


- [Community Perspectives](#)
- [Science in Society](#)
- [In the News](#)
- [In the Journals](#)
- [SPSP 2014](#)
- [About](#)
 - [Twitter Primer](#)
 - [Blogging Around Town](#)

Select Page:

Select Category:

 [Character and Context](#)
[Character and Context](#)

A blog of the Society for Personality and Social Psychology

Simone Schnall on her Experience with a Registered Replication Project

[Simone Schnall](#) May 23, 2014 [Simone Schnall on her Experience with a Registered Replication Project](#) 2014-05-25T08:22:13+00:00 [In the Journals](#) [53 Comments](#)

This post originally appeared [here](#) and is re-posted in its entirety below. Brent Donnellan, one of the authors of the attempted replication, has written a related post [here](#). For more insights into the new special replication issue of *Social Psychology*, see our weekly link round-up [here](#).

 [PUB_SP_08-01.indd](#)

Recently I was invited to be part of a “registered replication” project of my work. It was an interesting experience, which in part was described in an article in [Science](#). My commentary describing specific concerns about the replication paper is available [here](#).

Some people have asked me for further details. Here are my answers to specific questions.

Question 1: “Are you against replications?”

I am a firm believer in replication efforts and was flattered that my paper ([Schnall, Benton & Harvey, 2008](#)) was considered important enough to be included in the special issue on [“Replications of Important Findings in Social Psychology.”](#) I therefore gladly cooperated with the registered replication project on every possible level: First, following their request I promptly shared all my experimental materials with David Johnson, Felix Cheung and Brent Donnellan and gave detailed instructions on the experimental protocol. Second, I reviewed the replication proposal when requested to do so by special issue editor Daniel Lakens. Third, when the replication authors requested my SPSS files I sent them the following day. Fourth, I offered the replication authors to analyze their data within two week’s time when they told me about the failure to replicate my findings. This offer was declined because the manuscript had already been submitted. Fifth, when I discovered the ceiling effect in the replication data I shared this concern with the special issue editors, and offered to help the replication authors correct the paper before it goes into print. This offer was rejected, as was my request for a published commentary describing the ceiling effect.

I was told that the ceiling effect does not change the conclusion of the paper, namely that it was a failure to replicate my original findings. The special issue editors Lakens and Nosek suggested that if I had concerns about the replication, I should write a blog; there was no need to inform the journal’s readers about my additional analyses. Fortunately Editor-in-Chief Unkelbach overruled this decision and granted published commentaries to all original authors whose work was included for replication in the special issue.

Of course replications are much needed and as a field we need to make sure that our findings are reliable. But we need to keep in mind that there are human beings involved, which is what Danny Kahneman's [commentary](#) emphasizes. Authors of the original work should be allowed to participate in the process of having their work replicated. For the Replication Special Issue this did not happen: Authors were asked to review the replication proposal (and this was called "pre-data peer review"), but were not allowed to review the full manuscripts with findings and conclusions. Further, there was no plan for published commentaries; they were only implemented after I appealed to the Editor-in-Chief.

Various errors in several of the replications (e.g., in the ["Many Labs" paper](#)) became only apparent once original authors were allowed to give feedback. Errors were uncovered even for successfully replicated findings. But since the findings from "Many Labs" were already heavily publicized several months before the paper went into print, the reputational damage for some people behind the findings already started well before they had any chance to review the findings. "Many Labs" covered studies on 15 different topics, but there was no independent peer review of the findings by experts in those topics.

For all the papers in the special issue the replication authors were allowed the "last word" in the form of a rejoinder to the commentaries; these rejoinders were also not peer-reviewed. Some errors identified by original authors were not appropriately addressed so they remain part of the published record.

Question 2: "Have the findings from Schnall, Benton & Harvey (2008) been replicated?"

We reported two experiments in this paper, showing that a sense of cleanliness leads to less severe moral judgments. Two direct replications of Study 1 have been conducted by Kimberly Daubman and showed the effect (described on Psych File Drawer: [Replication 1](#), [Replication 2](#)). These are two completely independent replications that were successfully carried out at Bucknell University. Further, my collaborator Oliver Genschow and I conducted several studies that also replicated the original findings. Oliver ran the studies in Switzerland and in Germany, whereas the original work had been done in the United Kingdom, so it was good to see that the results replicated in other countries. These data are so far unpublished.

Importantly, all these studies were direct replications, not conceptual replications, and they provided support for the original effect. Altogether there are seven successful demonstrations of the effect, using identical methods: Two in our original paper, two by Kimberly Daubman and another three by Oliver Genschow and myself. I am not aware of any unsuccessful studies, either by myself or others, apart from the replications reported by Johnson, Cheung and Donnellan.

In addition, there are now many related studies involving conceptual replications. Cleanliness and cleansing behaviors, such as hand washing, have been shown to influence a variety of psychological outcomes (Cramwinckel, De Cremer & van Dijke, 2012; Cramwinckel, Van Dijk, Scheepers & Van den Bos, 2013; Florack, Kleber, Busch, & Stöhr, 2014; Gollwitzer & Melzer, 2012; Jones & Fitness, 2008; Kaspar, 2013; Lee & Schwarz, 2010a; Lee & Schwarz, 2010b; Reuven, Liberman & Dar, 2013; Ritter & Preston, 2011; Xie, Yu, Zhou & Sedikides, 2013; Xu, Zwick & Schwarz, 2012; Zhong & Liljenquist, 2006; Zhong, Strejcek, Sivanathan, 2010).

Question 3: "What do you think about "pre-data peer review?"

There are clear professional guidelines regarding scientific publication practices and in particular, about the need to ensure the accuracy of the published record by impartial peer review. The [Committee on Publications Ethics](#) (COPE) specifies the following [Principles of Transparency and Best Practice in Scholarly Publishing](#): "Peer review is defined as obtaining advice on individual manuscripts from reviewers expert in the field who are not part of the journal's editorial staff." Peer review of only methods but not full manuscripts violates internationally acknowledged publishing ethics. In other words, there is no such thing as "pre-data peer review."

Editors need to seek the guidance of reviewers; they cannot act as reviewers themselves. Indeed, "One of the most important responsibilities of editors is organising and using peer review fairly and wisely." ([COPE, 2014, p. 8](#)). The editorial process implemented for the Replication Special Issue went against all known publication conventions that have been developed to ensure impartial publication decisions. Such a breach of publication ethics is ironic considering the stated goals of transparency in science and [increasing the credibility of published results](#). Peer review is not censorship; it concerns quality control. Without it, errors will enter the published record, as has indeed happened for several replications in the special issue.

Question 4: “You confirmed that the replication method was identical to your original studies. Why do you now say there is a problem?”

I indeed shared all my experimental materials with the replication authors, and also reviewed the replication proposal. But human nature is complex, and identical materials will not necessarily have the identical effect on all people in all contexts. My research involves moral judgments, for example, judging whether it is wrong to cook and eat a dog after it died of natural causes. It turned out that for some reason in the replication samples many more people found this to be extremely wrong than in the original studies. This was not anticipated and became apparent only once the data had been collected. There are many ways for any study to go wrong, and one has to carefully examine whether a given method was appropriate in a given context. This can only be done after all the data have been analyzed and interpreted.

My original paper went through rigorous peer-review. For the replication special issue all replication authors were deprived of this mechanism of quality control: There was no peer-review of the manuscript, not by authors of the original work, nor by anybody else. Only the editors evaluated the papers, but this cannot be considered peer-review ([COPE, 2014](#)). To make any meaningful scientific contribution the quality standards for replications need to be at least as high as for the original findings. Competent evaluation by experts is absolutely essential, and is especially important if replication authors have no prior expertise with a given research topic.

Question 5: “Why did participants in the replication samples give much more extreme moral judgments than in the original studies?”

Based on the literature on moral judgment, one possibility is that participants in the Michigan samples were on average more politically conservative than the participants in the original studies conducted in the UK. But given that conservatism was not assessed in the registered replication studies, it is difficult to say. Regardless of the reason, however, analyses have to take into account the distributions of a given sample, rather than assuming that they will always be identical to the original sample.

Question 6: “What is a ceiling effect?”

A ceiling effect means that responses on a scale are truncated toward the top end of the scale. For example, if the scale had a range from 1-7, but most people selected “7”, this suggests that they might have given a higher response (e.g., “8” or “9”) had the scale allowed them to do so. Importantly, a ceiling effect compromises the ability to detect the hypothesized influence of an experimental manipulation. Simply put: With a ceiling effect it will look like the manipulation has no effect, when in reality it was unable to test for such an effects in the first place. When a ceiling effect is present no conclusions can be drawn regarding possible group differences.

Research materials have to be designed such that they adequately capture response tendencies in a given population, usually via pilot testing ([Hessling, Traxel, & Schmidt, 2004](#)). Because direct replications use the materials developed for one specific testing situation, it can easily happen that materials that were appropriate in one context will not be appropriate in another. A ceiling effect is only one example of this general problem.

Question 7: “Can you get rid of the ceiling effect in the replication data by throwing away all extreme scores and then analysing whether there was an effect of the manipulation?”

