Promoting Replicable Sexual Science: A Methodological Review and Call for Metascience

John Kitchener Sakaluk, Ph.D.

University of Toronto Mississauga

Department of Psychology

3359 Mississauga Road, Deerfield Hall Rm. 4098, Mississauga, ON L5L 1C6

email: jksakaluk@gmail.com

Author Note

This work has been supported by a Social Science and Humanities Research Council (SSHRC) Postdoctoral Fellowship awarded to John Sakaluk.

Abstract

Concerns have increased within the medical and social science communities about the replicability of scientific findings, and subsequently, assessments of replicability and proposals for how it may be increased have become more common. Sexual scientists, however, with few exceptions, have yet to formally participate in the published discourses about replicability. In this commentary, I begin by highlighting how replicability is important for science in general, and then arguing that sexual science could be uniquely and negatively impacted without more direct involvement in the replicability movement from those within our field. I then briefly review several mechanisms through which replicability can be undermined in research, and some of the proposals for addressing these issues. I conclude by offering some ideas for how sexual scientists might begin to evaluate and improve the replicability of our field, and stress the need for sexual scientists to add their voices to the ongoing discussions about the problem of replicability of scientific findings.

Key words: *metascience; methods; open science; replication; sexual science*

Promoting Replicable Sexual Science: A Methodological Review and Call for Metascience

Hold up a mirror and ask yourself what you are capable of doing, and what you care about. Then take the initiative—don't wait for someone else to ask you to act.

--Sylvia Earle (National Geographic, 2011)

Introduction

Sexual science is multidisciplinary (Wiederman & Whitley, 2002), and draws strongly from researchers trained in social and medical sciences. Within the social and medical sciences, concerns have emerged about the *replicability* of research findings. That is, to what extent findings will reasonably approximate those from original studies, when the methods and analyses from original studies are followed exactly with new, but comparable samples (though see Crandall & Sherman, in press, on the value of conceptual replications, in lieu of exact replications).

Widespread concerns about replicability among scientists in these disciplines were surely caused by multiple factors, but Ionnidis's (2005) highly-touted and damning analysis of the replicability of scientific findings, high-profile revelations of scientific fraud, and an increasing public awareness of these replicability woes (e.g., Carey 2011), have surely contributed. As such, much ink has been spilt, and many pixels spent, by social and medical scientists offering explanations for the limited replicability of scientific findings, and making suggestions for how replicability might be improved.

As a social psychologist who studies human sexuality, I have been struck by the absence of sexual scientists formally participating in the published academic discourse surrounding replicability. I therefore intend this commentary to serve several aims. I will first argue for why the replicability of scientific findings is generally important. I will then argue that sexual

scientists have a unique vested interest in formally participating in discourses about replicability. Finally, I will review some of the literature on *metascience*—the process of examining and evaluating scientific practices scientifically—and describe a few of the domains of metascientific inquiry and how they aim to improve replicability. I will conclude by offering a few ways in which sexual scientists could begin examining and improving the replicability of their science.

Why is Replicability Important?

Asking the question, "Why is replicability important?" might seem underwhelming or trite to some, but an appreciation for metascience cannot be garnered without first ensuring there is an appreciation of replicability as a cornerstone of good science. Replicability is perhaps most intrinsically important to scientists because we are fundamentally curious beings. We have a tremendous epistemic need to understand the sexual world around us—to whatever extent such a sexual world "exists" —and research that is conducted in such a way that compromises replicability threatens our desires for understanding and knowledge accumulation.

Replicability is also important because sexual scientists highly value applied science; we conduct research not only to understand, but also to improve the sexual experiences of those around us. But applied research is doubly vulnerable to not being replicable; not only must the basic research findings, upon which a given intervention is based, be replicable, but the intervention research itself must also be replicable. If we are to maximize our discipline's ability to positively impact the world around us, we must therefore strive to ensure the replicability of our scientific findings.

