L. E. (2014). Stereotype Spillover Effects on Women in Mathematics. *2014 NCUR*.
- Johnson, D. J., Cheung, F., & Donnellan, M. B. (2014). Does cleanliness influence moral judgments? A direct replication of Schnall, Benton, and Harvey (2008). *Social Psychology*, 45(3), 209.
- Koole, S. L., & Lakens, D. (2012). Rewarding replications: A sure and simple way to improve psychological science. *Perspectives on Psychological Science*, 7, 608-614.
- Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of personality and social psychology*, 37(6), 822.
- LeBel, E. (Dec 9, 2013). New “Reviewer Statement” Initiative Aims to (Further) Improve Community Norms Toward Disclosure. Retrieved from

- <http://osc.centerforopencscience.org/2013/12/09/reviewer-statement-initiative/>, March 13, 2015.
- LeBel, E. P. (July 11, 2014). New replication policy at flagship social psychology journal will not be effective. Retrieved from <https://proveyourselfwrong.wordpress.com/tag/replication-policy/> March 13, 2015.
- LeBel, E. P., & Campbell, L. (2013). Heightened Sensitivity to Temperature Cues in Individuals With High Anxious Attachment Real or Elusive Phenomenon?. *Psychological science*, 24, 2128-2120.
- LeBel, E. P., & Peters, K. R. (2011). Fearing the future of empirical psychology: Bem's (2011) evidence of psi as a case study of deficiencies in modal research practice. *Review of General Psychology*, 15, 371-379.
- LeBel, E. P., & Wilbur, C. J. (2014). Big secrets do not necessarily cause hills to appear steeper. *Psychonomic Bulletin & Review*, 21, 696-700.
- LeBel, E. P., Borsboom, D., Giner-Sorolla, R., Hasselman, F., Peters, K. R., Ratliff, K. A., & Smith, C. T. (2013). PsychDisclosure.org: Grassroot support for reforming reporting standards in psychology. *Perspectives on Psychological Science*. 8, 424-432.
- Lenggenhager, B., Hilti, L., Palla, A., Macaudo, G., & Brugger, P. (2014). Vestibular stimulation does not diminish the desire for amputation. *Cortex*, 54, 210-212.
- Lewkowicz, D. J., & Minar, N. J. (2014). Infants Are Not Sensitive to Synesthetic Cross-Modality Correspondences A Comment on Walker et al.(2010). *Psychological science*, 25(3), 832-834.
- Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, 70, 151-159.
- Lynott, D., Corker, K. S., Wortman, J., Connell, L., Donnellan, M. B., Lucas, R. E., & O'Brien, K. (2014). Replication of "Experiencing physical warmth promotes interpersonal warmth" by Williams and Bargh (2008). *Social Psychology*, 45(3), 216.

- Madurski, C., & LeBel, E. P. (2014). Making sense of the noise: Replication difficulties of Correll's (2008) modulation of 1/f noise in a racial bias task. *Psychonomic bulletin & review*, 1-7.
- Makel, M. C., Plucker, J. A., & Hagerty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, 7, 537-542.
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in Psychology Research How Often Do They Really Occur?. *Perspectives on Psychological Science*, 7(6), 537-542.
- Maslany, A. J., & Campbell, J. I. (2013). Failures to replicate hyper-retrieval-induced forgetting in arithmetic memory. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 67(1), 72.
- Matthews, W. J. (2012). How much do incidental values affect the judgment of time?. *Psychological science*, 23(11), 1432-1434.
- Matzke, D., Nieuwenhuis, S., van Rijn, H., Slagter, H. A, van der Molen, M. W, & Wagenmakers, E.-J. (in press). The effect of horizontal eye movements on free recall: A preregistered adversarial collaboration. *Journal of Experimental Psychology: General*.
- McCarthy, R. J. (2014). Close replication attempts of the heat priming-hostile perception effect. *Journal of Experimental Social Psychology*, 54, 165-169.
- Mitchell, J. (July 8, 2014). On the emptiness of failed replications. Retrieved from [http://wjh.harvard.edu/~jmitchel/writing/failed\\_science.htm](http://wjh.harvard.edu/~jmitchel/writing/failed_science.htm) March 13, 2015.
- Neuliep, J. W., & Crandall, R. (1990). Editorial bias against replication research. *Journal of Social Behavior and Personality*, 5, 85–90.
- Nieuwenstein, M., & van Rijn, H. (2012). The unconscious thought advantage: Further replication failures from a search for confirmatory evidence. *Judgment and Decision Making*, 7(6), 779-798.
- Nosek, B. A., & Lakens, D. (2014). Registered reports. *Social Psychology*, 45(3), 137-141.

- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7, 615-631.
- Open Science Collaboration, T. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, 7, 657-660.
- Open Science Collaboration. (2013). The Reproducibility Project: A model of large-scale collaboration for empirical research on reproducibility. *Implementing reproducible computational research (a volume in the r series)*. New York, NY: Taylor & Francis.
- Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, 7(6), 531-536.
- Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7, 528-530.
- Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of social distance? Failure to replicate effects on social and food judgments. *PloS one*, 7(8), e42510.
- Pashler, H., Rohrer, D., & Harris, C. R. (2013). Can the goal of honesty be primed?. *Journal of Experimental Social Psychology*, 49(6), 959-964.
- PERKINS, H.W., MEILMAN, P.W., LEICHLITER, J.S., CASHIN, J.S. AND PRESLEY, C.A. (1999). Misperceptions of the norms for the frequency of alcohol and other drug use on college campuses. *J. Amer. Coll. Hlth*, 47, 253-258.
- Popper, K. R. (1959). *The logic of scientific discovery*. Oxford, UK: Basic Books.
- Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets?. *Nature reviews Drug discovery*, 10(9), 712-712.

- Roney, J. R., & Simmons, Z. L. (2012). Men smelling women: Null effects of exposure to ovulatory sweat on men's testosterone. *Evolutionary Psychology*, 10(4), 703-713.
- Rosenthal, R. (1997). Some issues in the replication of social science research. *Labour Economics*, 4(2), 121-123.
- Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research*, 37(3), 409-425.
- Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, 17, 551-566.
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(2), 90.
- Schmiedeberg, C., & Schröder, J. (2014). Does Weather Really Influence the Measurement of Life Satisfaction?. *Social indicators research*, 117(2), 387-399.
- Schnall, S. (May 23, 2014). Simone Schnall on her experience with a registered replication project. Retrieved from <http://www.spsblog.org/simone-schnall-on-her-experience-with-a-registered-replication-project/> March 13, 2015.
- Schnall, S., Benton, J., & Harvey, S. (2008). With a clean conscience cleanliness reduces the severity of moral judgments. *Psychological science*, 19(12), 1219-1222.
- 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, 1359-1366.
- Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. "A 21 word solution." *Available at SSRN 2160588* (2012).
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, 9(1), 76-80.



- Simonsohn, U. (2013). Small telescopes: Detectability and the evaluation of replication results. *Available at SSRN 2259879*.
- Smith, J. K., Gerber, A. S., & Orlich, A. (2003). Self-Prophecy Effects and Voter Turnout: An Experimental Replication. *Political Psychology*, 24(3), 593-604.
- Srivastava, S. (September 27, 2012). A Pottery Barn rule for scientific journals. Retrieved from <https://hardsci.wordpress.com/2012/09/27/a-pottery-barn-rule-for-scientific-journals/> March 13, 2015.
- Steele, K. M. (2014). Failure to replicate the Mehta and Zhu (2009) color-priming effect on anagram solution times. *Psychonomic bulletin & review*, 21(3), 771-776.
- Steele, K. M., Bass, K. E., & Crook, M. D. (1999). The mystery of the Mozart effect: Failure to replicate. *Psychological Science*, 10(4), 366-369.
- Steele, K. M., Brown, J. D., & Stoecker, J. A. (1999). Failure to confirm the Rauscher and Shaw description of recovery of the Mozart effect. *Perceptual and Motor Skills*, 88(3), 843-848.
- Tal, I. R., Akers, K. G., & Hodge, G. K. (2008). Effect of paper color and question order on exam performance. *Teaching of Psychology*, 35(1), 26-28.
- Van Wolferen, J., Inbar, Y., & Zeelenberg, M. (2013). Magical thinking in predictions of negative events: Evidence for tempting fate but not for a protection effect. *Judgment and Decision Making*, 8(1), 45-54.
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J. & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7, 627-633
- Wicherts, J. (Jan 2, 2015). Will be 2015 be the year in which most Psychological Science articles have open data? Tweet retrieved from <https://twitter.com/jeltewicherts/status/551039830650785792> March 13, 2015.
- Wicherts, J. M. (2013). Science revolves around the data. *Journal of Open Psychology Data*, 1(1), 1-4. DOI: <http://dx.doi.org/10.5334/jopd.e1>

- Wolfe, J. M. (2013). Registered reports and replications in Attention, Perception, & Psychophysics. *Attention, Perception, & Psychophysics*, 75, 781–783.
- Yong, E. (March 10, 2012). A failed replication draws a scathing personal attack from a psychology professor. Retrieved from <http://blogs.discovermagazine.com/notrocketscience/2012/03/10/failed-replication-bargh-psychology-study-doyen/#.VQMV0454r3s> March 13, 2015.
- Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PloS one*, 7(12), e51382.

### Author's Note

I would like to thank Daniel Lakens, Hal Pashler, Brent Roberts, Rogier Kievit, and Brian Earp for valuable feedback on an earlier version of this manuscript.