There is no way to fix a ceiling effect. It is a methodological rather than statistical problem, indicating that survey items did not capture the full range of participants’ responses. Throwing away all extreme responses in the replication data of Johnson, Cheung & Donnellan (2014) would mean getting rid of 38.50% of data in Study 1, and 44.00% of data in Study 2. This is a problem even if sample sizes remain reasonable: The extreme responses are integral to the data sets; they are not outliers or otherwise unusual observations. Indeed, for 10 out of the 12 replication items the modal (=most frequent) response in participants was the top score of the scale. Throwing away extreme scores gets rid of the portion of the data where the effect of the manipulation would have been expected. Not only would removing almost half of your sample be considered “p-hacking,” most importantly, it does not solve the problem of the ceiling effect.

Question 8: “Why is the ceiling effect in the replication data such a big problem?”

Let me try to illustrate the ceiling effect in simple terms: Imagine two people are speaking into a microphone and you can clearly understand and distinguish their voices. Now you crank up the volume to the maximum. All you hear is this high-pitched sound (“eeeeee”) and you can no longer tell whether the two people are saying the same thing or something different. Thus, in the presence of such a ceiling effect it would seem that both speakers were saying the same thing, namely “eeeeee”.

The same thing applies to the ceiling effect in the replication studies. Once a majority of the participants are giving extreme scores, all differences between two conditions are abolished. Thus, a ceiling effect means that all predicted differences will be wiped out: It will look like there is no difference between the two people (or the two experimental conditions).

Further, there is no way to just remove the “eeeeee” from the sound in order to figure out what was being said at that time; that information is lost forever and hence no conclusions can be drawn from data that suffer from a ceiling effect. In other words, you can’t fix a ceiling effect once it’s present.

Question 9: “Why do we need peer-review when data files are anyway made available online?”

It took me considerable time and effort to discover the ceiling effect in the replication data because it required working with the item-level data, rather than the average scores across all moral dilemmas. Even somebody familiar with a research area will have to spend quite some time trying to understand everything that was done in a specific study, what the variables mean, etc. I doubt many people will go through the trouble of running all these analyses and indeed it’s not feasible to do so for all papers that are published.

That is the assumption behind peer-review: You trust that somebody with the relevant expertise has scrutinized a paper regarding its results and conclusions, so you don’t have to. If instead only data files are made available online there is no guarantee that anybody will ever fully evaluate the findings. This puts specific findings, and therefore specific researchers, under indefinite suspicion. This was a problem for the Replication Special Issue, and is also a problem for the [Reproducibility Project](#), where findings are simply uploaded without having gone through any expert review.

Question 10: “What has been your experience with replication attempts?”

My work has been targeted for multiple replication attempts; by now I have received so many such requests that I stopped counting. Further, data detectives have demanded the raw data of some of my studies, as they have done with other researchers in the area of embodied cognition because somehow this research area has been declared “suspect.” I stand by my methods and my findings and have nothing to hide and have always promptly complied with such requests. Unfortunately, there has been little reciprocation on the part of those who voiced the suspicions; replicators have not allowed me input on their data, nor have data detectives exonerated my analyses when they turned out to be accurate.

I invite the data detectives to publicly state that my findings lived up to their scrutiny, and more generally, share all their findings of secondary data analyses. Otherwise only errors get reported and highly publicized, when in fact the majority of research is solid and unproblematic.

With replicators alike, this has been a one-way street, not a dialogue, which is hard to reconcile with the alleged desire for improved research practices. So far I have seen no actual interest in the phenomena under investigation, as if the goal was to as quickly as possible declare the verdict of “failure to replicate.” None of the replicators gave me any opportunity to evaluate the data before claims of “failed” replications were made public. Replications could tell us a lot about boundary conditions of certain effects, which would drive science forward, but this needs to be a collaborative process.

The most stressful aspect has not been to learn about the “failed” replication by Johnson, Cheung & Donnellan, but to have had no opportunity to make myself heard. There has been the constant implication that anything I could possibly say must be biased and wrong because it involves my own work. I feel like a criminal suspect who has no right to a defense and there is no way to win: The accusations that come with a “failed” replication can do great damage to my reputation, but if I challenge the findings I come across as a “sore loser.”

Question 11: “But... shouldn’t we be more concerned about science rather than worrying about individual researchers’ reputations?”

Just like everybody else, I want science to make progress, and to use the best possible methods. We need to be rigorous about all relevant evidence, whether it concerns an original finding, or a “replication” finding. If we expect original findings to undergo tremendous scrutiny before they are published, we should expect no less of replication findings.

Further, careers and funding decisions are based on reputations. The implicit accusations that currently come with failure to replicate an existing finding can do tremendous damage to somebody’s reputation, especially if accompanied by mocking and bullying on social media. So the burden of proof needs to be high before claims about replication evidence can be made. As anybody who’s ever carried out a psychology study will know, there are many reasons why a study can go wrong. It is irresponsible to declare replication results without properly assessing the quality of the replication methods, analyses and conclusions.

Let’s also remember that there are human beings involved on both sides of the process. It is uncollegial, for example, to tweet and blog about people’s research in a mocking tone, accuse them of questionable research practices, or worse. Such behavior amounts to bullying, and needs to stop.

Question 12: “Do you think replication attempts selectively target specific research areas?”

Some of the findings obtained within my subject area of embodied cognition may seem surprising to an outsider, but they are theoretically grounded and there is a highly consistent body of evidence. I have worked in this area for almost 20 years and am confident that my results are robust. But somehow recently all findings related to priming and embodied cognition have been declared “suspicious.” As a result I have even considered changing my research focus to “safer” topics that are not constantly targeted by replicators and data detectives. I sense that others working in this area have suffered similarly, and may be rethinking their research priorities, which would result in a significant loss to scientific investigation in this important area of research.

Two main criteria appear to determine whether a finding is targeted for replication: Is a finding surprising, and can a replication be done with little resources, i.e., is a replication feasible. Whether a finding surprises you or not depends on familiarity with the literature —the lower your expertise, the higher your surprise. This makes counterintuitive findings a prime target, even when they are consistent with a large body of other findings. Thus, there is a disproportionate focus on areas with surprising findings for which replications can be conducted with limited resources. It also means that the burden of responding to “surprised” colleagues is very unevenly distributed and researchers like myself are targeted again and again, which will do little to advance the field as a whole. Such practices inhibit creative research on phenomena that may be counterintuitive because it motivates researchers to play it safe to stay well below the radar of the inquisitors.

Overall it’s a base rate problem: If replication studies are cherry-picked simply due to feasibility considerations, then a very biased picture will emerge, especially if failed replications are highly publicized. Certain research areas will suffer from the stigma of “replication failure” while other research areas will remain completely untouched. A truly scientific approach would be to randomly sample from the entire field, and conduct replications, rather than focus on the same topics (and therefore the same researchers) again and again.

Question 13: “So far, what has been the personal impact of the replication project on you?”

The “failed” replication of my work was widely announced to many colleagues in a group email and on Twitter already in December, before I had the chance to fully review the findings. So the defamation of my work already started several months before the paper even went into print. I doubt anybody would have widely shared the news had the replication been considered “successful.” At that time the replication authors also put up a blog entitled “Go Big or Go Home,” in which they declare their studies an “epic fail.” Considering that I helped them in every possible way, even offered to analyze the data for them, this was disappointing.

At a recent interview for a big grant I was asked to what extent my work was related to Diederik Stapel’s and therefore unreliable; I did not get the grant. The logic is remarkable: if researcher X faked his data, then the

phenomena addressed by that whole field are called into question.

Further, following a recent submission of a manuscript to a top journal a reviewer raised the issue about a “failed” replication of my work and therefore called into question the validity of the methods used in that manuscript.

The constant suspicions create endless concerns and second thoughts that impair creativity and exploration. My graduate students are worried about publishing their work out of fear that data detectives might come after them and try to find something wrong in their work. Doing research now involves anticipating a potential ethics or even criminal investigation.

Over the last six months I have spent a tremendous amount of time and effort on attempting to ensure that the printed record regarding the replication of my work is correct. I made it my top priority to try to avert the accusations and defamations that currently accompany claims of failed replications. Fighting on this front has meant that I had very little time and mental energy left to engage in activities that would actually make a positive contribution to science, such as writing papers, applying for grants, or doing all the other things that would help me get a promotion.

Question 14: “Are you afraid that being critical of replications will have negative consequences?”

I have already been accused of “slinging mud”, “hunting for artifacts”, “soft fraud” and it is possible that further abuse and bullying will follow, as has happened to colleagues who responded to failed replications by pointing out errors of the replicators. The comments posted below the [Science](#) article suggest that even before my commentary was published people were quick to judge. In the absence of any evidence, many people immediately jumped to the conclusion that something must be wrong with my work, and that I’m probably a fraudster. I encourage the critics to read my [commentary](#) so you can arrive at a judgment based on the facts.