Finally, replicability is important if for no other reason than because of the considerable resources scientists, participants, and the public, invest in the research process. Scientists and trainees invest uncountable hours designing studies, collecting and analyzing data, and

disseminating findings; each year, thousands of students and community members contribute their time and energy to participate in research; and funding for sexual science, though perhaps small by relative comparison, is still a considerable amount of money by any absolute standard. What is the point of such expenditures, if we are not "getting it right" with our research?

Why Should Sexual Scientists Care About the Replicability Discourse?

As I will soon discuss, the scientific literature discussing the issue of replicability and what might be done to improve it has exploded in the last decade. Sexual scientists have been largely absent from this discourse in the published record—Twitter exchanges and private discussions notwithstanding. Perhaps their absence is because many sexual scientists are not aware of the broader concerns about replicability, or because they are not interested in participating in these discussions. I believe, however, that sexual scientists have a unique vested interest in the outcome(s) of the replicability discourse, and for a couple of reasons, cannot afford to be sideline spectators as this discourse continues to shape scientific norms.

Sexual Science is Unique

As Wiederman and Whitley (2002) so aptly discussed, there is much that makes sexual science a unique domain of inquiry. We research topics that are personal, and to some, weird or even vulgar; we use methods of data collection that are intimately detailed (e.g., qualitative interviews), or invasive (e.g., physiological data collection of sexual arousal); and we recruit participants from diverse, and sometimes vulnerable populations. Given these qualities, I, for one, am greatly concerned about replicability concerns shaping scientific norms in a "one size fits all" fashion.

Consider, for example, the case of data sharing. Many metascience scholars promote open data sharing as one way to increase the replicability of findings (e.g., Nosek, Spies, &

Motyl, 2012; Simonsohn, 2013), and journals are beginning to encourage (e.g., *Association for Psychological Science* journals), and in other cases, require data sharing (e.g., the *Public Library of Science* journals). Subsequently, metascientific examinations of the prevalence of data sharing show that it is becoming more common than it used to be (Vanpaemel, Vermorgen, Deriemaecker, & Storms, 2015). And though much of sexual science data could—and probably should—be shared more openly, there are many instances where a norm of open data would be suboptimal. Investigations of vulnerable sexual populations in particular, would seem ill-suited to be subjected to a norm of open data, in order to protect the identities of participants who could be placed in very real harm if their identities were somehow revealed. This is but one example of the replicability discourse shifting scientific norms in a way that might be problematic for sexual scientists.

Some of the changing scientific norms will likely improve the replicability and overall quality of sexual science, but some will likely pose serious challenges to our valued paradigms for research. We, as sexual scientists, know the needs of our scientific community best, and we should therefore begin evaluating proposals for changing scientific norms from our disciplinary perspective(s), and doing so in the published scientific domain—or offering our own vision for replicable science. Those looking for examples to follow can consult papers by Finkel and colleagues (2015) and Campbell and colleagues (2014), who evaluate some replicability-related research practices from the unique perspective of relationship scientists.

Sexual Science is Vulnerable

Failures to replicate can take a number of different forms, and occur for numerous reasons: a specific effect may not hold across multiple studies (e.g., Shanks et al., 2015); a broader field may struggle with producing replicable findings (e.g., Open Science Collaboration,

2015); or in the most extreme of cases, effects may not replicate because researchers have deliberately engaged in fraud (see Simonsohn, 2013, for a discussion). And though sexual science likely struggles—knowingly or not—with each of these causes of unreplicable findings to a comparable extent, the fallout from each of these issues could be greater for sexual science.

It is my opinion, but one that I believe to be shared, that sexual science is one of the more—if not *the* most—politicized areas of research. As such, sexual science can be impacted by shifting political landscapes. Consider the muzzling of Canadian scientists that took place under the previous Conservative government (Renzetti, 2013), or the targeted removal of funding for political science research in the United States, by Republican politicians (American Psychological Association, 2013). These examples highlight the dramatic effects that the political systems in North America—and elsewhere—can have on the research process, censoring scientific communication, or in other cases, eliminating the funding necessary to conduct research in the first place.

