Multi-Site Replications in Social Psychology: Reflections, Implications, and Future Directions

Roy F. Baumeister[1] , Brad J. Bushman[2] and Dianne M. Tice[3]

[1] The University of Queensland (Australia)
[2] The Ohio State University (USA)
[3] Brigham Young University (USA)

Abstract. Social psychology findings have fared poorly in multi-site replication attempts. This article considers and evaluates multiple factors that may contribute to such failures, other than the "crisis" assumption that most of the field's published research is so badly flawed that it should be dismissed wholesale. Low engagement by participants may reduce replicability of some findings (while not affecting certain others). Incentives differ between original researchers and replicators. If multi-site replications are indeed biased toward failure, this may have a damaging effect on the field's ability to build correct theories.

Science seeks to discover and provide evidence for general principles. Most scientific publications provide some experiments that fit a theory about what causes what. Broad conclusions are however only justified if that theory applies in other settings as well. To verify them, replication is essential: Do other observations, typically collected by other scientists, support the same principle?

In light of the importance of replication, psychologists have recently begun putting increasing emphasis on studies that seek to reproduce prior findings. The initial results were not encouraging. A much publicized effort by Nosek and colleagues (Open Science Collaboration, 2015) sought to re-do 100 prominent findings from social and cognitive psychology. The overall rate of success was low, and indeed in social psychology it was considerably lower than in cognitive psychology. This finding, combined with several cases of fraud and other developments, led to widespread concern that social psychology's knowledge base was full of false positive findings (Simmons et al., 2011). Although all fields of science have encountered problems with replication, the problem was seen as particularly acute in social psychology, leading to a general sense of dismay that has been widely dubbed the "replication crisis." We are not disposed toward catastrophizing, which may be a common tendency based on the human penchant for overstating problems and negative developments (Baumeister et al., 2001), but we recognize that the brouhaha over replicability has prompted social psychologists to adopt somewhat drastic new directions, possibly limiting the scope of future work. Indeed, the adoption of crisis mentality may be an unfortunate tendency that often leads to a responsive solution that in the long run may be more damaging than the original problem itself (see Tierney & Baumeister, 2019). We fear that the replication brouhaha may produce just such a destructive overreaction, such as the broad abandonment of laboratory observations of human behavior —precisely what elevated social psychology to interdisciplinary prominence in the first place.

One positive response to the problem was the adoption of multi-laboratory replication projects. These proceeded by signing up a multitude of different laboratories to re-run a particular, previously published experiment.

Correspondence concerning this article should be addressed to Roy F. Baumeister. The University of Queensland. Saint Lucia Campus. 4072 Saint Lucia (Australia).

E-mail: r.baumeister@uq.edu.au

Conflicts of Interest: None.

Data Sharing: This article does not report original data. Nevertheless, it is based on a previous project, and the codings for that article are available to everyone at https://osf.io/t7u3x/

The results of the various laboratories could thus be combined to see how well the original finding fared. At first blush, this seemed like a methodological ideal. After all, conducting the same experiment in different places would provide good insight into how well others could reproduce the original finding, and the greater statistical power of combining results from multiple laboratories would enable a clear and precise appraisal of the strength and robustness of the effect. Indeed, many researchers were enthusiastic about the potential for multi-site replications not only to verify original results but to furnish more precise estimates of how strong the effect is. After all, there is broad consensus that effect sizes in the published literature are artificially inflated, because of publication bias: If the same hypothesis is tested many times, the failures will not be published, while the most successful ones may succeed at being published. A multi-site project can report how many labs found significant effects , —and how many did not.

Alas, these multi-site replications have not lived up to expectations for being the saviors of social psychology. If anything, they intensified the sense of crisis, especially insofar as some well established effects failed to find support. As noted below, (social) priming has been a particular sore point. Hundreds of publications have reported significant findings in favor of priming, and several meta-analyses have confirmed the reality of priming (Dai et al., in press; Weingarten et al., 2016). But priming's record in multi-site replications is dismal: Over a dozen projects have tested priming in multiple laboratories, with essentially zero success. (One mini-study reported a successful finding in support of the original, but it seems to suffer from a serious confound; see below.)

Concern over the frequent failures of multi-site replications prompted us to conduct a comprehensive review of all such studies in social psychology (Baumeister et al., in press). Our goal was to read the reports and seek some insight into why so many social psychology findings fail in multi-site replication.

To be sure, there are many who say there is no need for such close inspection, because the answer is obvious: They think most social psychology research in the past half century is fatally flawed. False positive findings will not replicate, after all. This view assumes that the field's knowledge base is hopelessly infected with false positive findings, and the only way forward is to discard and ignore all the research done up till now and start over with new and better methods. Indeed, many have suspected that most social psychologists up till now have been faking their data or at least engaging in p-hacking practices that disguise random, fortuitous trends as solid scientific findings.

Our approach was explicitly more upbeat and optimistic. We assumed both original researchers and multi-site replicators are honest scientists who are sincerely trying to advance knowledge, including advancing their careers by conducting good research. Admittedly, this put us into a puzzling situation: The most common result has been that the original researchers found significant results but replicators found little or no support for the same hypotheses.

The particular outcome that prompted us to conduct the review was a multi-site replication of the finding that contemplating one's mortality and death causes participants to defend their cultural worldview against criticisms. This has been a central tenet of terror management theory (Pyszczynski et al., 1997), but the multi-site replication by Klein et al. (2021) found no support for it at all.

One of us attended the 2020 conference where these null findings were first presented. It was implausible that the present author was biased in favor of terror management theory, because he has disputed their theory at multiple levels for several decades (e.g., Muraven & Baumeister, 1997). Yet his laboratory has also replicated the worldview defense findings, successfully, over a dozen times, in multiple publications (DeWall & Baumeister, 2007; Gailliot et al., 2006; Schmeichel et al., 2009). Thus, although he vigorously disputes the broader claims of terror management theory, he cannot doubt the genuineness of some of their laboratory findings. Why, then, would a genuinely correct result fail so utterly in the multi-laboratory replication? The disturbing implication was that multi-site replications contain some serious flaws that bias their results toward null results.