I will continue to defend the integrity of my work, and my reputation, as anyone would, but I fear that this will put me and my students even more directly into the cross hairs of replicators and data detectives. I know of many colleagues who are similarly afraid of becoming the next target and are therefore hesitant to speak out, especially those who are junior in their careers or untenured. There now is a recognized culture of “replication bullying:” Say anything critical about replication efforts and your work will be publicly defamed in emails, blogs and on social media, and people will demand your research materials and data, as if they are on a mission to show that you must be hiding something.

I have taken on a risk by publicly speaking out about replication. I hope it will encourage others to do the same, namely to stand up against practices that do little to advance science, but can do great damage to people’s reputations. Danny Kahneman, who so far has been a big supporter of replication efforts, has now voiced concerns about a lack of “[replication etiquette](#),” where replicators make no attempts to work with authors of the original work. He says that “this behavior should be prohibited, not only because it is uncollegial but because it is bad science. A good-faith effort to consult with the original author should be viewed as essential to a valid replication.”

Let’s make replication efforts about scientific contributions, not about selectively targeting specific people, mockery, bullying and personal attacks. We could make some real progress by working together.

A note on comments: All first-time commenters’ postings must be approved — I approve all comments as soon as I can get to them, regardless of content, as long as they are civil and on topic. If you are interested in writing a full length blog post, I’m happy to consider posting it; email me at [davenussbaum at gmail](mailto:davenussbaum@gmail.com), or tweet me ([@davenuss79](https://twitter.com/davenuss79))

Share this:

- 
- 
- [Pocket](#)
- 

- [Email](#)
- [More](#)
-

Like this:

Like Loading...

[replication](#)

Related Posts

 [Featured](#)

[Psychology News Round-Up \(May 23rd\)](#)

 [Featured](#)

[A Recipe for Replications](#)

 [Community Perspectives](#)

[The Need for Power in Psychology](#)

53 Comments Already

[Subscribe to comments feed](#)

1. [Nicholas A. Christakis](#) - [May 24th, 2014 at 10:18 am](#) none Comment author #16845 on [Simone Schnall on her Experience with a Registered Replication Project by Character and Context](#)

To me, these are the most telling features of this event, as Dr. Schnall notes:
 “Unfortunately, there has been little reciprocation on the part of those who voiced the suspicions; replicators have not allowed me input on their data, nor have data detectives exonerated my analyses when they turned out to be accurate. I invite the data detectives to publicly state that my findings lived up to their scrutiny, and more generally, share all their findings of secondary data analyses. Otherwise only errors get reported and highly publicized, when in fact the majority of research is solid and unproblematic. With replicators alike, this has been a one-way street, not a dialogue, which is hard to reconcile with the alleged desire for improved research practices. So far I have seen no actual interest in the phenomena under investigation, as if the goal was to as quickly as possible declare the verdict of ‘failure to replicate.’”
 This suggests that what the scientists doing the replication studies may be after is not the advancement of knowledge, but some other sort of gratification. This is not good for the field. For one thing, it obliges a careful and innovative scientist such as Dr. Schnall to spend what is clearly a huge amount of time on a boring rear guard action, when she could be discovering new things about cognition.

[Reply](#)

o

[Marek Vranka](#) - [May 24th, 2014 at 12:59 pm](#) none Comment author #16896 on [Simone Schnall on her Experience with a Registered Replication Project by Character and Context](#)

A study with strong claims (such as 40 sentences associated with purity will make subjects judge 6 highly heterogeneous acts as less morally wrong), low power (= only 40 participants) and no significant results should not have been published, period. It has no informational value whatsoever –

the only discovered thing is that the effect probably is not in the opposite direction (95%CI for effect size ranges from 0 to 1).

The ceiling effect criticism is very weak – see <http://www.bit.ly/1jeOBJBW> and moreover, in Schnall's study, the only significant results was in a scenario with a even stronger "ceiling effect" – <http://t.co/30mUSenBml>

Authors of the replication did everything for cooperation with the original author – but giving her right to veto publication of results that she does not like is, in my humble opinion, asking a bit too much.

[Reply](#)

2.

ViewFromNowhere - [May 24th, 2014 at 3:10 pm](#) none Comment author #16943 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

I think papers reporting replications should always include a meta-analysis of all published evidence on the phenomenon under consideration.

Nowadays, non-replications tend to be interpreted as proof that the phenomenon does not exist (although the authors often deny that this is what they mean to say). Therefore, it is important to put them into the broader context.

[Reply](#)

o

Marek Vranka - [May 25th, 2014 at 2:32 am](#) none Comment author #17195 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

Asking to do a meta-analysis is once again asking too much – why is it enough to have 40 participants in the original study, but the replication paper have to be so many more times better? (200+ new subjects per study, searching for other replication attempts – in this case all unpublished...)

I don't understand, why should the original finding enjoy such high levels of protection – it's just a one piece of evidence, same as any other.

When you read the replication paper, you will see it is extremely polite, mentioning a possible reason for different result ("One possibility is that there are unknown moderators that account for these apparent discrepancies. Perhaps the most salient difference between the current studies and the original SBH studies is the student population.") and concluding with "The current studies suggest that the effect sizes surrounding the impact of cleanliness on moral judgments are probably smaller than the estimates provided by the original studies. Researchers attempting future work in this area should use fairly large sample sizes to have the power to detect subtle but perhaps important effects (say a d of 0.10 or smaller). It is critical that our work is not considered the last word on the original results in SBH and we hope there are future direct replications of the original results using populations drawn from many different countries. More broadly, we hope that researchers will continue to evaluate the emotional factors that contribute to moral judgments."

I think this is what we can call "putting things into the broader context". How can someone call this bullying is beyond me.

[Reply](#)

■

Michael Cohn - [May 25th, 2014 at 5:17 am](#) none Comment author #17271 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

The original finding deserves "protection" if one sees a failed replication as an attack against the author's competence, honesty, or intelligence. In our current climate, reports of failed replications are often perceived that way. This is partly a problem with the field, which has ignored and ghettoized direct replication for so long that it's easy to see it as strange and vindictive behavior rather than routine science. Perhaps another reason is that the people who are sounding the alarm about the dangers of ignoring replication are in many cases also the

ones sounding the alarm about questionable research and publication practices. This may make it seem like they're out to identify QRPs when in fact they're just trying to improve the strength or precision of our inferences about a particular finding.

The two aims aren't completely separate. In the Reproducibility Project: Psychology, the way we've framed it is that any one replication might fail for a variety of reasons, and shouldn't be taken as anything except for additional evidence with which to update our beliefs about the phenomenon. However, a low replication rate across many studies does provide evidence of poor practices in the field as a whole. I personally see this lack of specificity as an advantage — it means that we can provide evidence that systemic reform is needed, without making accusations about any individual author's practices.

[Reply](#)

■

Marek Vranka - [May 25th, 2014 at 8:00 am](#) none Comment author #17316 on [Simone Schnall on her Experience with a Registered Replication Project by Character and Context](#)

Agreed, failed replication is often mistakenly confused with some kind of accusation, even when it is only an update about the size and generalizability of effect.

[Reply](#)

■

[Nicholas A. Christakis](#) - [May 25th, 2014 at 3:09 pm](#) none Comment author #17474 on [Simone Schnall on her Experience with a Registered Replication Project by Character and Context](#)

Yes, that is exactly correct. This is what additional studies do — update the size and generalizability of the effect, and perhaps, as Dan Gilbert has said elsewhere here, clarify the “fragility” of the effect (i.e., clarify what other factors are relevant to the ability to discern an effect, if it is real). (And, incidentally, I also note that the sample size in many psychology studies is unreasonably low, and the practice of continuing to accrue subjects until the desired effect is detected is equally problematic.) But, given this quite correct perspective articulated by MV, we must do three further things which often appear to be neglected by many:

1) Wait until several studies have been done exploring an area before determining that the initial effects were not borne out. This is what is typically done with drug discovery in medicine. Note that, with this so-called “replication effort,” the most recent result is somehow seen as the correct one, but who checks the replicators? Why should their work — on its own — get any special credence?

2) Not presume that the original author had either done anything wrong, nor besmirch the competence of the original author; this means, as a corollary, that the replicators must abide all the same rules as the original authors, including making all their data, etc, available, including to the original author. I think the editors of this special issue could easily have anticipated (and should have anticipated) that ‘failed replications’ are not a trivial matter to the authors, especially since the whole issue was configured as about replication, and that all the authors should have been allowed to respond, in print, in the same issue. This is related to the points Danny Kahneman has made about this fracas elsewhere.

3) Not grant special status to the replicators. In fact, we should also, in my view, think more highly of the scientists who take risk and blaze new areas than those who attempt to explore existing areas. To be clear, I believe that both sorts of work are important in science, and, of course, most of us do both sorts of work. In this regard, I am reminded by Theodore Roosevelt's famous ‘Man in the Arena’ speech, which is well worth reading in this discussion: <http://www.theodore-roosevelt.com/trsorbonnespeech.html>

■

Gregg Collins - [May 30th, 2014 at 1:58 pm](#) none Comment author #20605 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

In your comment above, you say the original study “should not have been published” and “has no informational value whatsoever.” (“Whatsoever”! My, my.) In light of that, perhaps you can explain what you mean when you say “failed replication is often mistakenly confused with some kind of accusation.” Where exactly does the “mistake” part come in?

