

## RESEARCH METHODS

# Experimenter gender and replicability in science

Colin D. Chapman,\* Christian Benedict, Helgi B. Schiöth

There is a replication crisis spreading through the annals of scientific inquiry. Although some work has been carried out to uncover the roots of this issue, much remains unanswered. With this in mind, this paper investigates how the gender of the experimenter may affect experimental findings. Clinical trials are regularly carried out without any report of the experimenter's gender and with dubious knowledge of its influence. Consequently, significant biases caused by the experimenter's gender may lead researchers to conclude that therapeutics or other interventions are either overtreating or undertreating a variety of conditions. Bearing this in mind, this policy paper emphasizes the importance of reporting and controlling for experimenter gender in future research. As backdrop, it explores what we know about the role of experimenter gender in influencing laboratory results, suggests possible mechanisms, and suggests future areas of inquiry.

## INTRODUCTION

Failure to replicate significant findings has become a recent concern across several disciplines of scientific inquiry. Some research groups report that attempts to replicate published data in biomedical science fail more often than they succeed, and a recent paper revealed that of 100 articles published in high-ranking psychology journals in 2008, only one-third to one-half of original findings were successfully replicated (1, 2). Here, we point to one important and overlooked factor likely perpetuating this ubiquitous problem: the role of experimenter gender. Experiments in humans are regularly carried out without any report of the experimenter's gender; however, there is a range of evidence supporting the influence of experimenter gender on a variety of psychological and physiological variables (3, 4).

Pioneering work into experimenter effects demonstrated that several aspects of the experimenter can have significant influence. Scientists such as Robert Rosenthal laid the groundwork for this understanding, revealing the importance of experimenter expectations in relation to participant performance and, among other things, the importance of experimenter gender (5). Since these initial investigations, the field has grown: From intelligence testing to pain sensitivity, participants demonstrate robust responses to manipulation of experimenter gender (6, 7). The range of effects is troubling because it is broad enough to influence many fields of scientific inquiry that are not accustomed to controlling for experimenter effects.

Variance in such prominent mental and physical variables could potentially encourage reporting of illusory effects in clinical biomedical trials, inducing potentially serious consequences for patient treatment. For instance, when testing the efficacy of antinociceptive drugs, males report less pain to nociceptive stimulation when supervised by a female experimenter, as demonstrated by Alabas *et al.* (8). If, when testing an antinociceptive drug, a disproportionate number of treatment trials with male participants are supervised by female experimenters, then this could result in overestimations of drug efficacy. Putting aside the possibility of false positives, false negatives could be holding back progress. If scientists have difficulty replicating findings because of excessive null results, then the resulting noise makes any broader analysis less conclusive and more likely to induce further inquiry and delays. For instance, the collaborators in the Open Science Collaboration unsuccessfully attempted to replicate the findings of Epley *et al.* (9). The original study had shown that lonely participants were more likely to restore their

sense of belonging through increased belief in supernatural agents and events (9). Meanwhile, the replication failed to find significance. Surprisingly, in both the original and the replicated studies, the authors failed to report experimenter gender.

Thus, this review aims to summarize a sampling of studies demonstrating the influence of experimenter gender in a plethora of contexts, to speculate about mechanisms, and to propose policy recommendations for improving experimenter gender reporting. To this aim, the paper examines—in successive sections—a sampling of the experimenter gender's established impacts on elements of mind, body, and behavior. Following this, the paper covers possible mechanisms and policy recommendations. Finally, the paper concludes by suggesting future areas of research to further reveal the extent of the biasing effects of experimenter's gender.

## IMPACT ON THE MIND

When an experimenter and participant interact, their genders influence a range of psychological and physical variables, in much the same way as when two friends or colleagues interact. Bearing this in mind, this paper highlights examples of experimenter gender bias within three broad categories of human research: mind, body, and behavior. These sections are further bracketed by areas of study that emphasize the range of experimenter gender effects.

## Intelligence

Before any interest was piqued as to the experimenter gender's role in biasing other measures, there was a wave of interest in its impact on higher-level cognitive functioning. In particular, scientists were curious about how an experimenter's gender could influence performance on intelligence testing. Early results suggested a variety of interactions. Studies in children revealed a significant effect of experiment gender on performance. Namely, female examiners appear to elicit higher full-scale intelligence quotient (IQ), verbal IQ, comprehension, similarities, and vocabulary scores on the Wechsler Intelligence Scale for Children for both boys and girls (10). These studies raise obvious concerns about the replicability of intelligence testing, but perhaps more alarming is the impact on the development of therapeutics to treat learning disabilities in children. Newer medications for attention deficit hyperactivity disorder may show results that are too favorable, or not significant enough, as a result of experimenter gender influence. Again, this could be holding back or delaying the development of newer, safer therapeutics for use in treating these conditions because more and

Copyright © 2018  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
NonCommercial  
License 4.0 (CC BY-NC).