Given the politicized nature of sexual science, I think that it would simply be too easy for an unfriendly government to use a replicability-related scandal to justify (albeit, speciously) limiting our ability to do the scientific work that we do, and must continue to do. Thus, I think that the sexual science community would be best served if we became more proactive, regarding replicability, as opposed to waiting for an embarrassing raison d'être to evaluate and promote the replicability of our field.

How Can Metascience Improve Replicability?

The field of metascience is generally one in which researchers turn the scientific method upon itself. Like any other field, metascience involves both descriptive explorations of how common particular scientific practices are, and correlational or quasi-experimental studies

(experiments are rarer) in order to understand how certain scientific practices are associated with particular scientific outcomes. Much of this metascience literature from social and medical science scholars in the last decade has revolved around examining various practices that are thought to promote or undermine the replicability of findings. I now turn to describing some of the more prominent replicability-related scientific practices highlighted in the metascience literature; I will cover both scientific practices that have been identified as problematic, as well as the changes proposed to address these problematic practices.

Rates of Replication

Perhaps some of the most fundamental questions to ask about replication in a given field is how often are replications attempted, and how how often are they 'successful'? Psychology has emerged as a disciplinary leader for confronting replication-related issues like these directly. Makel, Plucker, and Hegarty (2012), for example, reviewed all articles published since 1990 in the five psychology journals with the highest impact factors. Only 1.60% of these articles used the term "replication", and even fewer, upon closer examination, were actual replications. Further, and more concerning, a recent large-scale collaborative study on the replicability of psychology-related research findings found that only 36% to 47% of these findings successfully replicated, depending on which criterion of "success" was used (Open Science Collaboration, 2015).

In the face of the discouraging rates of replication and replication success, psychologists have developed a number of different methods to more strongly integrate replication into their research programs. Though the specifics of these methods vary in minor ways, they can largely be thought of falling into one of two types of approaches: large-scale replication collaborations, and individual lab-based replication. For the large-scale replication collaboration approach (e.g.,

Open Science Collaboration, 2012; Uhlmann et al., in press), collaborators agree in advance on what effect(s) are interesting and/or important to replicate, develop and execute the same study across their labs, and pool their data for analysis—sometimes meta-analyzing the replication results across different the laboratories that participated (see Fabrigar & Wegener, in press). In this way, the large-scale collaborative approach to replication helps to share some of the perceived costs of replication, as no one lab is solely responsible for spending the time, energy, and money to conduct research that might be thought of as less-than-cutting-edge.

Individual lab-based replication approaches (e.g., Sakaluk, in press; Stanton & Campbell, in press), alternatively, attempt to shift the burden of conducting replication studies to the researchers who are studying the original effect in the first place. With this approach, after discovering an effect they are interested in, researchers are compelled to either collect an additional superficially large sample, or multiple and more modest samples, in order to evaluate the robustness of their original effect. With the results of all samples synthesized in a single article, this approach would therefore help to increase the confidence a researcher could have in the replicability of an effect described in a particular research article.

Availability of Data and Study Materials

Calls for the open sharing of data and research materials—sometimes collectively described as *open science* methods—have been among the most widely discussed metascientific proposals for increasing replicability (see Simonsohn, 2013, for an example). Open sharing of data can help to promote replicability in a number of ways. For example, researchers will inevitably, on occasion, make mistakes in the course of data analysis (e.g., see 2015 corrigendum regarding analyses by Greitemeyer & Mügge, 2014) or scientific reporting (Bakker & Wicherts, 2011; Nuijten, Hartgerink, van Assen, Epskamp, & Wicherts, 2015); open access to data can

therefore greatly facilitate the detection and correction of such errors. And in extreme cases, open access to data can play a vital role in the detection and prevention of deliberately fraudulent research (see Simonsohn, 2013).