3.

[Daniel Gilbert](#) - [May 25th, 2014 at 12:22 am](#) none Comment author #17137 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

The replicability of the results is not the issue. With some careful cooperative work it would be easy to find out just how fragile or robust this effect is — and no one has been more cooperative than the author. The issue is the process, which resembles a witch hunt that is entirely in the hands of a bunch of self-righteous, self-appointed sheriffs, and that is clearly not designed to find truth. Simone Schnall is Rosa Parks — a powerless woman who has decided to risk everything to call out the bullies. This expose is damning. The replication police should apologize and explain what they are going to do to make sure this never happens again.

[Reply](#)

o

[JP de Ruiter](#) - [May 25th, 2014 at 8:21 am](#) none Comment author #17327 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

As Andrew Wilson wrote in another comment here, anyone can take someone else’s method section and try to replicate on the basis of that (that’s how real science works, or at least, is supposed to work) but the replication team nevertheless went out of their way to cooperate with the authors. That there are disagreements between them does not mean they were bullying.

Dan, you wrote that the way the people from this replication project treated Simone Schnall is similar to how black people were treated in the days of Rosa Parks. I think that is a very harsh accusation, and I invite you to either take it back, or provide arguments for it. By just insulting the team you are not helping the debate, or for that matter, Simone Schnall.

[Reply](#)

o

[Christian Battista](#) - [May 25th, 2014 at 4:23 pm](#) none Comment author #17500 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

I can’t tell whether you’re being serious here or whether you’re just pranking us all with a level commitment that rivals that of Andy Kaufman. If it’s the second thing, my hat is off to you sir.

[Reply](#)

o

[Timothy Ryan](#) - [May 25th, 2014 at 6:59 pm](#) none Comment author #17546 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

I don’t think this comment is nearly as insulting to the replicators as it is to the experience of African Americans during the civil rights era. I don’t think Mr. Gilbert has much of an appreciation for how brutal those years were and how silly comparing Ms. Schnall’s treatment to them is. Has it occurred to you, Mr. Gilbert, that the variables that Ms. Schnall was looking at (a feeling of cleanliness, moral outrage) are remarkably hard to quantify and that the sample sizes she was working with (80-something students) were quite small. Is it not at all conceivable that she might have gotten her results by chance or that the replicators might have gotten their by chance... is it not at all possible that the effect she found does not exist? It is very possible and I don’t read anything in the replicators’ paper that hints to me that they were acting as judge, jury and executioner. Replication is absolutely necessary. This does not at all call into question the work of Ms. Schnall. Her work is

important and necessary and was undoubtedly conducted well. Everyone assumes that to be true. To not give the replicators the same courtesy is absolutely unjust. You, sir, are being very silly indeed.

[Reply](#)

4.

[Daniel Lakens](#) - [May 25th, 2014 at 1:55 am](#) none Comment author #17179 on *Simone Schnall on her Experience with a Registered Replication Project by Character and Context*

As the editor of the manuscript by Johnson, Chueng, and Donnellan of the work by Simone's Schnall and colleagues, I first of all want to say I find it very unfortunate this experience was not a positive one for all the people involved. Both the original researcher, as the replication team, have been on the receiving end of a lot of opinions and comments that were not as nuanced as you should expect from scientists (myself included). The goal of the special issue was to get more data on important effects for which few direct replications were available in the literature. I think it is important to discuss whether this was done in the best possible way.

Just to clarify some things that might distract from this discussion. There was peer review (by original authors, if alive) of the introduction, method, and analysis plan (basically the results, without numbers). Final manuscripts only differed in that they had a short discussion. As editors, we expected a brief conclusion, stimulated authors to include meta-analyses, and a short interpretation of any surprises. These were reviewed by the editor for consistency with the data and whether the conclusions made sense (in all cases, revisions were made). This seems to have worked in 14/15 papers. In the specific case of Simone Schnall, peer review by experts after the discussion was written would not have changed the final manuscript, and we would have ended up with exactly the same disagreement. The original author would have argued against publication because of a ceiling effect, and I, as the editor, would have disagreed with this interpretation. See this post-publication peer review for one line of reasoning why the ceiling effect explanation was not convincing: <http://pigeer.wordpress.com/2014/05/24/additional-reflections-on-ceiling-effects-in-recent-replication-research/>. I don't think this example is a failure of how our special issue was peer reviewed. It's a disagreement that would have arisen under any circumstance. Unfortunate, but a scientific disagreement will happen. I also disagree that the rejoinder to Simone Schnall's article (which was, together with the commentary, handled by the editor of Social Psychology, Christian Unkelbach, to prevent any possibility of bias) still contains unaddressed errors.

I invite anyone interested to look at the validity of the ceiling effect argument (the data is shared on the Open Science Framework, although due to inexperience with the OSF, Simone Schnall still needs to make her data 'public' – I made a similar mistake the first time I used the OSF). I especially invite researchers to consider whether analyzing *percentages* of extreme scores (instead of just the extremity of the means) is valid, and whether neutral and experimental conditions should be grouped (given that if there's no true effect, experimental trials should be higher (and thus, more 'extreme'), since the hypothesis is that the manipulation reduces ratings on the moral vignettes compared to the control condition). Brent Donnellan has posted additional arguments here: <http://traitstate.wordpress.com/2014/05/21/random-reflections-on-ceiling-effects-and-replication-studies/>.

Simone Schnall refers to 'various errors' in the Many Labs project, but in essence this boils down to a difference in the anchoring paradigm used (see <https://osf.io/vka47/>) which led to a much larger effect than in the original studies. This is not ideal, and should be improved. We did not have 15 reviewers for all individual studies, and although the Many Labs team contacted many authors to check the procedures, one or two 'classroom examples' of studies that always replicate received less attention than they should have (the procedure has been imposed in the second round of the Many Labs project). But this is a criticism that should not weigh too heavy against the replication studies published in the special issue in general.

I do not agree with the suggestion raised in Simone's blog that the conclusions in the article are wrong due to a lack of peer review. However, I do think we should reconsider whether we want to always give original researchers the possibility to review and comment on replication studies. Looking at the literature, you see a mixed approach: Sometimes replications are posted as they are, sometimes authors are given the ability to reply. If we exclude studies that replicated researchers who are no longer alive in our Special Issue, I think that if one in ten researchers has a negative experience with a replication procedure, we should work collaboratively as a field to create procedures that make negative experiences with replication research as rare as possible. People (such as the journalist who wrote the Science article) should stop combining examples of fraud and examples of replications that did not reveal a statistical difference in the same paragraph – not every finding can be expected to replicate (especially in small studies, see Ioannidis, 2005), and all authors leave open the possibility there are unknown moderators at work and stress more research is needed, also in the case of the link between morality and cleanliness. At the same time, we

should not let this single negative experience distract from the fact that our field placed itself on the forefront of self-organized improvements to the way we do research by publishing the first journal issue in the history of science consisting exclusively of pre-registered replication studies.

[Reply](#)

5.

[Bill von Hippel](#) - [May 25th, 2014 at 4:19 am](#) none Comment author #17249 on *Simone Schnall on her Experience with a Registered Replication Project by Character and Context*

Simone's experience with the replication project is very disheartening. I'm a big believer in the replication project and in providing data when people request them, but the responsibilities go both ways. People who ask for someone else's data should report when their investigations support the original author, perhaps that would prevent numerous redundant requests. And the original author should be allowed to offer a review of replication attempts after the data are collected. Reviewers don't have veto power, they simply provide advice that an editor can use or ignore.

[Reply](#)

o

[Marek Vranka](#) - [May 25th, 2014 at 7:50 am](#) none Comment author #17308 on *Simone Schnall on her Experience with a Registered Replication Project by Character and Context*

"People who ask for someone else's data should report when their investigations support the original author, perhaps that would prevent numerous redundant requests." – I couldn't agree more. I think this is the only valid criticism offered by Schnall. However, it doesn't apply here – there were no unreported supportive findings.

Moreover, other papers from the special issue of Soc Psy report supportive findings (e.g. Many Labs).

But it could be a more general problem and we should do everything in order to prevent it – behaving according to guideline "publish everything" will probably be enough.

[Reply](#)

6.

[Andrew D Wilson](#) - [May 25th, 2014 at 6:20 am](#) none Comment author #17288 on *Simone Schnall on her Experience with a Registered Replication Project by Character and Context*

Authors of the original work should be allowed to participate in the process of having their work replicated. This caught my eye. Actually, it's not true. To replicate your work I don't need you or your permission; I (should) just need your Methods section. It's nice that the recent organised pushes for replications have tried to involve all parties, but once your work is out there it is fair game for replication and criticism. That's just the deal.