Department of Neuroscience, Uppsala University, Uppsala, Sweden.  
\*Corresponding author. Email: colin.chapman@neuro.uu.se

more studies are run to determine whether a particular compound's effects are consistent. Worse still, it could halt investigations altogether if early results are unfavorable.

### Creativity

Additional studies have investigated the impact of experimenter gender on creative problem solving. In general, male experimenters have been shown to elicit more solutions in a creative problem solving task (Remote Associates Test) for both genders of participants (11). However, female participants were significantly more affected by the gender of the experimenter, whereas men were only marginally affected. In other words, male experimenters improved results for both genders but much more so for females. In addition, female experimenters reduced results but also much more so for females. The researchers concluded that females are generally more sensitive to and responsive to other people than males. However, this conclusion should be tempered by the cultural context and timing of the research.

### Learning and memory

One of the first studies looking at experimenter gender demonstrated that verbal learning was influenced, such that female participants learned significantly faster in a serial trigram task with a male experimenter as opposed to a female experimenter (5). Other studies have taken these findings further. Experiments using simple sorting tasks reveal that participants performed significantly better, regardless of gender, when tested by an opposite-gender experimenter (12). It was speculated that this could follow from opposite-gender dynamics increasing competitiveness, anxiety, or the desire to please. Making the picture more intricate, however, another study found that, on a complex verbal conditioning task, while, as expected, low-anxiety men performed significantly better when tested by a female experimenter, highly anxious men actually performed worse (13). The authors theorized that this may have been due to an overload of stress for the high-anxiety men. Thus, although, in general, results support the conclusion that opposite-gender experimenters improve performance on learning and intelligence-related tests, this conclusion must be tempered because qualities specific to the participant appear to also modulate this effect. Finally, some research has revealed that even fundamental memory processes are sensitive to experimenter gender. Men paired with a female experimenter tend to provide more elaborate verbal autobiographical memories, and women with a male experimenter report fewer "internal states" such as emotional or cognitive states (14).

Again, these studies are significant in light of the recent surge in development of therapeutics designed to treat conditions such as Alzheimer's disease and other forms of cognitive impairment associated with aging. In some cases the same cognitive tests that demonstrated experimenter gender biases are used to determine whether these cognitive-enhancing therapeutics are efficacious. Imagine an experiment being run without the gender of the experimenter being stringently controlled, where a female directs the treatment participants and a male directs the placebo participants. Imagine further that the participants themselves are male. This design could easily lead to an exaggerated treatment effect.

### Neurological factors

More recently, some experimenters have ventured into the territory of neurobiology, looking for the correlates that one might expect to the behavioral differences that experimenter gender elicits. Evidence indicates that defensiveness is related to relative left frontal activation (LFA)

in women and right frontal activation (RFA) in men, as measured by electroencephalogram (EEG). LFA has been associated with "behavioral approach," whereas RFA has been associated with "behavioral withdrawal." Researchers have found that when an opposite-gender experimenter is in the room, participants who are highly defensive show greater LFA activation, and participants who were not defensive showed greater RFA activation (15). This suggests that when self-presentation is primed via the presence of the opposite gender, different parts of the brain are stimulated depending on the personality of the participant. Presumably more defensive individuals have greater LFA activation in the presence of an opposite-gender experimenter because they use more approach-related strategies to cope with their defensive dispositions, whereas less defensive individuals gravitate toward avoidance strategies. Most significantly, this study points to neurological differences in the reaction of participants to experimenter gender, which seem most pronounced in an opposite-gender context, demonstrating the possibility of bias in other neurobiological studies that fail to account for such effects.

### IMPACT ON THE BODY

Mental differences in response to interaction with different genders are natural to assume because many people experience these personally. However, less intuitive are the possible effects of experimenter gender on bodily functioning. To date, more research has been done on psychological or mental traits; however, there appear to also be several physical effects, partly mediated by central mechanisms. In addition, not only physical performance but also underlying biomarkers and physiological systems appear to be influenced, again underlining the significance of this bias for clinical therapeutic trials.

### Physical performance

A small series of studies has investigated the impact of experimenter gender on physical performance, and, again, significant results were observed. In one study, the effect of experimenter gender was investigated for participants performing a 50-yard dash, a shuttle run, and sit-ups. The study demonstrated that, for sit-ups, male experimenters elicited better scores for both genders of participants (16). On the other hand, both the 50-yard dash and the shuttle run participants performed significantly better when paired with an opposite-gender experimenter, regardless of their own gender. However, other studies have demonstrated a lack of effect with regard to physical performance. One study, for example, investigating the impact of experimenter gender on performance on grip strength and hand steadiness tests found no interaction for either task (17). Thus, much like intelligence and learning, physical performance appears to generally be enhanced by opposite-gender experimenters, although there are some inconsistencies and null results.