Results of Review and Meta-Analysis

An exhaustive literature search yielded 36 articles that reported a multi-site replication project in social psychology (Baumeister et al., in press). There were some additional papers filled with what we called mini-studies. These were brief studies, typically taking the participant just a minute or two to complete, and typically done in clusters (so that each participant at each session might do a dozen different experiments, one right after another).

Focusing on the main studies, each of which was dedicated to a single replication, we amply confirmed the poor replication record for social psychology. We defined successful replication as finding a significant result after combining data across the multiple laboratories, that was consistent with the original finding. Only four of the 36 were successful. More stringent criteria would reduce the success rate even further. In particular, only one of the 36 yielded an effect size comparable to the original. As a further blow to social psychology, that lone full success (Ito et al., 2019) did not even present itself as social psychology, instead being published as research in applied cognitive psychology.

Several additional studies were considered partial successes. That is, they did report both a significant and a nonsignificant result, depending on analysis strategy, exclusion of large amounts of data, and the like. Even so, 27 of the 36 (75%) were complete failures.

Our purpose in this article is to elaborate on the results and conclusions from that earlier paper (Baumeister et al., in press). We refer interested readers to that article for full details. Here, we seek to provide additional insights and reflections, including some points that could not be covered previously due to length limits or editorial negotiation, as well as some that have emerged since that article was completed. The goal is to understand the widespread discrepancies between the results of multi-site replications and original articles.

Incentives and Bias

One obvious hypothesis as to why original articles report significant results while replications report non-significant ones is that the scientific publishing system has built-in incentives that bias in opposite directions. The assumption of publication bias in original research is generally accepted. Few journals, and particularly elite journals, like to publish original investigations that report null results. We find this unsurprising. Null results are inherently ambiguous, as nearly every researcher learns in graduate school. Null results may indicate that the hypothesis is wrong —but could also indicate poor operationalization, high error variance, or other problems. The journals mainly publish significant findings. Editors sometimes instruct authors to delete weaker studies from their manuscripts, so that the eventual publication presents only the best results.

Less discussed, but potentially just as important, is the possibility that journals may prefer to publish unsuccessful than successful replications. Editors typically have limited page allotments and wish to use it to advance the field. They may consider that a successful replication merely confirms what is already known —so it does not advance knowledge. In contrast, a failed replication does advance the field in the sense of contributing to the self-correcting nature of science. If a well established finding fails to replicate, that may be a newsworthy sign that the field should reconsider its general theory. A survey of journal editors in economics confirmed that they preferred to publish failed rather than successful replications (Galiani et al., 2017). While conducting our review, we heard multiple anecdotes by replication researchers who said that they found it difficult to publish successful replications. The classic "file-drawer problem," consisting of unpublished findings, may contain a fair number of successful replications.

These conflicting editorial preferences may be quite understandable, but they do potentially set up a perverse incentive system. Original researchers know that their chances of publication are best served by getting significant results. In contrast, replicators may realize that their chances of publication are best served by reporting nonsignificant results.

We may speculate further on how the incentives translate into the daily activities of young researchers. (We are discussing these in terms of sincere research efforts and not implying any fakery.) Someone testing his or her new theory will be highly motivated to provide a careful test, in which the participants fully understand the instructions and manipulations and can respond in a relevant fashion. Such a researcher knows that the chances for publication depend on getting a significant result and therefore wants to provide an unambiguous test.

In contrast, the incentives are quite different for a young researcher participating in a multi-site replication. Publication does not depend on the results, and indeed in some cases publication has already been approved regardless of how the results will turn out. (Indeed, as already noted, the chances of publication may be improved by failing to replicate.) For such a researcher, the logical strategy is to get it done as quickly, easily, and efficiently as possible. Ensuring a careful and sympathetic test is not relevant. Moreover, even if one's conduct of the study is a bit casual, one knows that one's data are only a small part of the giant project and therefore unlikely to make or break the result. Social loafing would be understandable, even rational, in such a case (Karau & Williams, 1995).

We cannot prove the difference, but it certainly fits the evidence. Original studies may be done with loving care so as to provide the optimal test of the hypothesis. Replications may be conducted so as to get large amounts of data rapidly and efficiently, which often does not entail making sure to provide the optimal test. Anecdotally, one of the Terror Management Researchers told us that his graduate training emphasized starting each laboratory session with some small talk to forge an interpersonal bond with the participant, so as to ensure the participant was attentive and engaged (T. Pyszczynski, personal communication, August 9, 2022). A researcher oriented toward collecting large amounts of data efficiently might readily dispense with such niceties.

Was the Replication a Valid Test?

When original studies and replications point to different conclusions, one wants to know which one is more valid. The multi-site replication has several advantages, including nearly universal preregistration, a much

larger total sample, and multiple different places. On the other hand, we found quite a few of the replications — and none at all of the original studies— reported that their manipulations had failed. This took the form of a nonsignificant manipulation check.

As a hypothetical example, suppose the original study manipulated sadness and found that participants in the sad condition performed significantly worse than neutral-mood controls on a math test. Suppose, also, that the multi-site replication failed to replicate that result —but also reported that participants in the sad treatment condition were no sadder than those in the control. Such a result does not invalidate the original finding. It merely raises the question of why the manipulation of sadness failed to work as well in the replication as in the original study.

It seems essential to distinguish true failures to replicate from operational failures. A true failure to replicate falsifies the original finding. For that, it needs to have manipulated the independent variable successfully. If the manipulation fails, then the study has not tested the hypothesis. This was a common pattern. While operational failure is disturbing, it does not mean that the original finding and its theoretical point were wrong. Further work may investigate why the manipulation succeeded with the original sample but failed with the replication.

Obviously, if an experiment fails to manipulate its independent variable, it is unlikely to yield significant results on the dependent variable: It has failed to provide a test of the hypothesis, so even if the hypothesis is entirely and generally true, the experiment's results will not support it. The implications of manipulation failure are extensive and to some extent can restore confidence in the published literature. A failed replication only casts doubt on the original finding insofar as it provided a valid test of the hypothesis —and if the manipulation check was nonsignificant or even just small, then it was not a valid test, and null results do not justify such doubts.

Are Participants Engaged?

Two particular big findings emerged from our perusal of the multi-site replications in social psychology. Both of them point toward the conclusion that participants in these studies are less engaged than participants in many original studies.

One finding was the frequently high rate of data exclusion. (We defer the second finding to a later section.) Discarding data from some participants has been a problem throughout the history of social psychology. In the early years, when $n = 10$ per cell was the norm, a single outlier could make or break a study. Deceptions were often extensive, and probing for suspicion was a regular part of the procedure. Methodologists insisted that decisions to exclude data be made without knowing whether the participant's data fit the hypothesis, but sometimes that was impossible, and it is hard to know whether researchers scrupulously adhered to that guideline even when it was possible. Journal reviewers attended carefully to how many participants were discarded. As sample sizes increased, reviewers remained attentive to how many participants were discarded, and it was difficult to publish a study if more than a few participants were thrown out.

In contrast, we found the multi-site replications frequently discarded large proportions of data, often 20-40% and sometimes more than half. In a few unfortunate cases, exclusions were unequally distributed across conditions, such that a majority of participants were discarded in the experimental condition but only a tiny minority in the control condition —which makes the samples potentially very different in terms of personality and other factors.

Typically, the discarding of data from multi-site replications was the result of pre-registered criteria. Researchers had often stipulated that participants who made too many errors or failed attention checks should be discarded. The researchers were then often surprised and dismayed to find that a third of their data had to be discarded. They had pre-registered the exclusion criteria as a way of deleting the odd participant here and there who might not be paying attention.

The high rates of exclusion suggest that low engagement is a pervasive problem in multi-site replications. As the previous section noted, the incentives for the researcher in a multi-site project are to get the data collected as quickly and efficiently as possible. Failure to engage the participants may be a common byproduct of that strategy.

Manipulation checks were also often weak. These too suggest that the replications, while technically conforming to the details of the original procedure, may not be engaging participants as fully as those in the original study were.

Moreover, as the previous section noted, many of the replications reported manipulation check data indicating that the manipulation was unsuccessful, and for others it was quite weak. Low motivational engagement is one prominent cause of a failed manipulation check. To pursue the example from the previous section, participants who aren't paying close attention or who are indifferent to the study may not respond to a sadness manipulation.

The low engagement conclusion dovetails well with the previous point about researchers' incentives. Researchers conducting original experiments to test their new theories presumably exert themselves to get the most sympathetic test, and that includes ensuring

that participants are heavily engaged in the procedure. Researchers conducting multi-site replications have no such incentive, and if anything their incentives favor getting it over with. Failing to engage the participants in the procedure may be a common result.

We note also that this problem may be more acute for social psychology than other disciplines, because social behavior often involves motivational engagement. In this connection, it is instructive to consider some of the mini-studies. Given that these are often quick cognitive reactions, they may not require much engagement, and so their record of replication was somewhat better than the studies we focused on. As a vivid example, Klein et al. (2018) replicated the finding by Gray and Wegner (2009) that people blame a grown man who accidently hurts a baby more than they blame a baby who accidentally hurts a grown man. (In both cases, the accident involved knocking over some glasses.) Such a finding does not require highly motivated or emotional engagement.

Live Social Interaction

We noted that there were two big and surprising findings from the review by Baumeister et al. (in press). The second one involves live interaction and is thus relevant to the generational shift in social psychology's methods. Early work relied heavily on getting participants highly engaged by staging an engrossing live social interaction. Over the years, this has dwindled to some degree, replaced by having participants sit at computers and make ratings (e.g., Baumeister et al., 2007). Real experiences have been replaced with simulated or imaginary ones. Baumeister et al. (in press) coded the 36 multi-site replications as to whether the included live social interaction or not. (It proved necessary to have an in-between category of computer-mediated or computer-simulated interactions, in which participants were led to believe they were having a live social interaction via computer hookup, even though no live other person was present with them.)

The majority of the 36 multi-site replications had no live social interaction. Yet the ones that did include live interaction —and this included even just having the experimenter remain present with the participant and interact with him or her throughout the procedure— had a vastly better success rate. None of the pure failures included live social interaction, whereas most of the successes and partial successes did.

Only one of the 36 included unscripted live conversation among participants, and that one was also the single most successful one, indeed the only one to match the effect size of the original finding (Ito et al., 2019).

The implication is that social psychology replicates better when it includes social interaction. It is seductively tempting to collect large amounts of data by having a great many participants sit alone at computers and make ratings, and we do not dispute the value of such work. But quite possibly such procedures do not do justice to the social animal. People may be much more involved and engaged when they are dealing with a live human being than when they are merely sitting in a cubicle alone making ratings of on-screen stimuli. The recent COVID pandemic showed that when professors simply give the same lectures online, and students hear these recordings while alone in their dorms, without any interaction, the students learn less and rate their courses less favorably, as compared to being together in the same lecture hall with a live speaker (e.g., Tice et al., 2021). Listening alone to a pre-recorded lecture may be less engaging than sitting among an attentive audience for a live in-person lecture, not least because of the option of asking questions or in other ways interacting with the instructor.

A basic point about human nature may be at issue here. Humankind evolved to interact constructively with each other, far more so than in other apes. Communication and cooperation are the hallmarks of human evolution (e.g., von Hippel, 2018). Much of the human mind may be geared to respond to others in these ways. Sitting alone in a computer cubicle may not engage the full human being to the same extend (e.g., Oppenheimer et al., 2009).

All of this sets up a dilemma for future research in social psychology. The replication brouhaha has led to a widespread insistence on giant samples, based on the (unproven) assumption that they will yield more replicable results. In practice, getting large samples means streamlining the experimental procedures and especially using online raters. But such procedures typically minimize or eliminate any live social interaction. (And live social interaction does appear to improve replicability.) Will future research do better by increasingly emphasizing large online samples, or smaller ones with live interpersonal interaction? We hope it will try both.

Happy Reactions to Replication Failures

An intriguing dimension to the replication crisis has been the often positive if not outright gleeful reaction to replication failures. Social psychologists have long been recognized as highly critical of each other's work. Anecdotally, grant agency officers often commented with dismay at how negatively social psychologists review each other's proposals, with the broad result that it is more difficult for social psychologists than for researchers in other fields to obtain research funding.

We have no systematic data on this, but abundant anecdotal evidence supports the view that the poor replication record of social psychology has produced an outpouring of positive responses among researchers

in the field (at least those whose own work has not come under attack). Reviewers seem happy to reject new papers based on claiming that failed replications indicate that the phenomenon does not exist, so any new findings can be immediately discounted. The responses extend beyond occasional *Schadenfreude* (i.e., pleasure over another's misfortunes) to suggest an eagerness to dismiss and ignore most of the work that social psychologists published over its first half century.

The gleeful reactions to failed replications have a downside. Most obviously, senior researchers have devoted their lives to advancing social psychology and are understandably dismayed to see that the younger generation is happy to discredit and discard their life's work. Younger researchers may also be perturbed to think that if they spend the coming decades seeking to advance the field, their work may be summarily dismissed. If you work hard and become successful, then your colleagues will be hoping to see you brought down —will rejoice over your troubles and eagerly embrace flimsy evidence that you were wrong about everything.

Progress must involve correcting prior mistakes. Then again, a field that rejects its heritage is indeed arguably in a troubled condition. Ultimately, progress is presumably optimized by building on earlier work rather than dismissing it wholesale. To be sure, if social psychology's accumulated knowledge is indeed deeply, pervasively, and fatally flawed, then the best course is to dismiss it all and start over. (However, starting over with MTurk samples seems distinctively unpromising; see Webb & Tangney, in press.) But if one assumes that early generations of social psychologists were doing reasonably good work, then it would be best to build on rather than dismiss it.

Impact on the Field

Here we briefly consider the impact on the field of attempting multi-site replications of its most common findings, especially considering the high failure rate. We noted at the outset that this review was stimulated by Klein et al.'s (2021) failed replication of mortality salience effects —when such effects have been found many times (including by some of us).

We first comment on the "big four" findings that have been widely replicated in previous social psychology research but that encountered some negative results in multi-site replications: mortality salience, ego depletion, two routes to persuasion (elaboration likelihood model), and social priming. Each of these has well over a hundred published findings supporting it in the research literature, but each has had its image tarnished by the multi-site replication attempts.

Terror management and mortality salience. Mortality salience has had the one attempted multi-site replication (Klein et al., 2021). Given the abundant data in the literature, including some of our own studies, we think it is implausible to question the reality of the effect. The failure of the multi-site replication may point to the general weakness in the multi-site approach, though that requires explanation. The particular study may have been confounded by historical change. The original finding of worldview defense was reflected in condemnation of anti-American writers, but that depends on the participants having a strongly pro-American worldview. The time difference between the original 1994 study and the replication in 2020 coincides with a much discussed (yet still probably underappreciated) shift in American education at all levels to instill a less patriotic and much more critical and negative attitude toward America. It seems worth trying another multi-site attempt using a different measure, more in keeping with the mindset of today's students. Low engagement may also be a factor: if participants do not actively and emotionally contemplate their death, one cannot expect significant effects.

Ego Depletion. Social media have depicted ego depletion as a case of failed replication, and some scholars have begun to insist that there is no such phenomenon. Yet what if ego depletion is true? Crucially, only four social psychology findings have been significantly supported in a multi-site replication –and one of them is for ego depletion (Dang et al., 2021). Thus, the very short list of success stories in multi-site social psychology replications includes ego depletion. Invoking Lord et al.'s (1984) finding that considering the opposite is a useful heuristic strategy for overcoming bias, we consider: What is the case that ego depletion is the single best replicated finding in social psychology?

The case seems reasonably strong (Baumeister & Tice, 2022). To qualify as a well replicated effect, an effect would presumably have to point to support in the forms of (a) multiple significant findings, indeed preferably from different laboratories and with different methods; (b) pre-registered studies; (c) real-world or non-laboratory findings; and (d) multi-site replication. The last is quite rare in social psychology, but ego depletion stands out as unusually successful, given that Dang et al. (2021) supported ego depletion in one of the only four fully successful replications of anything in social psychology. Two additional multi-site studies provided mixed support. Hagger et al. (2016) was reported and highly publicized as a failed replication, but a reanalysis by Dang (2016) correcting for the manipulation check indicated that Hagger's study did show significant evidence of ego depletion, to the (slight) extent that the manipulation worked. Later, an ambitious study by Vohs et al. (2021) found significant evidence of a small

effect when the full sample was analyzed, but the exclusion of a third of the sample dropped this below significance. This could be a statistical power issue, given that the effect size remained the same (but became nonsignificant) after discarding over a thousand participants. They likewise found that for participants who reported more fatigue on the manipulation check, the ego depletion effect was significantly stronger.

Ego depletion's other credentials as a well replicated phenomenon would seem highly competitive. Published significant findings numbered around 600 already some years ago (Friese et al., 2019). There are preregistered successful replications (Garrison et al., 2019; Keller & Kiss, 2021) and multiple and diverse non-laboratory real-world findings (e.g., Danziger et al., 2011; Hurley, 2015; Philpot et al., 2018; Trinh et al., 2021). We are hard pressed to find many other findings in social psychology that can match that, except perhaps elaboration likelihood model (see below). And even if another one does come along, it seems clear ego depletion is among social psychology's few best replicated effects[1].

So what? Self-regulation is centrally important to self-theory, to relationship quality, to evolutionary theory, to positive outcomes, and to society. Insofar as psychology seeks to build a valid understanding of how the human mind works, it needs to get self-regulation right. If ego depletion is true, it cannot be omitted from that theory. Yet while having a reasonable case to be the best replicated finding in all social psychology, it has acquired the reputation of being incorrect because of adverse publicity associated with Hagger et al.'s (2016) multi-site replication. If multiple true and important patterns in human social behavior and judgment are discredited because of false-negative multi-site replication failures, the field's ability to end up with the truth will be seriously compromised.

Two routes to persuasion. The elaboration likelihood model mapping two different causal routes to persuasion has been tested in many ways. An early multi-site replication reported a true failure (Kerr et al., 2015; see also Ebersole et al., 2016), given that manipulation checks were significant (at least in some places) while the dependent measure was not. But the predictions regarding different routes to persuasion were supported in a replication by Ebersole et al. (2017). To be sure, one of the independent variables in the interaction was an individual difference, and in general our review avoided inclusion of individual difference findings.

(Our impression is that they generally replicate better than social psychology experiments; questionnaire-assessed individual differences produce reliable differences that are less dependent than situational variables on effective laboratory manipulations.) Nevertheless, the other variable was manipulated, and ELM likewise has a great many supportive findings in the literature. Given its multi-site successful replication by Ebersole et al. (2017), as well as other supportive evidence, it rivals ego depletion as one of social psychology's most frequently replicated findings. Again, if ELM is true but discredited on the basis of some multi-site failures, the field's ability to achieve a valid understanding of attitude change would be damaged.

It is noteworthy that, like ego depletion, the ELM's two routes to persuasion finding achieved its fully successful replication only after two earlier and less successful attempts. The failure by Ebersole et al. (2016) elicited constructive criticisms by Luttrell, Petty, and Xu (2017), which then led to the successful outcome by Ebersole et al. (2017).

Priming. Priming presents a complicated challenge. The hundreds of published findings constitute a formidable body of supportive evidence that cannot easily be dismissed (see Weingarten et al., 2016, and Dai et al., in press, for reviews). Yet over a dozen multi-site studies with different priming procedures have failed to show priming effects. Some of these included manipulation checks indicating successful operationalization, though others were operational failures or ambiguous. The mini-studies in the Klein et al. (2014, 2018) papers had two additional priming studies that failed to replicate — and one successful replication. The lone success out of these 17 multi-site attempts to replicate priming effects was described by authors as priming a consumer mindset, but the manipulation consisted merely of referring once to people as "consumers" rather than as "individuals." Calling them "consumers" led participants to estimate that these other people would not conserve water as extensively as they estimated when people were described as individuals. The finding may be confounded insofar as "consume" is the literal opposite of "conserve." Hence even this lone positive finding does not inspire confidence in the replicability of priming. The poor record of multi-site replication attempts with priming is especially surprising given that it is a cognitive effect, and in general the cognitive effects have replicated better (both within social psychology and elsewhere).

Two crucial points raise questions about making some sweeping negative judgment about priming based on the multi-site replication failures. First, priming operates presumably by activating some motivation in the participant. Weingarten et al.'s (2016) meta-analysis found that priming effects were larger and stronger in

---

[1] A multi-site replication of cognitive dissonance is currently underway. If it provides significant support, we would happily concede first place to cognitive dissonance as social psychology's best replicated finding. Even so, ego depletion and the ELM would be among the handful of best replicated findings.

proportion with participants' higher motivation. The present finding that low participant engagement is a common problem with multi-site replications would be consistent with concluding that priming typically fails in these settings because participants are not sufficiently motivated for the primes to have their effect. As a further sign, Corker et al.'s (2020) failure to replicate action priming effects with a thought-listing procedure noted that their participants across all labs and conditions generally listed far fewer thoughts than in the original. That indicates lower effort, consistent with the general pattern of low engagement.

Second, the manipulation checks for priming rarely confirm that the construct and motivation were actually activated. Phrases such as "break his leg" could be an aggressive prime —but might simply evoke a skiing accident. An original study by Williams and Bargh (2008) found effects by having participants hold a cup of warm coffee, but some follow-ups have used hot rather than warm stimuli (Bargh & Melnikoff, 2019) — and, crucially, the social implications of hot vs. warm may be quite different. It is thus unclear whether pro-social warmth was successfully primed, as opposed to "hot" antisocial impulses. (Furthermore, ambient temperature may moderate the effect: Warm physical sensations may promote prosocial feelings in cold weather but not in hot weather; Fay & Maner, 2020).

Thus, we have four findings that are prominent in recent social psychology research. Each of them has a hundred or more significantly supportive findings in the literature but at best a mixed record with multi-site replication. Priming is the extreme example, with hundreds of significant findings in the literature but over a dozen failures, and no clear successes, in multi-site replication. The multi-site method may well be biased toward false negative findings, particularly insofar as all the original findings depend on high participant engagement. Very plausibly, participants must get emotionally involved in contemplating death to obtain terror management effects, must exert high effort to deplete their willpower to obtain depletion effects, must care enough to reflect on persuasive sources' credibility to obtain persuasion effects, and must resonate personally with primed goals to exhibit priming effects.

These conclusions are supported by our evidence that replications were somewhat more successful with what we have called mini-studies, that is, studies that take only couple minutes or less and, crucially, that do not rely on participants becoming personally engaged (e.g., Klein et al., 2018). For example, the finding that people blame a (hypothetical) man who accidentally hurts a baby more than they blame a baby who accidentally hurts a grown man (replicated by Klein et al.) probably does not require deep emotional involvement or careful thought. The same goes for a false consensus effect, in which people estimate that many others would share their opinions. This line of thought suggests an alternative way forward, which is for social psychology to dispense with studying phenomena that engage people's motivations and limit research to quick thought-reaction procedures. Such an approach (which does appear to be the trend in the field at present) may have the benefit of improving replicability in multi-site online procedures, though some (including ourselves) would object that there are hidden costs in neglecting to study more highly involving, behavioral phenomena.

The broader implications suggest a pessimistic view of social psychology's future. If it is true that multi-site replications in social psychology are biased toward failure, then the course of new discoveries is likely to be curvilinear. Original researchers may identify a phenomenon and publish it. If it generates widespread interest and excitement, others may seek to build on that work. Once sufficiently established, however, it will be subjected to a multi-site replication, which is likely to fail. In a worst-case scenario, social psychology may continue to discredit its own best work, leaving new researchers uncertain what to believe and requiring them to rely mainly on findings that have not attracted replication interest, such as minor or rarely studied phenomena.

Ego Depletion, Anchoring-Adjustment, and the Engagement Issue

The previous section noted that the present authors were pleasantly surprised to realize, during the process of reviewing the literature, that ego depletion has an exceptionally strong record of replication, not just in multi-site replication but also in other work. Baumeister published a blog touting ego depletion as social psychology's best replicated finding and included a strong invitation to readers to nominate rivals for that title. Despite thousands of reads of the blog, only one other nomination was forthcoming (covered in a subsequent blog): anchoring and adjustment. The finding is that when people estimate a number after being given another number, their estimates are overly close to the first number (Jacowitz & Kahneman, 1995).

Anchoring and adjustment effects have been targeted in several multi-site replications, though typically as mini-studies. There was one reported failure and several successes (Klein et al., 2014, 2018). All in all, anchoring and adjustment has a reasonable claim to be even better replicated than ego depletion. Nevertheless, two important caveats must be acknowledged. First, although a broad definition of social psychology would include the anchoring-adjustment effect, it is not a very social phenomenon: It is merely a common mistake people make when estimating a number.

Second, anchoring-adjustment probably does not require participants to be all that engaged. In contrast, ego depletion is a social psychology phenomenon and only occurs if people are heavily engaged. With ego depletion, participants must exert themselves on the initial task sufficiently so as to use up some of their willpower energy, so that they perform worse on the dependent measure. Moreover, their engagement on the first (depleting) task must be sufficient to produce the level of mental fatigue that can impair their subsequent performance. In contrast, the anchoring effect may occur among participants who are only slightly engaged. Indeed, mistakes with numbers may actually increase when people are not paying close attention.

Converging anecdotal evidence is relevant. The first author has been named and thanked in two multi-site replication projects. His recollection is that in both cases, his suggestions were respectfully solicited —and then almost entirely disregarded. (In one case, his extensive and detailed comments led to changing one word in the final write-up.) The reason for disregarding his input for study design in both cases was that the suggested procedures were too difficult and labor-intensive for the multi-site replication. This reveals the priorities among multi-site replications: They need to be quick and easy to do. Getting research participants highly involved takes a fair amount of work, and so multi-site researchers favor procedures that do not require that.

Recommendations for Future Multi-Site Studies

A strong scientific field needs multiple methods. We anticipate that social psychology will continue to attempt multi-site replications. The following recommendations are intended to improve the value of such efforts for the field's progress toward building correct theory.

First, the multi-site replication procedure should be recognized as a very weak test of the hypothesis, possibly biased toward false negatives. Significant positive findings should be valued, while non-significant findings may be regarded as the norm. Some systemic change could help revise the editorial bias in favor of failed replications and its associated incentivization.

Smaller effect sizes are to be expected in multi-lab than in original investigations. A smaller effect size should not be the basis for declaring a replication to have failed, possibly except if it is truly miniscule. The notion of a true effect size for some social psychology variable, such as interpersonal rejection or cognitive dissonance, may be problematic if not absurd. After all, it is hard to imagine that being dumped by the great love of one's life would have the same effects as being excluded in an online "Cyberball" game by a pair of strangers. How badly theory-building is hampered by the inability to establish such a true effect size remains in question.

Operational failures need to be distinguished rigorously from falsification of the hypothesis. Both matter but the implications are quite different. Manipulation checks are therefore highly important and should be included wherever possible (even if absent from the original study). Conclusions about theory should only be revised if the multi-site replication yields a significant and large difference on the manipulation check but no difference on the dependent variable. Future studies should emphasize manipulation checking. The effect size on the manipulation check is an important marker of how well the hypothesis was tested.

Hauser et al. (2018) pointed out that manipulation checks may initiate processes that confound the study, such as by making people aware of their emotions (see also Kühnen, 2010). If there is concern that the manipulation check may alter the findings, researchers may try administering some of them after the measure of the dependent variable (as Hauser et al. suggest) or even using manipulation checks on only half the sample. One could then test whether results differ as a function of the presence versus absence of manipulation checks prior to the dependent variable.

Low engagement should be recognized as a common problem in multi-site replications. Possibly new statistical methods can be developed to ascertain whether effects are replicated among the participants whose manipulation check data indicate successful manipulation. If social psychologists continue the current trend of favoring online data collection with minimal interpersonal interaction, they should perhaps focus more heavily on phenomena that do not require participants to be engaged or motivated.

Live social interaction also appears conducive to successful replication. Future multi-site researchers might profitably consider how to include live social interaction in their procedures. Live interaction may help reduce the problem of low engagement.

The distribution of targets for multi-site replication attempts has been badly skewed, especially considering 16 attempts to replicate priming. It would be better for the field to focus on some basic and common findings, such as that people favor external attributions for failure, and that aggression is increased by interpersonal provocation.

More research on boundary conditions and moderator variables may help resolve inconsistencies between significant original findings and failed replications.

Concluding Remarks

Social psychology has not just one but several different problems with replications. Fortunately, not all of them

undermine the credibility of the accumulated research literature. Operational failures —replications that are unable to provide a valid test of the hypothesis— are common. While these are disturbing, they do not challenge the original finding. Low motivational engagement, failed manipulations, and other factors reduce the theoretical impact of some failures to replicate. To be sure, some failures to replicate do falsify the original conclusion and call for revising theoretical conclusions.

The broad trend toward abandoning live social interactions and instead relying on large samples of participants sitting alone at computers while making ratings is well underway but may have severe costs. These include loss of behavioral observation and abandonment of what made social psychology a noteworthy discipline in the first place. It is also far from clear that these large-sample online studies will replicate better than other methods.