[Reply](#)

o

[Dave Nussbaum](#) - [May 26th, 2014 at 4:03 pm](#) none Comment author #18140 on *Simone Schnall on her Experience with a Registered Replication Project by Character and Context*

Andrew wrote a lengthier post on this topic [here](#), specifically Kahneman's [commentary](#) in the Special Issue, to which I responded in the comments [here](#). We ended up agreeing on the broader points but emphasize different ways of improving the state of the field. My response follows:

On one level, I agree with your points, but I think you're far too quick to dismiss Kahneman as nonsense. We could use more humility in this debate altogether — I think a lot of the vitriol comes from people's failure to really listen to the other side and give them benefit of the doubt — as they usually deserve — before dismissing them.

Science has very strict ideals. Scientists are humans. We aspire to uphold these ideals, but we often come woefully short of them, even when we're not cutting corners or trying to gain some advantage at the expense of the truths we seek. As humans, scientists have real limitations and if we fail to take those into account then we may be making strong arguments about the ideals of science, but pragmatically we may be making little progress.

When Kahneman talks about the myth of perfect science he's acknowledging that things don't always go according to ideals and that we should be mindful of that. Perhaps researcher's whose work is not replicated should not feel threatened — some aren't, but should we really be surprised that some are? Sometimes formally, usually informally, their ability or integrity may be called into question, and their life prospects may be affected. If you're ready to run those people out of the field, then I think you'll find yourself pretty lonely.

So an "etiquette" that doesn't demand, but suggests that the original authors are consulted seems like a reasonable way to move things in the right direction. Does science demand it? No, absolutely not. Can it help improve science right now? I think it probably can.

What's more, if scientists write bad methods sections, then this approach has a good chance to change that too. It would certainly be embarrassing to have to send a 12-page addendum to researchers to tell them the additional steps needed to properly replicate your study.

By making the field more receptive to replications, and increasing their number and scope, the underlying elements are likely to be improved as well.

So what Kahneman says may not be something that is required for science to be done, but I certainly don't agree that it's nonsense.

In case the HTML doesn't work: <http://psychsciencenotes.blogspot.com/2014/05/psychologys-real-replication-problem.html>

[Reply](#)

7. *Anonymous - [May 25th, 2014 at 6:34 am](#) none Comment author #17292 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context*

I posted this comment anonymously because I fear retribution from replication bullies and data detectives.

I am happy to see the field discuss replication and methodology issues. It will help us produce better science in the long run. However, I am disturbed by the behavior of some individuals. It is likely that most people involved in replications are reasonable and seek to better understand the effects. Unfortunately, some individuals seem motivated by the desire to disprove an effect, kill a research area, or take down a researcher. I've seen mocking behavior firsthand on social media sites like Facebook and Twitter. It is surprising that some people do not even try to hide their identity when they mock others. Furthermore, it is also disturbing that some individuals seem more interested in failed rather than successful replications. I know of at least one example of an individual who was happy to publish failures to replicate but did not even try to publish a successful replication. These individuals seem to think that only they are doing "real" psychological science.

We can do better. I believe replications should focus on the science and not on researchers or research areas we do not like. We should replicate studies from a broad range of psychological science and not only studies that are easy to conduct online or fall within a certain domain. Replication practices will be better when we realize that most of us came to this field because we are excited about behavior and seek objective answers as much as possible.

[Reply](#)

8. *Epistemologist - [May 25th, 2014 at 8:11 am](#) none Comment author #17320 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context*

Once upon a time, I have learned that in science, critical issues should be addressed in an academic discourse to which all participants have the chance to equally contribute. This discourse was meant to be guided by specific questions that needed to be answered and specific problems that needed to be solved. Ideally, the outcome of such a discourse were new conceptual or methodological insights that afforded a deeper understanding of the underlying structures and processes. Guilt or innocence was not an issue.

In contrast, the current replication campaigns are not guided by specific questions or problems but by the zest of self-appointed judges to hand down a final verdict if a phenomenon is “real” or not. If the authorized replications fail, the original author is turned into a suspect who has either engaged in Questionable Research Practices, or even in fraud. From then on, s/he is under special observation and a candidate for further investigations.

I am glad that Simone Schnall has taken the courage to expose her experiences. Perhaps this may serve as a starting point to take a more critical look at these Questionable Replication Practices that are praised as a revolution in psychological science.

[Reply](#)

9.

[Michael J. Kane](#) - [May 25th, 2014 at 8:42 am](#) none Comment author #17335 on *Simone Schnall on her Experience with a Registered Replication Project by Character and Context*

Best practices in replication should be for replication authors to reach out to original authors, if only (and perhaps only) to discover any relevant information that may have been left out of Methods or Analyses sections, and to discuss any plausible a priori moderators of the effect that might be of concern or of interest (particularly if further work has been done that the replication authors are not aware of).

Beyond that, though, I have mixed feelings about providing the original authors any additional role in the work, particularly if the original work was underpowered (which, we all know, most psychological research is). As psychologists, we should understand very well the biases and motivated reasoning involved when we evaluate any research, and original authors — as human beings — are likely to be among the most biased evaluators of any replication attempts.

If you work in an active research area with competing theoretical views, as I do, it is regular and expected that other authors will publish work that claims to contradict your published findings or your views (whether via direct replication, conceptual replication, or an entirely new form of relevant study). In some cases, those authors do a good job of representing your findings or views, but in other cases they don't (and, of course, in both cases, one's assessment is likely to be biased to some extent). Although the former work would likely have benefited from formal or informal discussion beforehand, as a field we certainly don't require this, and I don't think that we should for “direct” or “conceptual” replications, either. If the work is important enough to the field and/or to the original author, then they/we should all be motivated to continuing to study and argue about the validity of our data and claims and others' data and claims.

I respectfully disagree with Dr. Schnall that the only valid form of peer review is that which includes Results and Discussion sections. I don't disagree that this form of peer review can work and can be important; I regularly review manuscripts that, in my view, have not analyzed the data entirely properly and/or have not appropriately interpreted their data. At the same time, particularly in research areas with active theoretical debate, reviewers may be biased in their evaluation of the quality of the work depending on how the results turn out. For a compelling and sobering demonstration of this (which is well cited but still probably not well enough known), please see Mahoney (1977; *Cognitive Therapy and Research*, vol 1(2), pp. 161-175). With Mahoney's findings in mind, I do think there's room for pre-data-collection peer review of manuscripts, particularly for work in contentious areas of study.

I do agree with Dr. Schnall that the field should develop norms or best practices regarding the use of other scientists' raw data. It distorts the literature if re-analyses only come to light when they contradict the original authors' claims and not when they support them. Perhaps sharing agreements might be written that affirm (if not guarantee) publication or posting of re-analysis results regardless of how they turn out?

In closing, I don't think there's much we can do (or much that should be done) about different areas of psychology being more or less prone to replication attempts — again by analogy to different areas of study eliciting different levels of theoretical disagreement and corresponding empirical work. I remember finding such disagreements and contention to be stressful when I was a graduate student, and I'll admit to finding it hard sometimes to separate my emotional responses from my intellectual responses in evaluating work that is critical of my own. But, in my view, that's integral to being a human being involved in the scientific enterprise. We should all be teaching our students and post-docs (and reminding ourselves) that we all will often be wrong (and, of course, replicating authors and original authors are equally likely to be wrong), and that's not an indictment of our intellect or ethics; however, as a field we are now being increasingly

reminded that we can also minimize being wrong more often than is necessary by running highly powered studies and replicating them internally before publishing.

I'm not a social psychologist, and I'm not an expert in the substantive domains at issue here, but I think that all psychologists can benefit from discussions like these. I've certainly learned a lot, and I thank Dr. Schnall for initiating such a discussion here.

[Reply](#)

○

Timothy Ryan - [May 25th, 2014 at 7:07 pm](#) none Comment author #17549 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

I think your tone and conciliatory words are very appropriate. That said, I sincerely doubt any of the findings here were due to investigator bias on either the part of Dr. Schnall or the replicators. I am sure that both ran very well conducted studies. (It is a courtesy absolutely necessary to extend to all scientists.) The small sample sizes scream of the variability of detected effects being due to chance. All this begs for is more and larger studies being done to investigate the effect.

[Reply](#)

■

Michael J. Kane - [May 27th, 2014 at 12:23 pm](#) none Comment author #18749 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

Yes, indeed, I agree with you. My points about bias were intended to refer to the ways in which reviewers and other scientists *interpret* studies and their findings. I'm sorry that I didn't make that clearer.

[Reply](#)

10.