### Testosterone

Where measurable physical performance is altered, one should of course expect biological systems underlying this to be modified as well. In particular, experiments reveal that—perhaps unsurprisingly—sex steroids such as testosterone are affected by experimenter gender, which, in turn, causes differences in physical performance. For instance, one study revealed that young male skateboarders take increased physical risks in the presence of an attractive female (18). This increased risk taking leads to not only more successes but also more crash landings in front of a female observer. Mediation analyses reveal that this effect is influenced in part by elevated testosterone levels in men who performed in front of the attractive female. In addition,

performance on a reversal-learning task predicted physical risk taking, and reversal-learning performance was also disrupted by the presence of the attractive female, and the female's presence moderated the observed relationship between risk taking and reversal learning. These data of course fit closely with earlier data suggesting an impact of experimenter gender on learning. Combined, these results suggest that men use physical risk taking as a sexual display strategy and that this may be moderated by elevated testosterone levels in the presence of a woman (be she an experimenter or otherwise).

Further evidence reveals not only that testosterone is selectively elevated in the presence of a female experimenter but also that it appears that this is quantifiable in perspiration. More specifically, men excrete higher levels of the sex steroids 17 $\beta$ -estradiol and testosterone when performing rigorous exercise in the presence of a female experimenter (19). In turn, these hormones are absorbed by the experimenter, surely having additional effects on the experimenter and his or her instructions and behavior. Combined, these papers reveal a critically important link: Experimenter gender affects hormonal substrates. The question of how far-reaching this is remains unanswered, but sex steroids could represent the tip of the iceberg. The implications for clinical therapeutics should be clear: There could be, for example, a huge biasing effect produced in estimates of the efficacy of testosterone boosting medications, if the tests are administered by females.

### Pain sensitivity

Starting in the 1990s, a growing body of literature on pain sensitivity revealed that experimenter gender was biasing results. Initial findings suggested that male participants demonstrate a significantly higher pain threshold (reporting significantly less pain) when tested by female experimenters (20). The same study found a trend toward women actually reporting higher pain when tested by a male experimenter, but this did not reach significance. Several years later, studies investigated the phenomenon of male participants demonstrating lower pain sensitivity when tested by females, and the early result has generally been supported (7, 21). A recent meta-analysis helps make sense of these findings. Alabas *et al.* analyzed 13 studies that looked at gender role and pain thresholds. The consensus finding was that participants who viewed themselves as more masculine and less sensitive to pain demonstrated higher pain thresholds and tolerance (8). Another study investigated whether these findings of reduced pain sensitivity for men with female experimenters were mirrored by alterations in autonomic pain response (as measured by heart rate variability and skin conductance levels). The study found that lower pain reports in male participants with female experimenters were not mediated by changes in autonomic parameters and the effect was thus likely more the result of psychosocial factors (22). For example, it could be that men in general tolerate higher levels of pain with a female experimenter as a function of their attempt to display higher degrees of masculinity.

### IMPACT ON BEHAVIOR

With the preceding sections, the cascade of mental and physical reactions to experimenter gender should reveal a system-wide effect on general functioning. That said, it should be unsurprising that behavior is also affected. Again, the extent of the effect is still understood only for a few dimensions of interpersonal interaction, but the results thus far provide fertile ground for future hypothesis testing. They also, unfortunately, create the same pervasive concern regarding study replicability for behavior-based research and interventions.

### Communication

A study investigating gender differences in the way marital couples interact with each other found a variety of somewhat predictable differences in nonverbal communication between men and women (such as the amount of smiling, laughing, and the average length of gazing at their spouse) (23). In addition, however, they found that some variables in both husbands and wives were dependent on the gender of the administering experimenter. In particular, husbands were more likely to speak first with a male experimenter, and discussions in general went on longer with a female experimenter present. The neurological evidence suggesting differences in the brains of men and women in targets such as Broca's area (known for its critical role in communicative behavior) suggest that there may be a plethora of other biasing effects of experimenter gender on variables that relate to communication; however, this remains largely uncharted territory. These data also relate back to memory performance, where, again, an effect of experimenter gender on verbal elaboration was discovered, which can be concerning in the context of Alzheimer's treatment research, for instance.

### Aggression

Several meta-analyses have revealed that males tend more toward physical aggression (24, 25). Conversely, females favor verbal or "relational" aggression (24). However, the gender of the experimenter appears to modulate these general trends. For instance, an early study revealed that, in male college age participants, female experimenters inhibited physical aggression in both genders of participants, whereas male experimenters potentiated it (25). However, another study demonstrated that the interaction is possibly more complex. Males in the presence of a male experimenter inhibited retaliatory aggression against a female "participant" (a study confederate) who had only mildly disagreed with them, but when the female confederate "participant" strongly disagreed, men tended toward more severe retaliatory insults (verbal aggression) and higher-intensity shocks (again, specifically in the presence of a male experimenter) (26). Similarly, men in the presence of a female experimenter showed higher levels of physical aggression against a male provocateur (also a confederate). The commonality appears to be that men will show more aggression when they are insulted or aggressed upon in the presence of both genders simultaneously, be they other participants, confederates, or experimenters. This is suggestive of dependence of experimenter gender-based effects on social context as well.