In our view, the strength of a discipline, especially in social science, resides in methodological diversity. Social psychology flourished in the late 20th century because researchers (and journal reviewers) sought to study each phenomenon with the best methods available, and preferably with a diversity of methods and operationalizations. That required adjusting methodological standards according to topic. Online data collection has its place but cannot study many things, and researchers who use other methods cannot compete for journal space if reviewers and editors insist on large samples. Social psychologists risk the health and progress of their field by abandoning their own methodological diversity.

References

Bargh, J. A., & Melnikoff, D. (2019). Does physical warmth prime social warmth? *Social Psychology*, *50*(3), 207–210. http://doi.org/10.1027/1864-9335/a000387

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 323–370. http://doi.org/10.1037/1089-2680.5.4.323

Baumeister, R. F., & Tice, D. M. (2022). Ego depletion is the best replicated finding in all of social psychology. *Scholarly Journal of Psychology and Behavioral Sciences*, *6*, 686–688. http://doi.org/10.32474/SJPBS.2021.06.000234

Baumeister, R. F., Tice, D. M., & Bushman, B. J. (in press). *A review of multi-site replication projects in social psychology: Is it viable to sustain any confidence in social psychology's knowledge base?* Perspectives on Psychological Science.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*, 396–403. http://doi.org/10.1111/j.1745-6916.2007.00051.x

Corker, K. S., Arnal, J. D., Bonfiglio, D. B. V., Curran, P. G., Chartier, C. R., Chopik, W. J., Guadagno, R. E., Kimbrough, A. M., Schmidt, K., & Wiggins, B. J. (2020). Many Labs 5: Registered replication of Albarracín et al. (2008), Experiment 7. *Advances in Methods and Practices in Psychological Science*, *3*(3), 340–352. http://doi.org/10.1177/2515245920925750

Dai, W., Yang, T., White, B. X., Palmer, R., Sanders, E., McDonald, J. A., & Albarracín, D. (in press). Priming behavior: A meta-analysis of the effects of behavioral and nonbehavioral primes on overt behavioral outcomes. *Psychological Bulletin.*

Dang, J. (2016). Commentary: A multilab preregistered replication of the ego-depletion effect. *Frontiers in Psychology*, *7*, Article 1155. http://doi.org/10.3389/fpsyg.2016.01155

Dang, J., Barker, P., Baumert, A., Bentvelzen, M., Berkman, E., Buchholz, N., Buczny, J., Chen, Z., De Cristofaro, V., de Vries, L., Dewitte, S., Giacomantonio, M., Gong, R., Homan, M., Imhoff, R., Ismail, I., Jia, L., Kubiak, T., Lange, F., … Zinkernagel, A. (2021). A multilab replication of the ego depletion effect. *Social Psychological and Personality Science*, *12*(1), 14–24. http://doi.org/10.1177/1948550619887702

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *PNAS*, *108*, 6889–6892. http://doi.org/10.1073/pnas.1018033108

DeWall, C. N., & Baumeister, R. F. (2007). From terror to joy: Automatic tuning to positive affective information following mortality salience. *Psychological Science*, *18*, 984–990. http://doi.org/10.1111/j.1467-9280.2007.02013.x

Ebersole, C. R., Alaei, R., Atherton, O. E., Bernstein, M. J., Brown, M., Chartier, C. R., Chung, L. Y., Hermann, A. D., Joy-Gaba, J. A., Line, M. J., Rule, N. O., Sacco, D. F., Vaughn, L. A., & Nosek, B. A. (2017). Observe, hypothesize, test, repeat: Luttrell, Petty, and Xu (2017) demonstrate good science. *Journal of Experimental Social Psychology*, *69*, 184–186. http://doi.org/10.1016/j.jesp.2016.12.005

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B.V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., … Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, *67*, 68–82. http://doi.org/10.1016/j.jesp.2015.10.012

Fay, A. J., & Maner, J. K. (2020). Interactive effects of tactile warmth and ambient temperature on the search for social affiliation. *Social Psychology*, *51*, 199–204. http://doi.org/10.1027/1864-9335/a000407

Friese, M., Loschelder, D. D., Gieseler, K., Frankenbach, J., & Inzlicht, M. (2019). Is ego depletion real? An analysis of arguments. *Personality and Social Psychology Review*, *23*(2), 107–131. http://doi.org/10.1177/1088868318762183

Gailliot, M. T., Schmeichel, B. J., & Baumeister, R. F. (2006). Self-regulatory processes defend against the threat of death: Effects of self-control depletion and trait self-control on thoughts and fears of dying. *Journal of Personality and Social*

Psychology, 91, 49–62. http://doi.org/10.1037/0022-3514.91.1.49

Galiani, S., Gertler, P., & Romero, M. (2017). *Incentives for replication in economics*. (Working Paper 23576). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w23576/w23576.pdf

Garrison, K. E., Finley, A. J., & Schmeichel, B. J. (2019). Ego depletion reduces attention control: Evidence from two high-powered preregistered experiments. *Personality and Social Psychology Bulletin, 45*(5), 728–739. http://doi.org/10.1177/0146167218796473

Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology, 96*, 505–520. http://doi.org/10.1037/a0013748

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., Brand, R., Brandt, M. J., Brewer, G., Bruyneel, S., Calvillo, D. P., Campbell, W. K., Cannon, P. R., Carlucci, M., Carruth, N. P., Cheung, T., Crowell, A., De Ridder, D. T. D., Dewitte, S., … Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science, 11*(4), 546–573. http://doi.org/10.1177/1745691616652873

Hauser, D. J., Ellsworth, P. C., & Gonzalez, R. (2018). Are manipulation checks necessary? *Frontiers in Psychology, 9*, Article 998. http://doi.org/10.3389/fpsyg.2018.00998

Hurley, P. J. (2015). Ego depletion: Applications and implications for auditing research. *Journal of Accounting Literature, 35*, 47–76. http://doi.org/10.1016/j.acclit.2015.10.001