Rob Folger - [May 25th, 2014 at 9:12 am](#) none Comment author #17345 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

I'm surprised that no one above has yet mentioned the ideas Danny Kahneman has expressed about grace, politeness, collegiality, and collaborative efforts. I'm doing this from memory, so it's possible the ones I have in mind might not have been published yet. As I recall, he has written about this issue twice: once pointing out the nature of the problem & subsequently speaking more to the issue of etiquette. Perhaps the latter has appeared only online; if so, I hope subsequent posts will cite the source (I'm not in a good position to look it up right now). Agree with Danny or not, it's at least worthwhile to include his thoughts in the discussion. In my opinion, they are relevant when we consider not only distributive justice (our verdict on the presumed correctness of one set of results or another) but also the other elements identified as concerns many people have (evidenced by a vast amount of empirical literature in psychology that I assume is replicable!): interpersonal justice, informational justice, and procedural justice—see especially the work by scholars such as Tom Tyler, E. Allan Lind, and Bob Bies. The interpersonal realm pertains to treating people with respect and dignity, and I think at this point we ought to bend over backwards in our attempts to live up to those ideals. Issues of informational justice apparently were centered on transparency at the outset, which is all to the good. The aftermath, however, has suggested we might not yet have worked out all the necessary details. I think that overlaps with procedural justice (e.g., see criteria suggested by Gerald Leventhal), and those “how to” details might take longer than we realize to achieve something like a decent amount of consensus. Like beauty, after all, fairness is in the eye of the beholder!

I'm identifying myself (Rob Folger) as others such as Dan Gilbert have done. An interesting test of interpersonal justice will be to see how people respond to my comment.

[Reply](#)

11.

Nick Brown - [May 25th, 2014 at 1:45 pm](#) none Comment author #17441 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

Dr Schnall writes:

>There is no way to fix a ceiling effect.
>It is a methodological rather than statistical problem, indicating
>that survey items did not capture the full range of participants'
>responses.

But the replicators had no choice in their use of the survey items that produced these ceiling effects, because they were specified by Schnall et al. in their original article. If these measures show ceiling effects, then the survey is at fault, unless it is specified in advance that it is not suitable for use with, say, Mid-West college students. That's why we test the reliability and validity of our instruments. The quoted statement effectively admits that the methods of Schnall et al. can only detect the effects reported by those authors in certain populations, i.e., those who express their moral judgements in a sufficiently nuanced way that they do not come too close towards the top end of a 0-9 or 1-7 numerical scale. That such a population was so easy to find — in the form of the most readily available population to the first team of researchers to attempt a replication, a population that at first blush would appear to have a substantial degree of similarity to Schnall et al.'s sample — is worrying.

Thus, Dr. Schnall's invocation of the ceiling effect does not provide her with a satisfactory escape from the problems posed by the failed replication of Schnall et al. To invoke the problem of Johnson et al.'s alleged ceiling effect in effect accepts that the original article is flawed by the use of one-item measures of moral judgement that are lacking in validity, or reliability, or both. (It is particularly ironic that, apparently, the population group that is "causing a problem" is the one studied more than any other in social psychology, i.e. American college students; cf Hinrich, Heine, & Noranzayan, 2010.)

[Reply](#)

12. [Political Science Replication](#) - [May 25th, 2014 at 2:06 pm](#) none Comment author #17447 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

I do understand that Simone Schnall took it personally when her study could not be replicated, and when she felt that she was not heard. But in general, I think the involvement of original authors is not something we should attempt when replicating work.

It should actually be possible to reproduce work WITHOUT the help or involvement of an original author. For that purpose, all journals should require authors to upload all data (not just ask authors to send them on request), full descriptions of the data and procedures etc. in an online supplement. Ideally, anyone should be able to reproduce the work independently without ever contacting the author.

In addition, it is not clear to me why an original author should have the right to analyze the replication and review the manuscript before publication. Once the original paper (and data) is published and enters the scientific dialogue, it is not the 'property' of the original author, but belongs to all. In the past, authors have answered to replications of their work, which I call a replication chain – but after publication.

The publicity around the project means that those authors whose work was put on the spot might suffer unnecessary negative attention. It also means that a discussion takes place, which is important.

I discuss what the case of Schnall means for the replication debate on my blog: "Replication Bullying": Who replicates the replicators? <http://wp.me/p315fp-tm>

[Reply](#)

13. [Brian Nosek](#) - [May 25th, 2014 at 2:30 pm](#) none Comment author #17457 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

Colleagues –

With permission from Daniel Lakens and Simone Schnall, I have compiled and made public all of the email correspondence regarding the replication by Johnson, Cheung, and Donnellan and her response. You can find it here: <http://bit.ly/1jNGiIM>

Regards,
Brian

[Reply](#)

o

JP de Ruiter - [May 26th, 2014 at 2:47 am](#) none Comment author #17730 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

I congratulate the editors and Simone Schnall with their decision to publish this email correspondence in the interest of transparency. What I personally see here unfolding is a very civilized disagreement about the interpretation of data and the right of the author to publish a final response (which is, in the end, granted). While I am sympathetic towards and worried about the negative consequences that were reported by Simone Schnall to have occurred on the basis of the failed replication, I do not believe that the editors were to blame for these unfortunate consequences.

[Reply](#)

14. *Jeff Bowers - [May 25th, 2014 at 3:36 pm](#) none Comment author #17487 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context*

I agree with the replication agenda, but it seems like the procedure could be greatly improved in one simple way. Insure that the lead author of any replicated paper is invited to review the submission.

I'm surprised that this is in any way contentious. When I write a paper that challenges the empirical or theoretical work of someone I expect the paper to be reviewed by that person. And when I'm acting as an action editor I send out any such paper to the person who is being challenged. After all, no one is more expert in the topic. The job of the editor is to consider the arguments of all the reviewers, not simply to look for consensus. If the person challenged recommends rejection that is no surprise. But the important issue is not his or her recommendation, but the reasons for his or her conclusion.

I would be annoyed if someone attempted to replicate my work and I only found out about it after the paper had been accepted. Who wouldn't? I think it is great that replications are on the agenda, and replications are being published. But let the published expert in the topic contribute his or her review. This should be standard procedure.

Jeff Bowers

[Reply](#)

o

Chris Chambers - [May 25th, 2014 at 4:08 pm](#) none Comment author #17493 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

"I would be annoyed if someone attempted to replicate my work and I only found out about it after the paper had been accepted. Who wouldn't?"

I wouldn't. I do not feel entitled to inspect the work of those attempting to reproduce my studies. The findings that my research produces are not my personal property – they belong to the community. I would instead be glad that anyone felt my work to be sufficiently important to attempt replicating it in the first place.

I also see it as my responsibility to provide all the details and materials that independent scientists need to attempt such replications. It is not their responsibility to chase me down for all the important stuff I forgot (or was too lazy) to mention in my Method section.

This is how physical sciences operate and, in my opinion, is how psychology should work as well. Anything less pollutes our enterprise with ego, presumption, entitlement, and ultimately a betrayal of the dispassionate principles that lead to true discovery.

[Reply](#)

■

Jeff Bowers - [May 25th, 2014 at 4:59 pm](#) none Comment author #17511 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

An action editor should look for the expert reviewers, and no one would be more expert than the author of the original paper. So that alone should lead the original author to be invited to review. But it is also a reasonable way to treat colleagues. No one is saying the original author owns the data, and they have no veto power in publishing a solid replication.

I'm not sure where you get the idea that the "physical sciences" operate in a different way. If a researcher/theorist in physics/chemistry, etc. has a paper that is being criticized, I would expect that he/she would have an opportunity to contribute in the review process. If not, why not? In purely scientific terms, the top expert should be consulted.

Jeff

[Reply](#)

■

JP de Ruiter - [May 25th, 2014 at 6:19 pm](#) none Comment author #17531 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

Jeff, you write that "An action editor should look for the expert reviewers, and no one would be more expert than the author of the original paper." I strongly disagree with that statement.

First of all, there is the obvious problem that original authors are naturally and understandably heavily biased against studies that fail to replicate their own work. This is already the case with studies that just happen to contradict other work. I know this from some studies that happen to falsify certain famous theories. Many journal editors, for reasons that elude me completely, assigned the authors of the very theory that the authors were busy falsifying as one of the reviewers. Not really surprisingly, these reviewers always found something devastatingly wrong with the study, and recommended a straight rejection without the possibility of resubmission. I suspect that assigning as a reviewer the original author that is just been non-replicated will have a similar result. In fact, I actually think it is even unethical to bring an original author in the position of reviewing failed replications of their own work.

But more importantly, you suggests that the relevant expertise that we are looking for here is primarily the ability to "get the effect" rather than the general scientific ability to do sound research. The idea that you need to be an expert in certain effects to find them is potentially dangerous for our accumulation of reliable knowledge. Rolf Zwaan addresses this issue in his blog here: <http://rolfzwaan.blogspot.co.uk/2014/01/why-social-behavioral-primers-might.html>

[Reply](#)

■

Chris Chambers - [May 25th, 2014 at 6:52 pm](#) none Comment author #17542 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

I agree with JP – expertise is one thing, but AEs must of course also be cognisant of bias and motivated reasoning.

I would add that one reason psychology is in such trouble is that we are, as a community, too thin-skinned. We are too concerned about the egos and reputations of those around us and not concerned enough about the credibility of the work we are doing.

In my view, the reasonable way to treat colleagues is not to necessarily consult with them in the process of conducting a replication but to ensure that when we publish original research we embrace maximum transparency in our published methods.

If I do this then it allows my colleagues to (a) trust that my results are in principle replicable and not the consequence of QRPs (conscious or unconscious); and (b) follow my experimental recipe without ever needing to communicate with me. If they want to talk to me that's fine and I should be happy to assist them. But I never have the right to feel aggrieved if they publish a replication without my knowledge.