### Prosociality

Trust and reciprocity research has gained a lot of traction recently, and a wave of increased interest has sprung fresh studies of human morality. In these studies, manipulating experimenter gender again revealed a robust impact on behavior, such that in the presence of a female experimenter, participants playing a trust game showed more trust and reciprocity (27). This is of particular interest in the light of recent issues replicating the links between oxytocin and trust. In a seminal study, Kosfeld *et al.* (28) seemingly revealed that intranasal oxytocin potently modulates trust behavior in the trust game. However, a host of newer research has shown profound difficulty in replicating these findings, using very similar methodology (29). One might wonder what characteristics the experimenters administering the task had in Kosfeld *et al.*—was a woman administering the treatment condition?

### Sexual behavior

Perhaps the most obvious domain for a biasing effect of experimenter gender is in the study of sex itself. This effect has been found, for example,

in questionnaires relating to sexual experience. In one study, male college students—who were primed with information about how women were becoming more sexually permissive—reported inflated numbers of sexual partners as compared to when they received no priming, but only when the questionnaire was administered by a female (30). The experimenters hypothesized that this was due to either a defensive reaction or a desire to perpetuate hegemonic masculinity. They supported this theory with the evidence that the significant results appeared to stem from the study participants who scored high on tests of hypermasculinity and ambivalent sexism.

Beyond questionnaires, experimenter gender can affect a participant's response to a variety of situations that implicate sexuality or sexual behavior. Early research into the impact of experimenter gender on sexual behavior found that both the gender and attractiveness of the experimenter could significantly influence experimentally induced sexual fantasies (31). In detail, an attractive female experimenter was shown to unsurprisingly promote sexual fantasies in heterosexual male participants in much the same way as other conditions that used different, more explicit stimuli. A later study revealed that experimenter gender could affect a participant's response to sexually explicit material. In detail, the study found that females who had an "informal" male experimenter felt more anxious after viewing sexually explicit material, whereas males who had an "informal" female experimenter rated the attractiveness of the sexually explicit material significantly higher. Thus, the study argues that experimenter gender may produce either a restraining or a permissive context, which, in turn, can account for a significant portion of the variance of a participant's response to sexual material (32). Consider, in this context, medications that could produce sexual dysfunctions as a side effect, such as exist for many antidepressants. It should be clear from the research pattern that if these studies are investigated using female experimenters and male participants, reporting of sexual dysfunction may be significantly underreported.

## WHY THE DIFFERENCES?

### Opposite-gender dynamics

There are a variety of possible reasons why men and women respond differently to experimenters of the same or opposite gender. One hypothesis focuses on the role of psychosocial stress in intergender scenarios. For heterosexuals, opposite-gender encounters can mediate social rewards that same-gender encounters cannot (33). The theory is that favorable perception by the opposite gender can result in romantic, sexual, or marital relationships, all of which have the potential to confer reward (33). In addition, when a person makes a favorable impression on another, it can result in self-affirming feedback that they are socially and sexually attractive. Although unquestionably valuable, this feedback generally cannot be obtained from same-gender interactions (again, for the sake of simplicity, we refer here only to heterosexuals). Supporting this line of reasoning, a study using daily interaction records from college students demonstrated that they tended to be more concerned with conveying an impression of being likeable, competent, ethical, and attractive when interacting with those of the opposite sex (33). Further studies on the interaction of a perceiver and a target individual have revealed that the more socially desirable rewards a perceiver controls, the more likely target individuals will attempt to create a favorable impression. Furthermore, the apparent value structure of a perceiver can influence a target's aggression, reward allocation, and helping behavior.

Thus, opposite-gender experimenters might, in general (again, principally in the case of heterosexuals—the effect should be the reverse for

homosexual participants), elicit improved responses on a variety of measures related to general mating "fitness," including the observed improvements in physical fitness, learning and intellectual abilities, and further alterations in beliefs and social behavior relating to aggression and altruism. Even alterations in pain sensitivity observed in male participants with a female observer could be explained by this phenomenon because experienced pain may not in fact differ, with male participants instead simply reporting less pain to produce a positive impression.

In this line of thought, it is important to recognize that it is not "opposite gender" that is significant per se but likely the psychosocial stress that often results from this scenario and the heightened reward potential, which, in aggregate, creates this trend. Theoretically, this could be manipulated by other circumstances, such as increasing the number of experimenters, manipulating their age and their professional status, and so on. In addition, this interpretation of results suggests that certain research areas will prove more vulnerable. Experimenter gender should have the greatest impact in areas of study where participants are in frequent and close contact with experimenters. In addition, experiments implicating characteristics important for mate selection—such as mental acuity, physical prowess, or morality—may be more influenced.