Ito, H., Barzykowski, K., Grzesik, M., Gülgöz, S., Gürdere, C., Janssen, S. M. J., Khor, J., Rowthorn, H., Wade, K. A., Luna, K., Albuquerque, P. B., Kumar, D., Singh, A. D., Cecconello, W. W., Cadavid, S., Laird, N. C., Baldassari, M. J., Lindsay, D. S., & Mori, K. (2019). Eyewitness memory distortion following co-witness discussion: A replication of Garry, French, Kinzett, and Mori (2008) in ten countries. *Journal of Applied Research in Memory and Cognition, 8*(1), 68–77. http://doi.org/10.1016/j.jarmac.2018.09.004

Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin, 21*(11), 1161–1166. https://doi.org/10.1177/01461672952111004

Karau, S. J., & Williams, K. D. (1995). Social loafing: Research findings, implications, and future directions. *Current Directions in Psychological Science, 4*, 134–140.

Keller, T., & Kiss, H. J. (2021). Do exhausted primary school students cheat more? A randomized field experiment. *PLOS ONE, 16*(12), Article e0260141. http://doi.org/10.1371/journal.pone.0260141

Kerr, G., Schultz, D. E., Kitchen, P. J., Mulhern, F. J., & Beede, P. (2015). Does traditional advertising theory apply to the digital world? A replication analysis questions the relevance of the elaboration likelihood model. *Journal of Advertising Research, 55*(4), 390–400. http://doi.org/10.2501/JAR-2015-001

Klein, R. A., Cook, C. L., Ebersole, C. R., Vitiello, C., Nosek, B. A., Hilgard, J., Ahn, P. H., Brady, A. J., Chartier, C. R., Christopherson, C. D., Clay, S., Collisson, B., Crawford, J., Cromar, R., Gardiner, G., Gosnell, C., Grahe, J., Hall, C., Howard, I., … Ratliff, K. (2021). *Many labs 4: Failure to replicate mortality salience effect with and without original author involvement*. PsyArXiv. http://doi.org/10.31234/osf.io/vef2c

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., … Nosek, B. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology, 45*(3), 142–152. http://doi.org/10.1027/1864-9335/a000178

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Jr., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., … Nosek, B. A. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science, 1*(4), 443–490. http://doi.org/10.1177/251524591881022

Kühnen, U. (2010). Manipulation checks as manipulation: Another look at the ease-of-retrieval heuristic. *Personality and Social Psychology Bulletin, 36*, 47–58. http://doi.org/10.1177/0146167209346746

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology, 47*(6), 1231–1243. https://doi.org/10.1037/0022-3514.47.6.1231

Luttrell, A., Petty, R. E., & Xu, M. (2017). Replicating and fixing failed replications: The case of need for cognition and argument quality. *Journal of Experimental Social Psychology, 69*, 178–183. https://doi.org/10.1016/j.jesp.2016.09.006

Muraven, M., & Baumeister, R. F. (1997). Suicide, sex, terror, paralysis, and other pitfalls of reductionist self-preservation theory. *Psychological Inquiry, 8*, 36–40. http://doi.org/10.1207/s15327965pli0801_7

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*, Article 6251. http://doi.org/10.1126/science.aac4716

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867–872. http://doi.org/10.1016/j.jesp.2009.03.009

Philpot, L. M., Khokhar, B. A., Roellinger, D. L., Ramar, P., & Ebbert, J. O. (2018). Time of day is associated with opioid prescribing for low back pain in primary care. *Journal of General Internal Medicine, 33*, 1828–1830. http://doi.org/10.1007/s11606-018-4521-8

Pyszczynski, T., Greenberg, J., & Solomon, S. (1997). Why do we need what we need? A terror management perspective on the roots of human social motivation. *Psychological Inquiry, 8*, 1–20.

Schmeichel, B. J., Gailliot, M. T., Filardo, E.-A., McGregor, I., Gitter, S., & Baumeister, R. F. (2009). Terror management theory and self-esteem revisited: The roles of implicit and explicit self-esteem in mortality salience effects. *Journal of*

Personality and Social Psychology, 96, 1077–1087. http://doi.org/10.1037/a0015091

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366. http://doi.org/10.1177/0956797611417632

Tice, D., Baumeister, R., Crawford, J., Allen, K.-A., & Percy, A. (2021). Student belongingness in higher education: Lessons for professors from the COVID–19 pandemic. *Journal of University Teaching and Learning Practice*, 18, Article 2.

Tierney, J., & Baumeister, R. F. (2019). *The power of bad: How the negativity effect rules us and how we can rule it*. Penguin.

Trinh, P., Hoover, D. R., & Sonnenberg, F. A. (2021). Time-of-day changes in physician clinical decision making: A retrospective study. *PLOS One*, 16(9), Article e0257500. http://doi.org/10.1371/journal.pone.0257500

Vohs, K. D., Schmeichel, B. J., Lohmann, S., Gronau, Q. F., Finley, A. J., Ainsworth, S. E., Alquist, J. L., Baker, M. D., Brizi, A., Bunyi, A., Butschek, G. J., Campbell, C., Capaldi, J., Cau, C., Chambers, H., Chatzisarantis, N. L. D., Christensen, W. J., Clay, S. L., Curtis, J., … Albarracín, D. (2021). A multi-site preregistered paradigmatic test of the ego depletion effect. *Psychological Science*, 32(10), 1566–158. http://doi.org/10.1177/0956797621989733

von Hippel, W. (2018). *The social leap*. HarperCollins.

Webb, M., & Tangney, J. (in press). *Too good to be true: Bots and other bad data from Mechanical Turk*. Perspectives on Psychological Science.

Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., & Albarracín, D. (2016). From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words [Supplemental material]. *Psychological Bulletin*, 142(5), 472–497. http://doi.org/10.1037/bul0000030.supp

Williams, L. E., & Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science*, 322(5901), 606–607. http://doi.org/10.1126/science.1162548