And if their replication fails because of some undisclosed but critical detail in my experiment that I didn't mention in my original paper? That's on me, not them.

[Reply](#)

■

Jeff Bowers - [May 26th, 2014 at 3:27 am](#) none Comment author #17763 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

Chris Chambers, you write:

"And if their replication fails because of some undisclosed but critical detail in my experiment that I didn't mention in my original paper? That's on me, not them."

Is that the best way to advance knowledge? It would have been better if the replicators had found out about the undisclosed details no? This could have informed the replication. Or at least have been included in the discussion once the information was brought to light.

JP, you are concerned that the original authors will recommend rejection without possibility of resubmission. But they are not the editors. If they make good points, then the editor may reject. Otherwise not.

Indeed, a failed replication will have more force if the authors of the original paper were given the opportunity to review.

I just don't see any reason not to invite the original authors to review. It can only make a reviews more informed, and it might reduced hard feelings (speaking for myself, I would feel better if that was the process if someone was replicating me).

15.

[Fred Hasselman](#) - [May 25th, 2014 at 10:19 pm](#) none Comment author #17606 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

Excuse me.

This is really an inappropriate argument:

A ceiling effect means that responses on a scale are truncated toward the top end of the scale. For example, if the scale had a range from 1-7, but most people selected "7", this suggests that they might have given a higher response (e.g., "8" or "9") had the scale allowed them to do so.

What is described here is the case of a ceiling effect in which I would use a measurement stick of 1.5 meters to measure how tall people are and subsequently realise that my measurement instrument was unable to detect any lengths above 1.5 meters.

The scale was: 0 (perfectly OK) to 9 (extremely wrong).

A ceiling effect as described above that would be caused by measurement resolution can simply not occur on a rating scale. (Unless the categories are ambiguous, or instruction is flawed.)

By saying that people would have wanted to respond with "8" or "9", but couldn't, one claims that there exists an absolute measurement scale for moral judgements (such as length measured in centimetres) onto which all the participants somehow project their opinion. I am quite certain that if the max scale value would have been 99, the participants would NOT have wanted to answer "8" or "9".

The only objection one could make, is that an ordinal scale variable should not be considered as observed at an interval scale. A nonparametric rank order test should have been used.

More importantly than the methodological details:

Even if it were true that this a ceiling effect is the identified cause of the non-replication and not the falsity of the theory itself, why isn't everybody celebrating that science has advanced?

If still unconvinced, it will be very easy to settle this. The replication authors take another sample and add 8 and 9 to the top of their rating scale and everything should pan out as expected...

[Reply](#)

o

samD - [May 27th, 2014 at 4:43 am](#) none Comment author #18557 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

Memories of Spinal Tap: <http://www.youtube.com/watch?v=4xgx4k83zzc>

[Reply](#)

o

Barry Cohen - [May 30th, 2014 at 4:02 pm](#) none Comment author #20656 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

"A ceiling effect as described above that would be caused by measurement resolution can simply not occur on a rating scale."

It is true that regardless of the number of points on the rating scale the top number is going to be something like "extremely wrong," so it does not seem likely that increasing the number of points would help. But the inherent limitations of this type of scale can be a problem whenever the top category is often used. For example, if you asked people how wrong it was for someone to murder another human being because s/he had stolen something from him/her, many would use the top category of the "wrongness" scale, no matter how it was anchored. Then, if you asked the same person, or a different person, how wrong it would be to murder a person and his/her entire extended family, because that person had insulted you, they would surely use the top category again. However, if given a chance to compare the "wrongness" of the two acts, the participant might be inclined to rate the latter case a bit more wrong. In a between-groups design, in which the majority of participants in two different conditions are using the top category, you wouldn't have the ability to detect differences in extreme wrongness judgments that might exist. I don't know enough about the data in the study under discussion to judge whether this argument applies to it, but given that a more general point was being made about rating scales and ceiling effects, I thought it was worth making this point.

[Reply](#)

16.

Simon Farrell - [May 26th, 2014 at 12:31 am](#) none Comment author #17656 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

I agree with the many measured comments above that replications are good for science, and should not be taken as a direct attack on someone's work or reputation. However, the *should* in the previous sentence is rather contentious. In an ideal world, failures to replicate would not be taken as negatively reflecting on the original research, but we don't live in an ideal world. If it is the case that a failure to replicate is taken by peers and institution to reflect poorly on the original author(s) (and comments above suggest that at least some people share this assumption), then it is understandable that those original authors are going to feel like their reputation has suffered.

Another issue is that the choice of studies to attempt to replicate is not random, and so has some directed element to it. While [some](#) have proposed randomly selecting studies to replicate, as far as I am aware most (all?) published replication attempts to date have involved some element of deliberate choice on the part of the replicators (the Reproducibility Project being an obvious example). Whilst having your study replicated might feel good to some ("Yay, my study was considered important enough to replicate!"), it can also be taken to send a signal that those results are suspect for some reason. The intention of replicators doesn't always align with the signals that are sent by the act of running a replication study.

[Reply](#)

o

Epistemologist - [May 26th, 2014 at 8:46 am](#) none Comment author #17952 on *Simone Schnall on her Experience with a Registered Replication Project by Character and Context*

Good point. The ambiguity why a given study was selected gives rise to all kinds of speculations. Thus, *ceterum censeo*, why not make a replication contingent on its potential to answer a specific question, be it methodological or conceptual? This question should be made explicit at the outset and the results should be evaluated from this perspective.

[Reply](#)

■

question - [May 26th, 2014 at 10:16 am](#) none Comment author #17992 on *Simone Schnall on her Experience with a Registered Replication Project by Character and Context*

I wonder if there is an underlying fear of replications in this whole discussion. Perhaps because they (will?) show that a large percentage of (some?) psychological research can not be replicated (due to whatever reasons)?

I am amazed by the discrepancy between mentioning that replication is important, but countless nuances on how they should be conducted, and the fact that they have hardly ever been conducted in the past 100 years or so.

Could it be that the non-replication of (some?) psychological studies is the big elephant in the room that nobody wants to directly address? Are we all fooling ourselves, and have been fooling ourselves for decades?

[Reply](#)

■

ViewFromNowhere - [May 26th, 2014 at 11:34 am](#) none Comment author #18022 on *Simone Schnall on her Experience with a Registered Replication Project by Character and Context*

There's a pattern. Critical comments regarding recent replication practices are often countered by questioning the scientific integrity of those who dare to criticize ("the lady doth protest too much") or of the field as a whole, as just witnessed. And this is part of the problem that is being discussed here.

[Reply](#)

17.

David Johnson - [May 26th, 2014 at 3:57 pm](#) none Comment author #18137 on *Simone Schnall on her Experience with a Registered Replication Project by Character and Context*

Many commenters have expressed concerns about our motivations for choosing to replicate Schnall, Benton, and Harvey (2008). I would like to clear the air with why we chose to replicate this study. Many of these reasons are detailed in our actual manuscript (<http://bit.ly/1mtQv7f>) but I will concisely list them here.

1) The the study was subjectively interesting to the first and second authors (Johnson & Cheung). It was also highly cited (over 150 times since Decemeber 2013), which was a requirement for the special issue. We then approached Dr. Donnellan and asked if he was willing to supervise the project.

2) There were opposing predictions from other researchers (i.e., Zhong, Strejcek, Sivanathan, 2010) that proposed that self-cleanliness would be linked to **harsher** moral judgment, rather than **less harsh** moral judgment as predicted by Schnall and colleagues. These competing predictions allow for tests of strong inference (Platt, 1964), which are desirable as they help avoid problems of confirmation bias.

3) Because this project was published in *Psychological Science* in 2008, it would also be eligible to be included into the Replicability Project even if it was not accepted for the special issue. We followed through with this, and our replication has since been included into the Replicability Project (<https://osf.io/apidb/>).

4) There was a large degree of imprecision in the estimates provided by Schnall and colleagues, due to their small sample sizes (Exp. 1 $d = -0.60$, 95% CI [-1.23, 0.04], $N = 40$; Exp. 2 $d = -0.85$, 95% CI [-1.47,

-0.22], N = 43). Notice in particular that the effect size estimate for Experiment 1 includes 0, making it all the more important to provide information about an effect that could range from 0 to greater than 1.

Perhaps just as important to mention is that one of our motivations was **not** to demonstrate that this particular effect does not exist. Given the method of strong inference, we would have been pleased to find evidence in either direction (i.e., cleanliness is linked to harsher/less harsh moral judgments). In addition, we also explicitly mention in our manuscript that:

“Our work simply provides more information about an interesting idea. The current studies suggest that the effect sizes surrounding the impact of cleanliness on moral judgments are probably smaller than the estimates provided by the original studies...It is critical that our work is not considered the last word on the original results in SBH and we hope there are future direct replications of the original results using populations drawn from many different countries.”

So, we do not argue that there is no effect of cleanliness on moral judgments. Rather, we conclude that the effect sizes are probably overestimates, and that more work to be done to ferret out possible boundary conditions of this effect.