### Psychosocial stress

Further evidence from studies of stress support this general conceptualization of the experimenter gender effect and add an additional layer. Stress is regulated in the body through two primary pathways—the hypothalamic-pituitary-adrenal (HPA) axis and the sympathetic-adrenal-medullary axis. These systems work to increase the body's vigilance in response to a stressor by increasing circulating levels of stress-regulating hormones such as glucocorticoids, epinephrine, and norepinephrine. The HPA axis in particular is especially sensitive to nonphysical stressors involving a social context, and its activation is therefore considered a strong indicator of exposure to psychosocial stress (34). There are a variety of paradigms commonly used in experimental settings to induce a stress response in participants. One of the most popular is the Trier Social Stress Test (TSST), which requires participants to present a free speech in front of a panel of "experts" (experimenters in laboratory coats) and afterward to perform a mental arithmetic challenge (35). Another, the Maastricht Acute Stress Test (MAST), also involves social evaluation but is less time- and resource-intensive than the TSST. Recent evidence indicates that the experimenter's gender can influence the results of such tests. For example, males tested by a female experimenter in the MAST demonstrated higher systolic blood pressure, whereas females tested by a male experimenter in the TSST showed higher subjective stress ratings (35). Stress can improve or degrade both physical and intellectual performance, depending on the degree. Thus, opposite-gender dynamics generate performance-enhancing effects through moderate increases in stress, whereas individuals who have high basal anxiety levels may actually perform worse under such circumstances, as discussed previously (12). Similarly, as discussed previously, these effects should be most pronounced in heterosexuals; other sexual orientations likely produce different effect patterns.

### POLICY RECOMMENDATIONS

To improve the prevalence of experimenter gender reporting, first and foremost, individual scientists must take upon themselves the task of tracking and reporting their experimenter and/or research assistant genders going forward. Furthermore, where appropriate, statistical analysis should test for experimenter gender effects. Research group

leaders have the strongest influence in this sense; however, it is ultimately each scientist's personal obligation to maintain reporting standards.

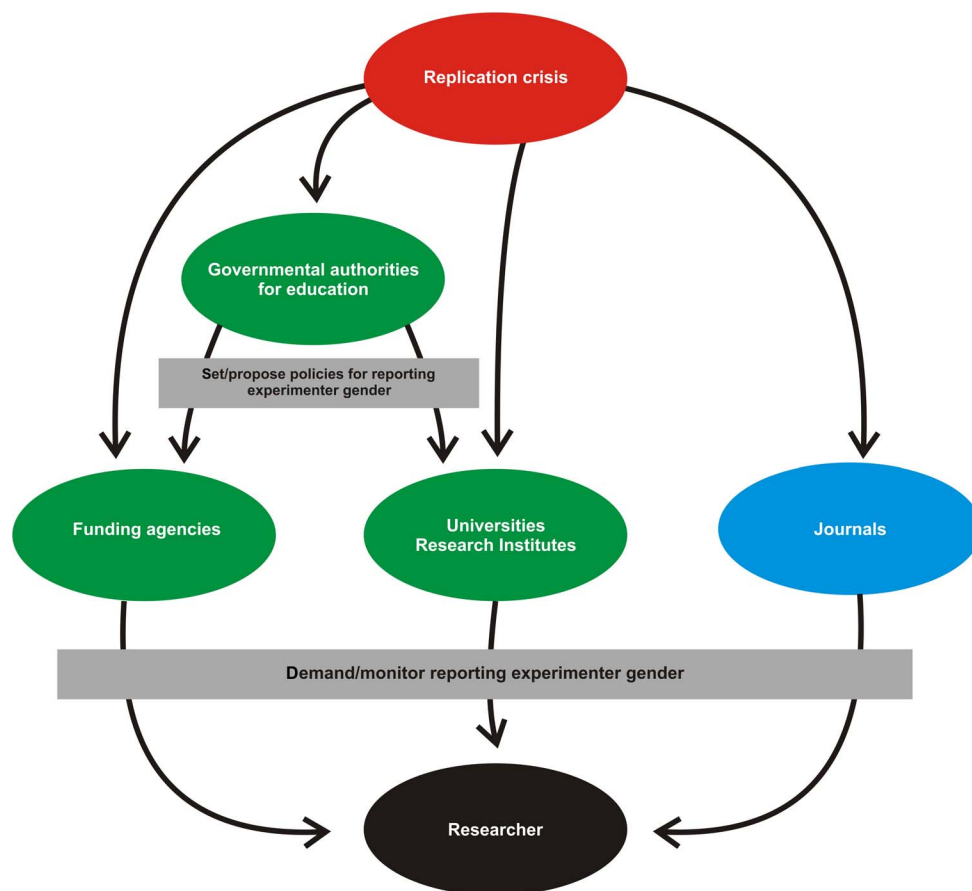
Looking speculatively toward the future of policy changes intended to improve replicability, there are several key players involved in the process that could promote change. For instance, universities or research institutes could take a top-down approach to the issue: It is not uncommon for universities to disseminate policy changes directly to laboratories under their umbrella. Because of the weight that universities have in setting the trajectory of individual scientists and ethical scientific standards, any guidance from them to report experimenter gender could be impactful.

Similarly, funding institutions could play a role. Every researcher is dependent on grants for survival—this gives grant issuers and private industry (such as the pharmaceutical industry) immense influence over research policy. Funding sources such as these could hypothetically augment their policies with a requirement for reporting experimenter gender. Stepping further back in the chain of influencers, governmental authorities could be the most significant potential influencer. Departments of higher education the world around are responsible for significant amounts of funding, both directly to universities and research institutes and indirectly through third-party organizations. Similarly, other divisions of government have large research budgets—for instance, in the United States, the Department of Agriculture alone budgets approximately \$1.8 billion to nutritional research (36). Finally, governmental regulators can influence policy with regard to private in-

dustry funding sources. Thus, with the significant amount of funding and influence that governments project toward the sciences, they are well positioned to assist in improving scientific standards.

Finally, journals, while relatively independent of the other key figures in this system, also have a powerful voice. If grants are necessary for survival, then so too are journal publications, which grant issuers evaluate critically in determining how to appropriate funds. Researchers are thus obliged to comply with any policy a journal sets out. The interplay between these various institutions and a roadmap for potential policy change is shown in Fig. 1.

Finally, it is worth addressing why reporting experimenter gender is an excellent jumping-off point in improving replicability. There are many other characteristics of experimenters that have demonstrable impact on participant performance, including age, height, and personality. However, gender has the unique qualities of being both (i) easy to record and report and (ii) categorical. Consider age for instance. Although it is similarly easy to record and report, it is much more difficult to interpret because it is not categorical. In other words, the extent of age differences (on a case-by-case basis) could create subtle differences that are not well understood. Gender, on the other hand, is both easy to record and report and relatively straightforward to demarcate. Thus, although controlling for more variables related to both the gender and the laboratory environment more broadly would be valuable to improving replicability, these changes would require significantly more



**Fig. 1. Flowchart identifying the key players responsible for policy changes within science.** As shown, the initiation of a crisis can induce change through several mechanisms. Prominent among these are changes in policy recommendations from government funding sources, in addition to policy changes at journals, universities, and independent funding agencies.

time, attention, thought, and resources to initiate. Finally, although this paper does argue that experimenter gender should be controlled and reported, this does not imply that every study should use an equal balance of male and female experimenters because this is similarly resource-intensive. Laboratories simply have a duty to report what gender their experimenters are, not to alter their staffing.

### THE FUTURE OF THE EXPERIMENTER GENDER EFFECT

Studies investigating psychometric variables and the newer research looking at differences in pain sensitivity have been instructive; however, a wide range of variables remain unexplored. To date, there is limited information on how biological and neurological measures are affected, such as genes, circulating hormones, neuropeptides, or brain activity as captured by functional magnetic resonance imaging (fMRI) or EEG. Where there are differences in psychological responses, there should be corresponding differences in neurobiology. For example, if participants who work with an opposite-gender experimenter are more likely to seek social reward through conveying a positive impression, then there are likely changes in their neurological responses. Studies have revealed that not only the acquisition of social reward but also the mere anticipation of it increases activity in mesolimbic brain structures (37, 38). Exposure to social reward also recruits a cohort of neuropeptides—for instance, in mice, the rewarding properties of social interaction have been shown to require the coordinated activity of oxytocin and serotonin (5-HT) in the nucleus accumbens. That said, opposite-gender experimenters are likely causing differential effects—through their impact on social reward processing—that could lead to significant differences in the results of fMRI responses and neuropeptide levels. There is a need to investigate differences that might appear in paradigms using EEG or fMRI or that look at circulating neuropeptide levels to determine where else there is systematic bias occurring.

Furthermore, there is good reason to believe that peripheral biological systems should be affected by changes in the central nervous system (CNS). A recent study in rats demonstrated that the animals' stress response was heightened in the presence of male experimenters (39). This stress response involves initial activation in the CNS, but via the HPA axis, activity proliferates to the periphery, and this pattern of effects is mirrored in humans. Thus, there is strong reason to believe that experimenter gender could be influencing a plethora of peripheral biological responses as well.

Some have recently suggested the concept of a “virtual experimenter.” The idea is to create a computer program that delivers treatment and instructions, which should theoretically increase standardization and reduce biasing effects and noise, such as those that come from experimenter gender (40). This standardized avatar would likely produce several advantages—in addition to controlling gender, other biasing influences such as personality, behavior, physical size, and, in general, human errors would be eliminated as confounders. However, the technology to support this becoming a ubiquitous and fail-safe tool could take some time to develop. Meanwhile, scientists can improve their own standards and practices to combat the issue.

### CONCLUSION

As this paper suggests, there is ample evidence, accumulated over decades of exploration, demonstrating that the gender of an experimenter has significant effects on a range of variables. It is also clear that the variables thus far investigated have been largely behavioral or psycho-

logical in nature, whereas biological and neurological responses remain largely unexplored. Given the strong connection between psychological and behavioral responses on the one hand and biological and neurological responses on the other, it stands to reason that this biasing effect should be similarly prevalent in these realms of study. It is common practice for studies in the fields of biology and neuroscience to not report experimenter gender, and yet, there is reason to believe that it could be significantly affecting results, including those of clinical trials. Note that research assistant positions are increasingly held by women, which could also potentially contribute to these replication issues. Combating the issue will be most effective if the major institutions of science—journals, funding sources, government, and universities—work in concert with individual scientists to encourage improved reporting standards. If these efforts are successful, then it could help clarify conflicting results in many subdisciplines and make sense of otherwise unusual data sets. It could pave the way for science to be more empirical, reduce noise in findings, increase the power of study designs, and generally improve the quality of scientific inquiry in these areas. With any luck, it will also aid in rebuilding the credibility of science by improving replicability.

### REFERENCES AND NOTES

1. F. Prinz, T. Schlange, K. Asadullah, Believe it or not: How much can we rely on published data on potential drug targets? *Nat. Rev. Drug Discov.* **10**, 712 (2011).
2. Open Science Collaboration, Estimating the reproducibility of psychological science. *Science* **349**, aac4716 (2015).
3. D. K. Rumenik, D. R. Capasso, C. Hendrick, Experimenter sex effects in behavioral research. *Psychol. Bull.* **84**, 852–877 (1977).
4. L. E. Carter, D. W. McNeil, K. E. Vowles, J. T. Sorrell, C. L. Turk, B. J. Ries, D. R. Hopko, Effects of emotion on pain reports, tolerance and physiology. *Pain Res. Manag.* **7**, 21–30 (2002).
5. R. Rosenthal, L. Jacobson, Teachers' expectancies: Determinates of pupils' IQ gains. *Psychol. Rep.* **19**, 115–118 (1966).
6. E. J. Archer, J. A. Cejka, C. P. Thompson, Serial-trigram learning as a function of differential meaningfulness and sex of subjects and experimenters. *Can. J. Psychol.* **15**, 148–153 (1961).
7. I. Kállai, A. Barke, U. Voss, The effects of experimenter characteristics on pain reports in women and men. *Pain* **112**, 142–147 (2004).
8. O. A. Alabas, O. A. Tashani, G. Tabasam, M. I. Johnson, Gender role affects experimental pain responses: A systematic review with meta-analysis. *Eur. J. Pain* **16**, 1211–1223 (2012).
9. N. Epley, S. Akalis, A. Waytz, J. T. Cacioppo, Creating social connection through inferential reproduction: Loneliness and perceived agency in gadgets, Gods and greyhounds. *Psychol. Sci.* **19**, 114–120 (2008).
10. R. Back, R. H. Dana, Examiner sex bias and Wechsler Intelligence Scale for Children scores. *J. Consult. Clin. Psychol.* **45**, 500 (1977).
11. M. Gall, G. A. Mendelsohn, Effect of facilitating techniques and subject-experimenter interaction on creative problem solving. *J. Pers. Soc. Psychol.* **5**, 211–216 (1967).
12. H. W. Stevenson, S. Allen, Adult performance as a function of sex of experimenter and sex of subject. *J. Abnorm. Soc. Psychol.* **68**, 214–216 (1964).
13. C. F. Etaugh, Experimenter and subject variables in verbal conditioning. *Psychol. Rep.* **25**, 575–580 (1969).
14. A. Grysman, A. Denney, Gender, experimenter gender and medium of report influence the content of autobiographical memory report. *Memory* **25**, 132–145 (2017).
15. J. P. Kline, G. C. Blackhart, T. E. Joiner, Sex, lies, and electrode caps: An interpersonal context for defensiveness and anterior electroencephalographic asymmetry. *Personal. Individ. Differ.* **33**, 459–478 (2002).
16. R. Rickli, Physical performance scores as a function of experimenter sex and experimenter bias. *Res. Q.* **47**, 776–782 (1976).
17. R. Rickli, Effects of experimenter expectancy set and experimenter sex upon grip strength and hand steadiness scores. *Res. Q.* **45**, 416–423 (1974).
18. R. Ronay, W. von Hippel, The presence of an attractive woman elevates testosterone and physical risk taking in young men. *Soc. Psychol. Pers. Sci.* **1**, 57–64 (2010).
19. B. Elliott, C. Muir, D. deCatanzaro, Sources of variance within and among young men in concentrations of 17 $\beta$ -estradiol and testosterone in axillary perspiration. *Physiol. Behav.* **173**, 23–29 (2017).
20. F. M. Levine, L. L. De Simone, The effects of experimenter gender on pain report in male and female subjects. *Pain* **44**, 69–72 (1991).

21. K. Gijbbers, K. F. Nicholson, Experimental pain thresholds influenced by sex of experimenter. *Percept. Mot. Skills* **101**, 803–807 (2005).
22. P. M. Aslaksen, I. N. Myrbakk, R. S. Høifødt, M. A. Flaten, The effect of experimenter gender on autonomic and subjective responses to pain stimuli. *Pain* **129**, 260–268 (2007).
23. C. Weisfeld, M. Stack, When I look into your eyes. *Psychol. Evol. Gender* **4**, 125–147 (2002).
24. J. Archer, Sex differences in aggression in real-world settings: A meta-analytic review. *Rev. Gen. Psychol.* **8**, 291–322 (2004).
25. J. S. Hyde, Gender similarities and differences. *Annu. Rev. Psychol.* **65**, 373–398 (2014).
26. R. J. Borden, Witnessed aggression: Influence of an observer's sex and values on aggressive responding. *J. Pers. Soc. Psychol.* **31**, 567–573 (1975).
27. G. L. Shope, T. E. Hedrick, R. G. Green, Physical/verbal aggression: Sex differences in style. *J. Pers.* **46**, 23–42 (1978).
28. M. Kosfeld, M. Heinrichs, P. J. Zak, U. Fischbacher, E. Fehr, Oxytocin increases trust in humans. *Nature* **435**, 673–676 (2005).
29. G. Nave, C. Camerer, M. McCullough, Does oxytocin increase trust in humans? A critical review of research. *Perspect. Psychol. Sci.* **10**, 772–789 (2015).
30. T. D. Fisher, Sex of experimenter and social norm effects on reports of sexual behavior in young men and women. *Arch. Sex. Behav.* **36**, 89–100 (2007).
31. R. A. Clark, The projective measurement of experimentally induced levels of sexual motivation. *J. Exp. Psychol.* **44**, 391–399 (1952).
32. P. R. Abramson, P. A. Golberg, D. L. Mosher, L. M. Abramson, M. Gottesdiener, Experimenter effects on response to explicitly sexual stimuli. *J. Res. Pers.* **9**, 136–146 (1975).
33. M. R. Leary, J. B. Nezelek, D. Downs, J. Radford-Davenport, J. Martin, Anne McMullen, Self-presentation in everyday interactions: Effects of target familiarity and gender composition. *J. Pers. Soc. Psychol.* **67**, 664–673 (1994).
34. L. Goff, N. Ali, J. Pruessner, Stress reactivity during evaluation by the opposite sex: Comparison of responses induced by different psychosocial stress tests. *McGill Sci. Undergrad. Res. J.* **8**, 11 (2013).
35. C. Kirschbaum, K.-M. Pirke, D. H. Hellhammer, The 'Trier Social Stress Test'—A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* **28**, 76–81 (1993).
36. S. Rowe, N. Alexander, F. Clydesdale, R. Applebaum, S. Atkinson, R. Black, J. Dwyer, E. Hentges, N. Higley, M. Lefevre, J. Lupton, S. Miller, D. Tancredi, C. Weaver, C. Woteki, E. Wedral, Funding food science and nutrition research: Financial conflicts and scientific integrity. *Nutr. Rev.* **67**, 264–272 (2009).
37. K. N. Spreckelmeyer, S. Krach, G. Kohls, L. Rademacher, A. Irmak, K. Konrad, T. Kircher, G. Gründer, Anticipation of monetary and social reward differently activates mesolimbic brain structures in men and women. *Soc. Cogn. Affect. Neurosci.* **4**, 158–165 (2009).
38. K. Izuma, D. N. Saito, N. Sadato, Processing of social and monetary rewards in the human striatum. *Neuron* **58**, 284–294 (2008).
39. R. E. Sorge, L. J. Martin, K. A. Isbester, S. G. Sotocinal, S. Rosen, A. H. Tuttle, J. S. Wieskopf, E. L. Acland, A. Dokova, B. Kadoura, P. Leger, J. C. S. Mapplebeck, M. McPhail, A. Delaney, G. Wigerblad, A. P. Schumann, T. Quinn, J. Frasnelli, C. I. Svensson, W. F. Sternberg, J. S. Mogil, Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nat. Methods* **6**, 629–632 (2014).
40. B. Horing, N. D. Newsome, P. Enck, S. V. Babu, E. R. Muth, A virtual experimenter to increase standardization for the investigation of placebo effects. *BMC Med. Res. Methodol.* **16**, 84 (2016).

#### Acknowledgments

**Funding:** This work was supported by the Swedish Research Council. The funding sources had no input in the design and conduct of this study; in the collection, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript. **Author contributions:** C.D.C., C.B., and H.B.S. contributed to the conceptualizations. C.D.C. drafted the manuscript. C.D.C. and C.B. created the figure. C.D.C., C.B., and H.B.S. edited the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the articles cited herein.

Submitted 2 May 2017

Accepted 5 December 2017

Published 10 January 2018

10.1126/sciadv.1701427

**Citation:** C. D. Chapman, C. Benedict, H. B. Schiöth, Experimenter gender and replicability in science. *Sci. Adv.* **4**, e1701427 (2018).