Open data sharing in the social sciences has typically been rare. Wicherts, Borsboom, Kats, and Molenaar (2006), for example, requested datasets for reanalysis from 141 articles published in journals of the American Psychological Association (APA). Importantly, authors of APA-published articles sign an agreement indicating, amongst other ethical principles, that they will share data for the purpose of reanalysis—their response rate may therefore be thought of as a reasonable "upper bound" of the rate of data sharing. After waiting for six months, and offering an assortment of reassurances attesting to their legitimacy (e.g., REB approval, researcher CVs, signed agreements to not share data with others), Wicherts and colleagues (2006) received only 25% of the datasets they requested. Further analysis of their response rates to data requests revealed that authors were less likely to share their data when their article contained more errors of reported statistical values (Wicherts, Bakker, & Molenaar, 2011). Nevertheless, in recent years, an increasing number of journals have begun to encourage (e.g., Association for Psychological Science journals, like *Psychological Science*) or require (e.g., Public Library of Science journals, like PLoS ONE) open sharing of data, sometimes attempting to incentivize the practice (e.g., via "badges" on articles acknowledging open science practices). A follow up to Wicherts and colleagues' investigation showed that open data sharing, though far from ubiquitous, is becoming more common (Vanpaemel et al., 2015).

Open sharing of research materials, alternatively, helps to facilitate replication by ensuring that researchers can use the exact same procedures and measures when studying an effect discovered by somebody else. Indeed, sometimes the pragmatics of the publication

process—like page limits—prevents methodological details from being reported in full. An ad hoc measure of a particular construct, for example, may not be fully described, or important, but subtle details of a novel experimental manipulation may be omitted from a scientific report. And should researchers fail to replicate or extend an effect using these incompletely-described materials, it begs the question of whether the effect itself is not replicable, or alternatively, whether the researcher did not use the proper methodology to study it. As in the case of open sharing of data, some journals are now encouraging or requiring open sharing of all research materials (e.g., *Personality and Social Psychology Bulletin*), or at the very least, ensuring that word limits for articles permit complete reporting of methodological details (e.g., Research Articles and Research Reports in *Psychological Science*).

Statistical Practices

If open science initiatives have been the most widely discussed type metascientific proposals, then calls for reform to the ways that social and medical scientists have been the mostly heatedly debated. Metascience-based proposals for statistical reform span the range of relatively small-scale initiatives attempting to improve relatively straightforward statistical errors, to large-scale proposals for sweeping changes to the ways scientists carry out their statistical analysis, and everything in between. I have chosen to review three areas of statistics-related metascientific concern that span this range.

Misreporting of statistical values. Though scientists might disagree as to what extent the misreporting of statistical values in scientific reporting is a problem for replicability, I submit that most would generally agree that, all else being equal, accurately reported statistical values are better for replicable science than are misreported statistical values. Within psychology it appears as though roughly 50% of all articles reporting quantitative analyses contain a

misreported *p*-value, and somewhere between 10%-20% of articles contained at least one *grossly* misreported *p*-value that impact conclusions about statistical significance (Bakker & Wicherts, 2011; Nuijten et al., 2015; Veldkamp, Nuijten, Dominguez-Alvarez, van Assen, & Wicherts, 2014). These grossly misreported *p*-values are particularly troubling, as they may lead researchers to mistakenly invest their time, money, and human resources to further study effects that are not actually replicable.

As mentioned, greater unwillingness to share data appears to be associated with more frequently misreported statistical values (Wicherts et al., 2011). This finding must be interpreted carefully, however, because though it is intuitive that scientists who knowingly misreported statistical values would be less likely to share their data, it seems equally plausible that researchers who are more disorganized would be both more likely to make more statistical errors and less responsive to data sharing requests. Veldkamp and colleague (2014) hypothesized that misreporting might be less frequent in collaborations in which colleagues share data and jointly conduct statistical analyses, but their data did not support this association. Some researchers—myself included—have therefore proposed that routinely incorporating checks for misreporting into the peer-review process could be one way to dramatically reduce rates of misreported statistical values in the scientific literature (Nuijten et al., 2015; Sakaluk, Williams, & Biernat, 2014).

Use of best practices for statistical analysis. The evaluation and promotion of best practices for statistical analysis has been a broad area of metascientific study. Some, for example, have examined how a variety of statistical analyses are (mis)interpreted and reported in research (e.g., Belia, Fidler, Williams, & Cumming, 2005), and have offered guidelines to increase the transparency—and thereby replicability—of statistics-based reporting (e.g., Kashy,

Donnellan, Ackerman, & Russell, 2009). Others have stressed the need to put less emphasis on the outcomes of individual quantitative studies, which may be less replicable, and instead place greater emphasis on the role of meta-analysis in the development of cumulative knowledge (Chan & Arvey, 2012). And some, alternatively, have elected to focus on a particular types of statistical analysis (e.g., exploratory factor analysis), and have evaluated its use and provided best practices that minimize the probability of biased or otherwise unreplicable outcomes of these analyses (e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999).

Currently, however, metascientific debates about statistical best practices have centered more around much larger, far-reaching, and dramatic proposals for changes to the way that social and medical scientists analyze their data (e.g., the seemingly never-ending Frequentist vs. Bayesian paradigm debate; see Goodman, 1999a and 1999b). In particular, none seem to have generated so much recent debate as Cumming's (2014) "*The New Statistics*" proposal, which, among other more tempered and agreeable recommendations, strongly argues for the abandonment of null-hypothesis significance testing. It is important to note that metascientific proposals—and Cumming's (2014) in particular—can, and do shape editorial policies at scientific journals. *Basic and Applied Social Psychology*, for example, has banned the use of null-hypothesis significance testing (Trafimow & Marks, 2015), and its use is discouraged at *Psychological Science* (Eich, 2014). Still, while acknowledging that null-hypothesis significance testing may incentivize novelty over replicability (see Nosek et al., 2012), I remain somewhat skeptical about the grandiose advertised benefits of the larger propositions of *The New Statistics*, and see value in considering alternative approaches to replicable research (Sakaluk, in press).

Questionable research practices and "p-hacking". Among metascientific issues in general, and statistical ones specifically, the discussion of methods used to artificially cajole *p*-

values below a certain threshold (usually $p < .05$)—often referred to as "questionable research practices" (QRPs; John, Loewenstein, & Prelec, 2012) or "*p*-hacking" (Simmons, Nelson, & Simonsohn, 2011)—has been exceptionally heated, and sometimes, quite personal. John and colleagues (2012) developed a list of ten QRPs, and surveyed psychologists about which practices they had used in their own research. Many of these QRPs may seem relatively innocuous, such as deciding whether to collect more data after looking to see whether the results were significant, or not reporting all measures used in a study. Simulations by Simmons and colleagues, however, show QRPs like these, and others (e.g., using covariates) can produce impossible findings (see their real-data example of listening to "When I'm Sixty-Four" 'causing' participants to be younger in age) by unknowingly inflating Type I (i.e., false-positive) error rates well-beyond what is considered acceptable (for a typical $\alpha = .05$, inflated to $\alpha = .10$-$.61$). Moreover, John and colleagues' (2012) survey suggests that many researchers use these methods (most QRPs used by >35% of their respondents). As such, *p*-hacking may very well represent one of the most common causes of unreplicable science, though whether researchers who use these methods are generally aware of how they negatively impact the replicability of their research remains an empirical question. One straightforward proposed solution to QRPs/*p*-hacking is to encourage researchers against using these practices, and for researchers to subsequently disclose that they have not used them in the main text of their manuscripts. (see Simmons, Nelson, & Simonsohn, 2012, for a 21-word disclosure).

A more comprehensive alternative proposal is to combine open science methods along with preregistration of hypotheses, materials, and data analysis methods (see Stanton & Campbell, in press, for a laudable example). Using this approach, researchers make explicit their predictions, planned sample size, materials (predictors, mediators, moderators, and all dependent

variables), and analytic methods (e.g., type of analysis, interpretation of one vs. two-tailed significance tests, data transformations, covariates to be included, criteria for outlier removal, etc.,), and then save an un-editable version of these disclosures in a time-stamped online repository (the Open Science Framework provides effective preregistration capabilities). Then, the researchers could link to their preregistration in the course of publishing their research, allowing editors, reviewers, and readers to have maximum confidence that the presented results are not simply a product of QRPs/*p*-hacking (an approach currently incentivized at *Psychological Science*).

A final proposed solution to QRPs/*p*-hacking is to used statistical means of detecting their presence when evaluating research; two better-known approaches to doing so are the use of *p*-curve methodology (Simonsohn, Nelson, & Simmons, 2014a; Simonsohn, Nelson, & Simmons, 2014b; Simonsohn, Simmons, & Nelson, 2015), and the calculation and interpretation of a descriptive indexes of the likelihood of an effect to replicate, or that a given set of studies was *p*-hacked (see Schimmack, 2012, 2014, for examples).

Recommendations for Improving the Replicability of Sexual Science

The ways that replicability in a field can be undermined are numerous, and it may seem difficult to establish a deserving starting point for sexual scientists to begin making improvements to our field. I offer the following suggestions, for how this process might begin.

Collectively, sexual scientists should begin pursuing an initiative akin to the Open Science Collaboration (2015), in order to begin evaluating the general level of replicability for our field. Of course, such an initiative would first need to select a subgroup of effects to attempt to replicate. I would suggest selecting a study (or group of studies) that is highly cited, or particularly important, either theoretically (e.g., the malleability of gender differences in

sexuality; see Conley, Moors, Matsick, Ziegler, & Valentine, 2011 for a review) or practically (e.g., the existence of the sexual double standard; see Crawford & Popp, 2003; Muehlenhard, Sakaluk & Esterline, 2015, for reviews).

Among individual sexual scientists, alternatively, there are many ways that replicability could begin to be evaluated and improved. Sexual scientists, for example, could begin routinely incorporating direct replications of their new effects, into their manuscripts (see Sakaluk, in press; Stanton & Campbell, in press, for different approaches). Additionally, sexual scientists could begin moving towards more of an open science approach—sharing their materials and, when appropriate, their data via the Open Science Framework (http://osf.io) — or begin familiarizing themselves with and avoiding questionable research practices (John et al., 2012).

Most of all, however, I would encourage sexual scientists of any level of training to begin contributing to the published domain of replicability- and metascience-related discourses. We need to begin discussing and evaluating existing metascientific proposals that will affect our field, with our field's particular needs in mind (Wiederman & Whitley, 2002). And we should also begin taking stock of metascientific issues that may be more unique to sexual science; research exploring the extent of volunteer bias in sexual science (Strassberg & Lowe, 1995) and evaluating the harm posed to participants partaking in sexuality surveys (Yeater, Miller, Rinehart, & Nason, 2012) are excellent examples of this more sexual-science-specific form of metascience.

However, regardless of the forum in which one discusses replicability- and metascience-related issues, I would encourage sexual scientists to strive towards communication that is respectful, trusting, and understanding of their colleagues when doing so. In my opinion, the adoption of replication-friendly change in other disciplines—namely social psychology—has

been delayed as a result of some unproductive accusatory, disrespectful, and defense exchanges. Should we pursue a more replicable sexual science—and I hope that we will—we must resist the temptation to engage in "witch-hunts", and the pull towards a cynical view of each others' work. All have, at one point or another, likely engaged in suboptimal research practices that could have compromised the replicability of research findings—be it accidently misreporting a statistical value, employing a QRP, unawares of its impact on Type I (false-positive) error rates, or something else. We must therefore forgive ourselves and each other, and look more towards the future, as opposed to dwelling on the past. Each sexual scientist must be free to express agreement or concern about particular replicability-focused initiatives, without having their integrity immediately called into question; any replicability movement will not be broadly accepted, and therefore successful, otherwise.