[Reply](#)

18. [Harry Reis - May 26th, 2014 at 5:35 pm](#) none Comment author #18178 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

Bravo, Simone, for having the courage to speak out. It is all too easy to accuse when a replication fails — “gotcha” has been popular in politics for years and now it has come to social psychology. There are many reasons why replications fail and many of them have nothing to do with the integrity of the original researcher or her/his work. Instead of delighting about the failed replication, and announcing it to the world, a little humility might be nice.

[Reply](#)

19. [Epistemologist - May 26th, 2014 at 11:56 pm](#) none Comment author #18388 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

A replicates b), C does not replicate d), E does not replicate f), G replicates h), etc., etc.,

What have we learned?

[Reply](#)

20. [Anon 2 - May 29th, 2014 at 11:08 pm](#) none Comment author #20226 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

There are many comments here that deserve comment. One, it is certainly true that a non-trivial portion (minority? majority?) of scholars attempting to replicate findings are bitter, unhappy, jealous individuals who seem hell-bent on not replicating. When that happens (or if; but with these folks, it does seem to be when), thinly-disguised accusation of unethical, immoral behavior on the part of the original authors emerge. That damages the morale of the field, not to mention what it does to the original authors and their colleagues/students. On that: PHD students are leaving our field in droves and who would blame them? Deans are opting to fund areas other than social, and granting agencies are moving their monies elsewhere. We want good science – absolutely, no question – but the manner and tone in which replications often are done (definitely not always but often) is quite damaging. I have heard PHD students say (from different students): if they cannot replicate an effect, everything in the field is fake; if they get a predicted result that is between $p = .04$ -.049, they will not publish it for fear of being charged with p-hacking (even with ample sample size)..... This is getting to be madness. Colleagues, wake up and realize that work that was published under prior regimes might not show the same effect size as what the original authors showed (or even be significant) but that is a far cry from those authors being immoral. It's unfair and unwise to judge research and papers developed under the prior standards by today's standards. The key is to be forward-looking and to get our colleagues to adopt new standards. Together we can make our field better and assuage the damage that has been done. Divided, though, we'll be in dire straights for a long time.

[Reply](#)

21. anon 3 - [May 30th, 2014 at 9:43 am](#) none Comment author #20494 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

Anon 2 made several excellent points, although the characterization of a non-trivial portion of replicators as 'jealous & bitter' is NOT helpful and needs to stop!!!

The main excellent point of Anon 2's that I want to reiterate is to halt the circular firing squad this has become and to focus instead on applying the new standards.

The changes that have happened are a huge benefit to us. I am so satisfied now to BELIEVE the results that are reported with large sample sizes and full disclosure of null effects, p-curves, etc. But as emphasized by anon 2, for how long do we need to keep looking back? Do we need exact replication of Festinger & Carlsmith?! (They had a very small sample size.)

Can we not have more pre-registered reports of NEW RESEARCH?! As researchers, let's focus on building a solid empirical foundation (and stop focusing on exposing the p-hackers of the past who admittedly deserve it in some cases but NOT if doing so requires burning down our house).

To those at the forefront of the reproducibility movement, show us how it is done: go ahead and carry out some important new work that is of the highest methodological standards (I am right here with you trying as well to do the same thing...)

To journal reviewers and editors, let's get away from demanding unrealistically novel conceptual advances in each paper and focus more on publishing solid incremental contributions (I am also trying to do this in each of these roles).

[Reply](#)

22. anon 3 - [May 31st, 2014 at 8:05 am](#) none Comment author #21050 on Simone Schnall on her Experience with a Registered Replication Project by Character and Context

One more point I wanted to add to anon 2: Maybe it is OK that grad students do not want to publish results with p values betw .04 and .05 before replicating. In my view the apparent imbalance in replicability of results from JPSP and JEP: LMC as reported by the reproducibility project reflect a confound: probably the p values from the articles in JEP: LMC are lower than from the articles in JPSP. Let's have more self-replication BEFORE publishing (this will lead of course to a norm of publication quantity lower than in the recent past--see, e.g., Betsy Levy Paluck's blog on this). This will firm up our own foundation, so we can then challenge the skeptics to have fun validating our ideas. In this way, we can begin to think of the skeptics as our research assistants.

[Reply](#)

Leave a Reply

Your email address will not be published. Required fields are marked *

Name *

Email *

Website

Comment

You may use these HTML tags and attributes: <abbr title=""> <acronym title=""> <blockquote cite=""> <cite> <code> <del datetime=""> <i> <q cite=""> <strike>

Post Comment

- ☐ Yes, add me to your mailing list.
- ☐ Notify me of follow-up comments by email.
- ☐ Notify me of new posts by email.

« [The Role of Ability Beliefs in Academic Gender Gaps Psychology News Round-Up \(May 23rd\)](#) »

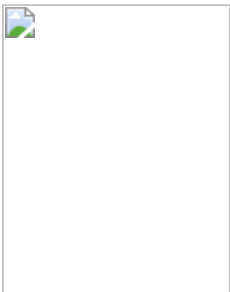
[Facebook](#)



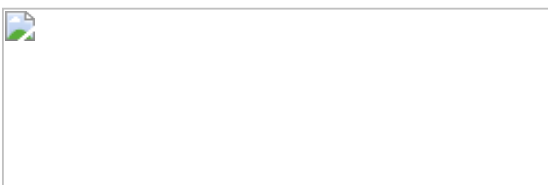
Twitter Primer

 [Getting Started with Twitter](#)

Who we are



Newsletter





Follow us #SPSPblog



Follow @SPSPnews

12.8K followers

Tweets by @SPSPnews



SPSP

@SPSPnews

Catch the whole recap here: wke.lt/w/s/bthV-s
<https://twitter.com/AlexWestPsych/status/1248290734257684486>

#SPSPchat 28: Early Career Challenges During Coronavirus
wakelet.com

13s

[Embed](#)

[View on Twitter](#)

Recent Posts

- [Psychology News Round-Up \(May 30th\)](#)
- [The Mystery of the Hot Hand](#)
- [Overprecision in the Accuracy of Private Belief](#)
- [APS 2014: Coping with Cognitive Conflict](#)
- [Psychology News Round-Up \(May 23rd\)](#)

Follow #SPSP2014 on twitter

[Tweets about "#SPSPblog"](#)

Blogging Around Town

[Notes on Replication from an Un-Tenured Social Psychologist](#)

Last week the special issue on replication at the Journal of Social Psychology arrived to an explosion of debate (read the entire issue here and read original author Simone Schnall's commentary on her experience with the project and Chris Fraley's subsequent examination of ceiling effects). The debate has been happening everywhere--on blogs, [...]

Tue, May 27, 2014

Psych Your Mind

[Does the replication debate have a diversity problem?](#)

Folks who do not have a lot of experiences with systems that don't work well for them find it hard to imagine that a well intentioned system can have ill effects. Not work as advertised for everyone. That is my default because that is my experience. - Bashir, Advancing How Science [...]

Sun, May 25, 2014

Hardest Science

[The Top Ten Worst Reasons to Stay Friends With Your Ex](#)

Your ex is your ex for a reason. But he or she was also an important part of your life for a significant amount of time, and it's understandable to want to hold on to that relationship in some capacity. Many former couples, whether they were dating partners or spouses, try [...]

Sat, May 24, 2014

Psych Your Mind










[Can Personality Change?](#)

Can personality change? In one respect, the answer is clearly "yes," because ample evidence shows that, overall, personality does change. On average, as people get older (after about age 20), they also become less neurotic and more conscientious, agreeable, and open, until about age 60 or so (Soto, John, Gosling [...])

Tue, May 06, 2014

Funderstorms

Top Posts & Pages

-  Simone Schnall on her Experience with a Registered Replication Project
[Simone Schnall on her Experience with a Registered Replication Project](#)
-  Psychology News Round-Up (May 30th)
[Psychology News Round-Up \(May 30th\)](#)
-  The Mystery of the Hot Hand
[The Mystery of the Hot Hand](#)
-  The Role of Ability Beliefs in Academic Gender Gaps
[The Role of Ability Beliefs in Academic Gender Gaps](#)
-  Psychology News Round-Up (May 23rd)
[Psychology News Round-Up \(May 23rd\)](#)
-  Does all morality require a victim?
[Does all morality require a victim?](#)
-  Constructing Self-Esteem Across Cultures
[Constructing Self-Esteem Across Cultures](#)
-  Changing Mindsets to Raise Achievement
[Changing Mindsets to Raise Achievement](#)
-  How is your sleep affecting your relationship?
[How is your sleep affecting your relationship?](#)

Search for:

Subscribe via email

 [Subscribe](#)

Sign up for instant notifications of new posts and our weekly digest.

Email *

Select list(s):

☐ Instant Notifications

☐ Weekly digest

Subscribe via Wordpress

Enter your email address to subscribe to this blog and receive notifications of new posts by email.

Subscribe via RSS



Copyright © 2014. Character and Context

- [SPSP](#)
- [About](#)

%d bloggers like this: