PSYCHOLOGICAL INQUIRY
2021, VOL. 32, NO. 1, 17–23
https://doi.org/10.1080/1047840X.2021.1889317

A Train Wreck by Any Other Name

Christine Harris[a], Doug Rohrer[b], and Harold Pashler[a]

[a]Department of Psychology, University of California San Diego, La Jolla, California, USA; [b]Department of Psychology, University of South Florida, Tampa, Florida, USA

This interesting paper by Sherman and Rivers (in press) seems to have a narrower point (which its authors argue for very explicitly) and a broader point (which they argue for much less explicitly.) The narrower point is that the phrase 'social priming' (and some other alternative labels that have cropped up in this context recently) are exceptionally poor labels for the body of priming research that has been called into question in a torrent of failed replication attempts appearing within the last 10 years or so (sometimes referred to as "the train wreck"). We agree with this narrower point, with a few important qualifications. Sherman and Rivers' broader point is a suggestion, which runs through the paper but is never quite explicitly stated, that the onslaught of non-replications has not really brought to light anything terribly worrisome or misguided about the field and its prevailing research practices. The authors' view seems to be that the field was doing about as well as one should expect of a scientific field, but suffered the misfortune of having a good number of outsiders wander in and stir up trouble in various ways. The biggest problem, they seem to suggest, was basically one of public relations rather than substance (especially, that a Nobel Laureate regrettably used somewhat inflammatory language to describe his frustration with research in the area, an event that got fairly wide coverage).

We are one of the groups who have been viewed by some as marauding outsiders, although one of us is a social psychologist. Beginning around 2010, we tried to replicate some of the most (to us) surprising and fascinating priming results coming out of the social cognition field, as did numerous other labs around the same time (e.g., Doyen, Klein, Pichon, & Cleeremans, 2012; Harris, Coburn, Rohrer, & Pashler, 2013, Klein et al., 2014; Pashler, Rohrer, & Harris, 2013; Shanks et al., 2013). Our direct replication attempts, usually with larger n's than the original studies, resulted in a stream of completely negative results, and left us with a rather different perspective on the situation than Sherman and Rivers. We think that the unreproducibility of such a high fraction of the well-known results on a purported phenomenon (essentially 100%, depending on what is counted) is extremely troubling and should, as it has, engender deep concern about research practices in any field. We see it as reflecting systemic problems that were (and perhaps still are) very widespread in social cognition (and perhaps a far wider swath of research topics), including a lack of recognition of the crucial importance of direct replications, the erroneous belief that conceptual replications offer an adequate substitute for direct replications, and a lack of an ethic of personal responsibility for authors of questioned research to reproduce their own findings and report candidly on their ability or inability to do so whenever possible.

Social Priming: The Label

We start first with the authors' narrower point about terminology. As noted above, we are largely in agreement with Sherman and Rivers (2020) that "social priming" is a strange phrase to use for the studies that have been drawn into question. In most cases these studies involved just a single subject tested alone. In the case of online studies, subjects were tested without engaging in any social interaction with an experimenter or anyone else (e.g., Caruso, Vohs, Baxter, & Waytz, 2013). As Sherman and Rivers suggest, the term 'social priming' may have gained popularity as a sort of lazy short-hand for "the kind of priming research done by social psychologists working in the field of social cognition."

However, as Sherman and Rivers allude to, in fact the term 'social priming' was used frequently by at least one well-known researcher within the field over time (e.g., Bargh, 2005, see p. 49; Bargh, 2006). The phrase was not the invention of outsiders. But as they say, it is not very descriptive.

Sherman and Rivers also criticize the term "behavioral priming," which has also been used from time to time, making the reasonable point that even the classic finding of associative perceptual priming in lexical-decision task involved a behavioral change – subjects are quicker to recognize DOCTOR as a word if it follows NURSE rather than BREAD (Meyer & Schvaneveldt, 1971)). Obviously, all of the priming effects reported in the literature, whatever their authors' home discipline, involve some modulation or change in behavior. If they didn't, there would be no effect to discuss.

Be that as it may, we were surprised that Sherman and Rivers never even mention what seems to us to be the most interesting and substantive difference that separates the priming effects that seem to have fascinated so many people, and which have been disconfirmed on so many occasions,

CONTACT Harold Pashler hpashler@gmail.com 4202 E Fowler Ave, Tampa, FL 33620, USA.

on the one hand, and the widely-established perceptual priming effects that are easily reproduced, on the other. That difference is this. The perceptual priming effects seem to reflect a bias toward perceiving an ambiguous stimulus as (or recalling) an instance of a kind of stimulus that would be more likely to occur given the prime (Johnston & Hale, 1984; Ratcliff & McKoon, 1988). This kind of threshold modulation based on conditional probabilities is likely to be entirely rational from a Bayesian perspective. If doctors and nurses are often encountered together (in texts as well as in real life), and if you just saw a doctor and now you see something that might be a nurse (or a letter string that might be 'nurse'), it is sensible to process new information with priors that favor that interpretation. The bias is a "feature not a bug."

By contrast, the priming results of John Bargh, Kathleen Vohs, and others that drew so much attention, and pretty much all of the ones that Kahneman discussed in his book (Kahneman, 2011) would not seem to represent any such rational self-adjustment in the cognitive apparatus. These purport to show something much more puzzling—that the mind is wired to choose actions and enhance motivational tendencies that are associated with concepts that have been serendipitously activated by recent sensory inputs. Consider a few examples that we tried and failed to replicate. Take "rudeness priming" of Bargh, Chen, and Burrows (1996, Exp. 1) (where the subject is said to interrupt the experimenter more quickly if she has been exposed to words like brazen,' 'infringe' and 'obnoxious'), or honesty priming of Rasinski, Visser, Zagatsky, and Rickett (2005) (where the subject discloses embarrassing information more fully if he was just exposed to words like 'open,' 'sincere' and 'truthful'), and the "money priming" effects of Vohs, Mead, and Goode (2006) (the subject chooses behaviors reflecting greater self-reliance after seeing a screensaver showing a lot of money, perhaps behaving as the subject might imagine an independently wealthy person would do.)

It seems to us that all these findings embody a fascinating but counterintuitive hypothesis about how the mind is constructed. One reason it may be counterintuitive is because it seems like it would be more of a "bug not a feature." As we pointed out (Harris et al., 2013), it would seem to make people prone to manipulation, something that evolution might have been assumed to minimize. As quite a few studies in the past 10 years have shown, ordinary intuitions are often quite a good guide to behavioral-science reality (Dreber et al., 2015; Forsell et al., 2019).

There may be no perfectly suitable label for this substantive category we are discussing, and various authors (ourselves included) seem to have struggled to find a term for it. For example, Bargh referred to "trait construct activation", Harris et al. (2013) referred to "Social/Goal Priming", Rasinski, Visser, Zagatsky, and Rickett (2005) referred to "goal priming", and Kahneman (2011) referred to "ideomotor effects" (p. 53). All of these terms seem to us preferable to 'social priming,' and they denote a meaningful psychological hypothesis.

To us, the more significant question is not what this category should be called but whether the underlying phenomenon is experimentally confirmable.

The Train Wreck: How Bad was it and Why Did it Happen?

The second two-thirds of Sherman and Rivers' paper consists of a number of extended remarks on the repeated replication failures that have afflicted priming research in the field of social cognition. The authors do not quite say "nothing to see here, move along" but the tone seems in line with that.

From our perspective, the failure to reproduce one after another counterintuitive priming effect involving purported changes in motivation or choice of actions (whatever term one uses for this category of studies) was startling and disturbing to watch as it unfolded in our own lab. In our first steps into this topic, we tried to replicate four studies from the Bargh lab and several other studies of "money priming" from the work of Vohs and colleagues. None of the studies produced anything that seemed to us to be remotely confirming the claimed effects despite fairly large sample sizes and strenuous efforts to duplicate the original procedures. We had begun the work with a desire to "see the effects for ourselves", but no strong conviction either way about the validity of the results. Unable to find first one effect and then another, we kept on trying, figuring we'd probably find at least a few results we could confirm. This never happened. Putting all our results together, the meta-analytic synthetic mean effect size was extremely close to zero. Unbeknownst to us, at the very same time we were involved in this research, several other cognitive labs were undertaking extremely similar (sometimes overlapping) investigations, also driven (we were told) more by curiosity than by a conviction that the effects would prove to be unreal (Doyen et al., 2012; Shanks et al., 2013). None of these groups were able to confirm any of the priming results they set out to reproduce. It still remains a mystery to us how Bargh and colleagues replicated their own effect in the elderly priming study and how Vohs et al. (2006) found a money priming effect in 9 out of their 9 studies, while no independent labs could reproduce such effects.

When we discussed our early failures to replicate with Kahneman and others around 2013, we encountered skepticism. We were not, after all, specialists in the field of social cognition. Were our methods subtly different from the original methods? Did we lack some "artistry" that was needed to produce the results? Did we bring some disbelief into the lab that was somehow "jinxing" the effects? We did not feel it likely that we had either deviated from the original protocols in any way or "jinxed" the effect (the research assistants running the studies often did not even know what was being measured, much less what outcome we might possibly have preferred). Nonetheless, in response to these challenges, we switched to emphasize online data collection with computer-delivered instructions, to rule out any possible role for

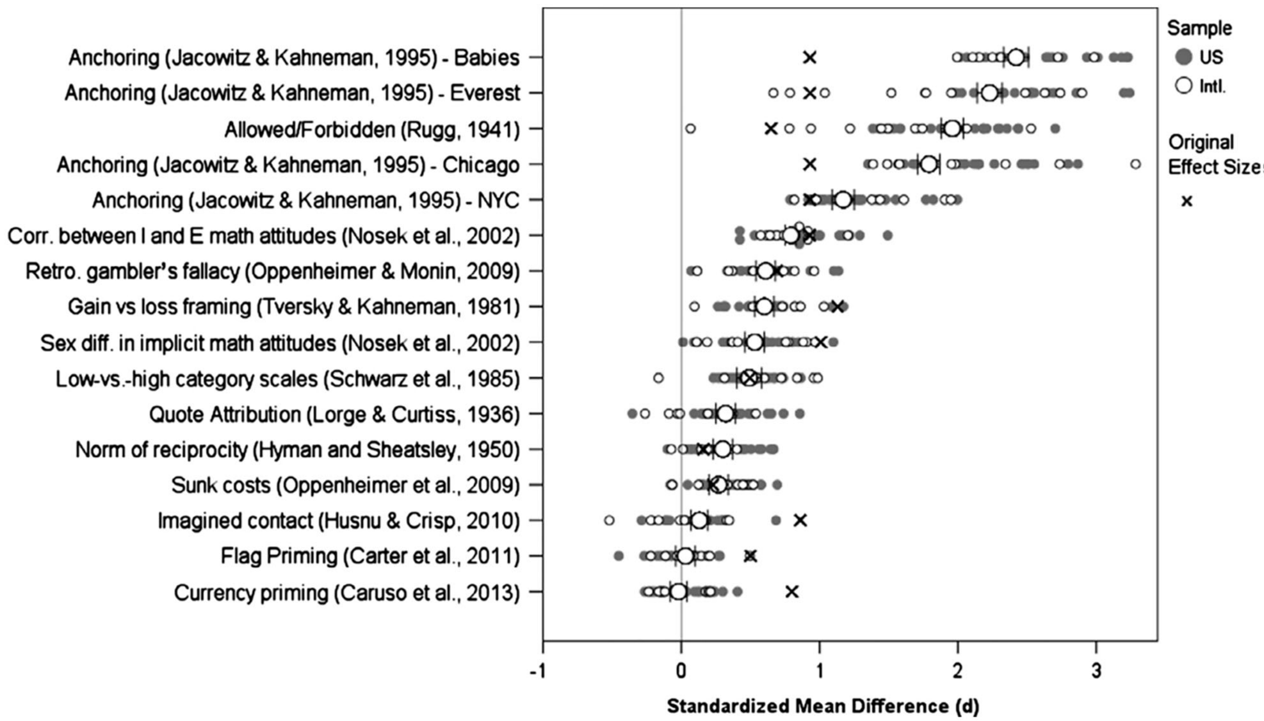

Figure 1. Many Labs Replication. Thirty-six laboratories attempted to replicate each of 13 findings, and the only two findings that showed essentially zero mean effect size were the two "social priming" effects: flag priming and money priming (Klein et al., 2014).

human-interaction moderators. The results continued to be a complete "wipe-out" (Rohrer, Pashler, & Harris, 2015).

Soon our efforts and those of our colleagues in the cognitive field were dwarfed (and complemented) by several new giant multi-lab studies that appeared as the Replication Crisis gathered steam. For example, the so-called Many Labs I study, in which 36 labs all sought to replicate a set of findings, included two social priming effects (Klein et al., 2014). As seen in Figure 1, the results were convincing failures to replicate, with between-lab heterogeneity scarcely exceeding what would be expected from sampling error.

To us, finding that result after result from many different labs would prove irreproducible was startling and disturbing. As we saw it, this train wreck suggested that a fairly large and quite visible scientific enterprise, one that was celebrated in textbooks and articles by science journalists alike, was somehow not operating correctly—with implications potentially going well beyond the social cognition field. There are several specific observations that convinced us of this—things that Sherman and Rivers do not discuss.

First, there was no evidence of any self-correction process taking place within the social cognition field itself. This is shown partly in the fact that it seemed to have taken an influx of outsiders for it to become known that the dramatic priming effects Kahneman had written about did not really work as reported. This lack of self-correction was especially highlighted when a large-scale fakery committed by Diederik Stapel came to light (see Wicherts, 2011). Due to very thorough investigation in Holland, it became clear that Stapel had produced and published many dozens of findings that were apparently faked. To our knowledge, not even a single failure to replicate ever appeared in print or was presented at a conference while the fraud was underway. It seems disappointing that in a field where self-correction had become basically nonexistent, Sherman and Rivers would fail to discuss the problem and outline measures to improve it.

Second, when we began speaking about our own failures to replicate priming effects, what was even more startling to us than the failures themselves was the phlegmatic and seemingly uninterested responses we got from some original investigators. We had assumed, from our experience in the cognitive psychology field, that any lab whose reputation was built on a spectacular result would, upon learning of a failure to replicate, jump into action and re-confirm that they themselves could obtain the results again. And then, we assumed, they would be eager to show would-be replicators what they were doing wrong. Instead, the "social priming" investigators have more often than not indicated that their interests had changed and they had no interest in further pursuit of the topic. In one case, these reactions came from a lab that over a period of years appeared to have received several million dollars in funding from the US National Institutes of Health to pursue research following up on the findings. A failure of original investigators to grapple seriously with replication failure when possible strikes us as an abdication of responsibility—one that makes it harder for the field to figure out whether a result is real or not.

We should mention here a few creditable exceptions, however. Eugene Caruso and Ap Dijksterhuis both participated in attempts to replicate their earlier findings; see Caruso, Shapira and Landy (2017) and O'Donnell et al. (2018). (See also Pashler & De Ruiter, 2017, for some further discussion of personal responsibility and replication failures in empirical science.)

In the remainder of this commentary, we discuss some of the lines of commentary and argument that Sherman and Rivers offer in their paper.

The Studies were Underpowered so it is Unsurprising that they would Fail

To us, one rather amusing suggestion in this paper is that it is Kahneman who should be embarrassed to see so many of the results he publicized in his book turn out to be irreproducible, rather than the field of Social Cognition. According to Sherman and Rivers, Kahneman should have realized that underpowered studies often fail for innocuous statistical reasons. This suggests that Sherman and Rivers view it as just a normal thing that famous literatures covered in every introductory textbook and graduate textbook of a field would be comprised of fascinating results none of which ever work when you try them again.

There are several problems with this. First, the judgment that these studies were underpowered presumes that one knows what effect size such an effect can be expected to have if it is real. Of course, this is never known in advance, which is one of the problems with efforts to reform science through more widespread calculation of a priori power. In fact, the large effect sizes associated with the counterintuitive priming results were, when noticed, interpreted as evidence of large actual effects. Indeed, the size of the effects seems to have been part of the importance that convinced Kahneman to emphasize the results in his book (e.g., he says money-primed people "persevered almost twice as long" and placed a chair "much farther" than unprimed people, p. 55). It matters if reading elderly-related words makes you walk 10% slower (about 1 standard deviation), or a tenth of a percent slowers. If seeing screenshots of money would make you place your chair 50% farther from other subjects than you would otherwise have done, the effect deserves more attention than if it makes you place it a half centimeter further (Vohs et al., 2006). The former would be fascinating, the latter quite a bit less so. Moreover all else equal, large effect sizes are also more likely to have theoretically pertinent causes rather than uninteresting "nuisance" causes, further arguing for the importance of effect size (Vul, Harris, Winkielman, & Pashler, 2009).

A second problem is that some of the most prominent original studies that were not able to be replicated in fact had substantial numbers of subjects, so the generalization proposed by Sherman and Rivers is not completely accurate. For example, Caruso, Vohs, Baxter, and Waytz reported five dramatic money-priming results with sample sizes as large as 275 subjects. Our own attempts to replicate these findings (Rohrer et al., 2015) did not verify any of them (nor did the massive follow-up study by the 36 laboratories participating in the Many Labs replication, Klein et al., 2014). To be clear, we agree with Sherman and Rivers that studies with particularly small samples should be viewed with caution. However, small sample sizes do not characterize all the studies that failed to be reproducible.

Direct Replications are Unnecessary and Conceptual Replications are Better Anyway

Glaringly absent from Sherman and Rivers' paper is anything about the need for direct replication. This absence would seem to reflect the authors' considered opinion, given the arguments presented by Crandall and Sherman (2016) in a paper entitled "On the scientific superiority of conceptual replications." In that paper, Crandall and Sherman maintain that conceptual replications—in which a number of features are "changed up" but the same basic phenomenon is tested—are better than direct replications. They argue that conceptual refutations offer a strong test of a hypothesis, potentially confirming the reality of an effect, but with a test of generalizability thrown in as a bonus.

Before the priming train wreck, we had sympathy for this view, and based on many conversations with colleagues over the years, it is our impression that this has been more or less a consensus view of psychologists who write literature review articles and textbooks. One hears things said like "I look for and take seriously effects that hold across some variety of situations, materials, and subjects", which certainly sounds reasonable enough. However, we have come to believe that, unfortunately, it is deeply misguided as a gauge of the reality of an effect. The problem is that taking conceptual replications as an adequate (or superior) substitute for direct replication *interacts insidiously with publication bias* to allow large literatures to emerge that seem to confirm the existence of completely nonexistent effects (Pashler & Harris, 2012).

Over and over again in the Replication Crisis, one encounters literatures that are resplendent with varied and imaginative conceptual replications–and yet somehow, whenever any particular experiment is tried, it does not work. How does that happen? In our view, the explanation has a sociological aspect as well as a statistical aspect.

When investigators attempt/conduct/run a conceptual replication and it fails, not only are they very unlikely to publish it—they are unlikely to even think that anything is amiss with the original phenomenon. The investigator may well reason, "I shouldn't have changed up so many things, my mistake", instead of wondering whether the *underlying result is nothing but a type 1 error.* For that reason conceptual replications are unable to convince people that effects do not exist.

The opposite is not the case, however. If a conceptual replication works, the change is likely to be convincing to author and reader alike, and also to be highly publishable. For one thing, the paper is almost certain to have at least one highly enthusiastic reviewer: the original investigator. Thus, there is no outcome of a conceptual replication that will lead the field to doubt the original finding.

Consider what happens, then, when you have a type 1 error that has found its way into the literature (hardly a rare event; see Ioannidis, 2005). Suppose the type 1 error pertains to the existence of an intriguing effect (like incidental priming of goals and motivations). Many people will step forward to try their hand at the effect, particularly if the studies seem simple enough to perform (as priming studies usually do). If the culture of the field demands that the first step is a direct replication, most attempts will fail and these failures may often become fairly widely known through informal channels. On the other hand, if the culture of the field encourages skipping the direct replication and going directly to conceptual replications (as Crandall and Sherman advocated), again most will fail but a few will work (perhaps many, if the field also has widespread p-hacking). The difference, however, is that the ones that work will readily find their way into the published literature and the failures will not dent anyone's confidence in the result.

The implication is that all one needs to have a non-effect sweep a field and produce a pseudo-literature of pseudo-confirmation that will be convincing to textbook authors and general readers alike is to have an appealing pseudo-phenomenon and a culture that views conceptual replications as a fine way to test the original hypothesis.

The policy implications recommended by this way of thinking are simple and (we suspect) may even have been fairly widely understood in some fields. Any follow-up on a published phenomenon should begin with a direct replication of that phenomenon. For scientific self-correction to work well, the results of such investigations should be deemed publishable, regardless of the outcome. Moreover, *only* by this sort of direct replication can the field obtain any credible estimate of the underlying effect size (see Wilson, Harris, & Wixted, 2020, for simulations on this point that put aside the convenient but unrealistic assumption that underlying effects are dichotomous). Once an investigator has succeeded in a direct replication, then by all means, tests of generality are appropriate and useful. At that stage in the process, conceptual replications have all of the virtues that Crandall and Sherman describe.

Critics of "Social Priming" have Overlooked Large Swaths of Replicable Research

Sherman and Rivers list about a half dozen priming designs that they assert to be highly replicable (p. 3). They claim that all of these rely on "the same underlying logic of priming information in memory prior to measuring some associated behavioral response" (p. 3). Moreover, these measures, they say, "produce robust results that have been widely replicated."

Several things should be noted here. First, none of the results seem to involve priming of goals or motivations. For the most part, the studies seem to involve activation of one concept increasing the availability of information that are presumably associated in some way with the prime. This is interesting, but more akin to the priming of Meyer and Schvaneveldt (1971) than to the theoretical ideas of "social priming" investigators.

Second, it is rather awkward that the authors cite individual papers, but not series of papers where the later papers directly replicate earlier papers, much less the sort of large-scale pre-registered replication reports that would be fully convincing. The main citation they offer on the reliability of the effects they mention (Cameron, Brown-Iannuzzi, & Payne, 2012) is a meta-analysis of studies, but it has become clear in recent years that unreplicable work (even in areas like ESP) can receive strong validation from meta-analysis (even when publication bias "corrections" are incorporated; Bem, Tressoldi, Rabeyron, & Duggan, 2015). We have no idea whether the findings they cite have been or could be confirmed with direct replication.

In an email discussion list that included social cognition researchers as well as critics from 2012 – 2014, critics repeatedly asked priming researchers to nominate specific priming effects that they were confident would survive direct replication. As far as we know, no strong candidates were offered—at least, none that involved priming of motivations or behavioral choices. If there were cases where investigators could show others how to reproduce priming effects that produce a surprising change in motivation or behavioral choice, we think we would have heard about it by now (2020).

People Need to Focus on Theory not Failures of Specific Effects

Sherman and Rivers argue that "it is time to move past arguments about the reliability of specific effects and shift our energy to building theories that help us to better understand the mechanisms underlying priming effects" (p. 1).

This is but one of many examples of where Sherman and Rivers seem to move back and forth as to whether they think any of the debated priming studies represent real effects or not. It seems strange to suggest that establishing the veracity of empirical phenomena is a separate enterprise from finding theoretical support for those phenomena. They suggest a focus on building theory, but what kind of useful theory can be built around phenomena that do not exist? This seems to be taken as self-evident in cases like cold fusion or ESP. It would surely be inefficient and misleading for anyone to spend a great deal of time exploring potential mechanisms for something whose existence cannot be confirmed.

Indeed, in his 2011 book, Kahneman engaged in an extended meditation on general principles of mental organization which, he said, entailed and were confirmed by the behavioral priming studies he described in Chapter 4 of his book (Kahneman, 2011). While Sherman and Rivers seem aggravated by Kahneman's impact on their field, they should keep in mind that this prominent outsider was complying with their admonition to develop broad theory, and moreover he celebrated and popularized the work of social cognition researchers–until new research led him to grave doubts about the literature he had been relying upon.

The Doubtful Priming Effects are Characterized by Use of Between Rather than Within-Subject Designs

We basically agree with Sherman and Rivers that this is an important distinction with methodologically critical consequences. In some of our papers, we offered some quantitative comparisons of priming effect sizes drawn from within and between designs. The difference, namely bigger (purported) effect sizes for the more indirect motivational priming effects (i.e., the "social priming"), was one of the things that made us initially suspicious. We noted that highly reproducible perceptual priming effects "often have involved statistically powerful studies, using within-subject comparisons with many trials collected for each participant in each of the experimental conditions" (Harris et al., 2013). In our view, within-subjects designs make it far more economical for a research community to obtain high statistical power, which in turn promotes scientific cultures that recognize the importance of direct replication. This is an important advantage in areas like psychophysics, attention and performance, memory, and other subfields of behavioral science that seem to be functioning normally—subfields where everyone agrees that effects can be (and are) routinely reproduced (see Rouder & Haaf, 2018 for a very instructive statistical analysis.) As a general prescription for social psychology, however, a wholesale conversion to the use of within-subject designs would be likely to squelch many important and interesting avenues of research where interventions have persisting effects on the individual—something that Sherman and Rivers acknowledge. Unfortunately, for the core topics of social psychology, obtaining reliable results probably entails doing a lot of large and labor-intensive studies. To avoid a any repeat of the "train wreck", expectations about "productivity" (by universities, granting agencies, and peers) need to be adjusted. Productivity that consists in producing a stream of findings that do not hold up to scrutiny is the sort of productivity that science can live without.

References

Bargh, J. A. (2005). Bypassing the will: Towards demystifying the nonconscious control of social behavior. In R. Hassin, J. Uleman, & J. Bargh (Eds.), *The new unconscious* (pp. 37–58). New York, NY: Oxford.

Bargh, J. A. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology*, *36*(2), 147–168. doi:10.1002/ejsp.336

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype-activation on action. *Journal of Personality and Social Psychology*, *71*(2), 230–244. doi:10.1037//0022-3514.71.2.230

Bem, D., Tressoldi, P., Rabeyron, T., & Duggan, M. (2015). Feeling the future: A meta-analysis of 90 experiments on the anomalous anticipation of random future events. *F1000Research*, *4*, 1188 doi:10.12688/f1000research.7177.2

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review*, *16*(4), 330–350. doi:10.1177/1088868312440047

Caruso, E. M., Shapira, O., & Landy, J. F. (2017). Show me the money: A systematic exploration of manipulations, moderators, and mechanisms of priming effects. *Psychological Science*, *28*(8), 1148–1159. doi:10.1177/0956797617706161

Caruso, E. M., Vohs, K. D., Baxter, B., & Waytz, A. (2013). Mere exposure to money increases endorsement of free-market systems and social inequality. *Journal of Experimental Psychology. General*, *142*(2), 301–306. doi:10.1037/a0029288

Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*, *66*, 93–99. doi:10.1016/j.jesp.2015.10.002

Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS One*, *7*(1), e29081. doi:10.1371/journal.pone.0029081

Dreber, A., Pfeiffer, T., Almenberg, J., Isaksson, S., Wilson, B., Chen, Y., … Johannesson, M. (2015). Using prediction markets to estimate the reproducibility of scientific research. *Proceedings of the National Academy of Sciences*, *112*(50), 15343–15347. doi:10.1073/pnas.1516179112

Forsell, E., Viganola, D., Pfeiffer, T., Almenberg, J., Wilson, B., Chen, Y., … Dreber, A. (2019). Predicting replication outcomes in the Many Labs 2 study. *Journal of Economic Psychology*, *75*, 102117. doi:10.1016/j.joep.2018.10.009

Harris, C. R., Coburn, N., Rohrer, D., & Pashler, H. (2013). Two failures to replicate high-performance-goal priming effects. *PLoS One*, *8*(8), e72467 doi:10.1371/journal.pone.0072467

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. doi:10.1371/journal.pmed.0020124

Johnston, J.C., & Hale, B.L. (1984). The influence of prior context on word identification: Bias and sensitivity effects. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 243–255). Hillsdale, NJ: Erlbaum.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: MacMillan.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., … Nosek, B. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology*, *45*(3), 142–152. doi:10.1027/1864-9335/a000178

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology: General*, *90*(2), 227–234. doi:10.1037/h0031564

O'Donnell, M., Nelson, L. D., Ackermann, E., Aczel, B., Akhtar, A., Aldrovandi, S., … Zrubka, M. (2018). Registered replication report: Dijksterhuis and van Knippenberg (1998). *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *13*(2), 268–294. doi:10.1177/1745691618755704

Pashler, H., & De Ruiter, J. P. (2017). Taking responsibility for our field's reputation. *APS Observer*, *30*(7).

Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of social distance? Failure to replicate effects on social and food judgments. *PLoS One*, *7*(8), e42510. doi:10.1371/journal.pone.0042510

Pashler, H., Rohrer, D., & Harris, C. R. (2013). Can the goal of honesty be primed? *Journal of Experimental Social Psychology*, *14*, 959–964.

Rasinski, K. A., Visser, P. S., Zagatsky, M., & Rickett, E. M. (2005). Using implicit goal priming to improve the quality of self-report data. *Journal of Experimental Social Psychology*, *41*(3), 321–327. doi:10.1016/j.jesp.2004.07.001

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, *95*(3), 385–408. doi:10.1037/0033-295x.95.3.385

Rohrer, D., Pashler, H., & Harris, C. R. (2015). Do subtle reminders of money change people's political views? *Journal of Experimental Psychology: General*, *144*(4), e73–e85. doi:10.1037/xge0000058

Rouder, J. N., & Haaf, J. M. (2018). Power, dominance, and constraint: A note on the appeal of different design traditions. *Advances in Methods and Practices in Psychological Science*, *1*(1), 19–26. doi:10.1177/2515245917745058

Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., … Moore, C. (2013). Priming intelligent behavior: An elusive phenomenon. *PloS One*, *8*(4), e56515 doi:10.1371/journal.pone.0056515

Vohs, K. D., Mead, N. L., & Goode, M. R. (2006). The psychological consequences of money. *Science (New York, N.Y.)*, *314*(5802), 1154–1156.

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Reply to comments on "puzzlingly high correlations in fMRI studies of emotion, personality, and social Cognition." *Perspectives on Psychological Science*, *4*(3), 319–324. doi:10.1111/j.1745-6924.2009.01132.x

Wicherts, J. M. (2011). Psychology must learn a lesson from fraud case. *Nature*, *480*(7375), 7–7. doi:10.1038/480007a

Wilson, B. M., Harris, C. R., & Wixted, J. T. (2020). Science is not a signal detection problem. *Proceedings of the National Academy of Sciences*, *117*(11), 5559–5567. doi:10.1073/pnas.1914237117