Conclusion

The replicability woes of other medical and social science fields, as well as concerning examples of unreplicable findings close to our field (e.g., LaCour & Green, 2014; Shanks et al., 2015), should serve as a warning to sexual scientists. Replicability is important for good science, in general, and I have tried to present a case for why I believe it is especially important for sexual scientists to take a greater active role in contributing to the published record regarding replicability and how it may be improved. There are any number of ways we might begin improving the replicability of our field, including conducting more replications, using more open science methods, and changing some of ways we analyze and report data. Whichever we ultimately choose to pursue, we as a discipline should begin the process of openly discussing avenues to increasing replicability, and taking the initiative to implement them, rather than waiting until some scandal demands that we act. The result of doing so will be a more replicable

sexual science that can inspire greater trust in research findings, among both scientists and lay-public alike.

References

American Psychological Association. (2013). *Coburn amendment restricts NSF political science funding.* Retrieved from http://www.apa.org/science/about/psa/2013/04/political-science-funding.aspx

Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, *43*, 666-678.

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, *10*, 389-396.

Campbell, L., Loving, T. J., & Lebel, E. P. (2014). Enhancing transparency of the research process to increase accuracy of findings: A guide for relationship researchers. *Personal Relationships*, *21*, 531-545.

Carey, B. (November, 2011). *Fraud case seen as a red flag for psychology research*. Retrieved from http://www.nytimes.com.

Chan, M. E., & Arvey, R. D. (2012). Meta-analysis and the development of knowledge. *Perspectives on Psychological Science*, *7*, 79-92.

Conley, T. D., Moors, A. C., Matsick, J. L., Ziegler, A., & Valentine, B. A. (2012). Women, men, and the bedroom: Methodological and conceptual insights that narrow, reframe, and eliminate gender differences in sexuality. *Current Directions in Psychological Science*, *20*, 296-300.

Crandall, C. S., & Sherman, J. W. (in press). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*.

Crawford, M., & Popp, D. (2003). Sexual double standards: A review and methodological critique of two decades of research. *Journal of Sex Research*, *40*, 13-26.

Cumming, G. (2013). The new statistics: Why and how. *Psychological Science*, *25*, 7-29.

Eich, E. (2014). Business not as usual. *Psychological Science*, *27*, 3-6.

Fabrigar, L. R., & Wegener, D. T. (in press). Conceptualizing and evaluating the replication of research results. *Journal of Experimental Social Psychology*.

Fabrigar, L. R., & Wegener, D. T. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272-299.

Finkel, E. J., Eastwick, P. W., Reis, H. T. (2015). Best research practices in psychology: Illustrating epistemological and pragmatic considerations with the case of relationship science. *Journal of Personality and Social Psychology*, *108*, 275-297.

Goodman, S. N. (1999a). Toward evidence-based medical statistics. 1: The *p* value fallacy. *Annals of Internal Medicine, 130*, 995-1004.

Goodman, S. N. (1999b). Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine, 130*, 1005-1013.

Greitemeyer, T., & Mügge, D. O. (2014). Video games do affect social outcomes: A meta-analytic review of the effects of violent and prosocial video game play. *Personality and Social Psychology Bulletin*, *40*, 578-589.

Ionnidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*, e124.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth testing. *Psychological Science*, *23*, 524-532.

Kashy, D. A., Donnellan, M. B., Ackerman, R. A., & Russell, D. W. (2009). Reporting and interpreting research in *PSPB*: Practices, principles, and pragmatics. *Personality and Social Psychology Bulletin*, *35*, 1131-1142.

Makel, M. C., Plucker, J. A., Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, *7*, 537-542.

Muehlenhard, C. L., Sakaluk, J. K., & Esterline, K. M. (2015). Double standard. In P. Whelehan & A. Bolin (Eds.), *The International Encyclopedia of Human Sexuality* (Vol. 1, pp. 309-311). Hoboken, NJ: Wiley & Sons, Inc.

National Geographic. (2011). *Ocean Hero: Sylvia Earle.* Retrieved from http://ocean.nationalgeographic.com/take-action/ocean-hero-sylvia-earle/

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*, 615-631.

Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (in press). The prevalence of statistical reporting errors in psychology (1985-2013). *Behavior Research Methods.*

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, *7*, 657-660.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, 943.

R Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Renzetti, E. (2013). *Censorship is alive and well in Canada—just ask government scientists.* Retrieved from http://www.theglobeandmail.com/globe-debate/censorship-is-alive-and-well-in-canada-just-ask-government-scientists/article8996700/

Sakaluk, J. K. (in press). Exploring small, confirming big: An alternative system to the new statistics for advancing cumulative and replicable psychological research. *Journal of Experimental Social Psychology*.

Sakaluk, J. K., Williams, A. J., & Biernat, M. (2014). Analytic review as a solution to the misreporting of statistical results in psychological science. *Perspectives on Psychological Science*, *9*, 652-660.

Schimmack, U. (2014). *Quantifying statistical research integrity: The replication index*. Unpublished manuscript, retrieved from: http://www.r-index.org/uploads/3/5/6/7/3567479/introduction_to_the_r-index__14-12-01.pdf

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, *17*, 551-566.

Shanks, D. R., Vadillo, M. A., Riedel, B., Clymo, A., Govind, S., Hickin, N., Tamman, A. J., Puhlmann, L. M. (2015). Romance, risk, and replication: Can consumer choices and risk-taking be primed by mating motives? *Journal of Experimental Psychology: General*, *144*, e142-e158.

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

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Dialogue, The Official Newsletter for Personality and Social Psychology*, *26*, 4-7.

Simonsohn, U. (2013). Just post it: The lesson learned from two cases of fabricated data detected by statistics alone. *Psychological Science*, *24*, 1875-1888.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). P-curve: A key to the file drawer. *Journal of Experimental Psychology: General*, *143*, 534-547.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). P-curve and effect size correcting for publication bias using only significant results. *Perspectives on Psychological Science*, *9*, 666-681.

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better p-curves: Making p-curve analysis more robust to errors, fraud, and ambitious p-hacking, a reply to Ulrich and Miller (2015). *Journal of Experimental Psychology: General*, *144*, 1146-1152.

Stanton, S. C. E., & Campbell, L. (in press). Attachment avoidance and amends-making: A case of advocating for the need for attempting to replicate one's own work. *Journal of Experimental Social Psychology.*

Strassberg, D. S., & Lowe, K. (1995). Volunteer bias in sexuality research. *Archives of Sexual Behavior*, *24*, 369-382.

Trafimow, D., & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology*, *37*, 1-2.

Uhlmann, E. L., et al. (in press). The pipeline project: Pre-publication independent replications of a single laboratory's research pipeline. *Journal of Experimental Social Psychology*.

Vanpaemel, W., Vermorgen, M., Deriemaecker, L., & Storms, G. (2015). Are we wasting a good crisis? The availability of psychological research data after the storm. *Collabra*, *1*, 1-5.

Veldkamp, C. L. S., Nuijten, M. B., Dominguez-Alvarez, L., van Assen, M. A. L. M., & Wicherts, J. M. (2014). Statistical reporting errors and collaboration on statistical analyses in psychological science. *PLoS ONE*, *9*, e114876.

Vianello, M., Sommer, S. A., & Jordan, J. (in press). The pipeline project: Pre-publication independent replications of a single laboratory's research pipeline. *Journal of Experimental Social Psychology*.

Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of evidence and the quality of reporting of statistical results. *PLoS ONE*, *6*, e26828.

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of research data for reanalysis. *American Psychologist*, *61*, 726-728.

Wiederman, M. W., & Whitley, B. E. (2002). A preview: The unique nature of sexuality research. In M. W. Wiederman & B. E. Whitley (Eds.), *Handbook for conducting research on human sexuality* (pp. 1-6). Mahwah, NJ: Lawrence Erlbaum.

Yeater, E., Miller, G., Rineheart, J., & Nason, E. (2012). Trauma and sex surveys meet minimal risk standards. *Psychological Science*, *23*, 780-7