Replicability and the Psychology of Science

Cory J Clark

University of Pennsylvania

Nathan Honeycutt

Rutgers University

Lee Jussim

Rutgers University

This chapter is forthcoming and may be cited as:

Clark, C. J., Honeycutt, N., & Jussim, L. (in press). Replicability and the psychology of science. In S. Lilienfeld, A. Masuda, & W. O'Donohue (Eds.), *Questionable Research Practices in Psychology*. New York: Springer.

"*They all pose as though their real opinions had been discovered and attained through the self-evolving of a cold, pure, divinely indifferent dialectic… whereas, in fact, a prejudiced proposition, idea, or 'suggestion,' which is generally their heart's desire abstracted and refined, is defended by them with arguments sought out after the event. They are all advocates who do not wish to be regarded as such, generally astute defenders, also, of their prejudices, which they dub 'truths,'—and VERY far from having the conscience which bravely admits this to itself...*"

--Friedrich Nietzsche, Beyond Good and Evil

Scientists are humans. They are smart, ambitious humans, with a peculiar desire to explain and understand the world and a set of principles and procedures that help steer them toward truth. They are humans nonetheless. Their psychology is therefore human psychology.

Psychological discoveries in the social sciences—human errors, heuristics, biases, motivations, psychological needs—all apply to scientists in similar if not equal (or possibly even greater) measure. For example, people with greater education and science literacy are *more* polarized in their views of scientific controversies (such as climate change), raising the possibility that education increases the extent to which reasoning is influenced by preferred worldviews (Drummond & Fischoff, 2017).

Although such biases and errors of reasoning are frequently investigated *by* scientists, they are rarely applied *to* scientists to understand how the reasoning patterns discovered by scientists likely influence scientists' own reasoning and discoveries. The present chapter will apply psychological science to explain why, when, and how scholars engage in QRPs, advance dodgy or erroneous conclusions (sometimes for decades on end), and suppress accurate or useful information. Although certainly some scholars consciously and purposefully engage in fraud or data suppression, we suspect the vast majority of these non-optimal truth-seeking strategies occur outside of researchers' awareness in the sense that they genuinely believe their research practices are more optimal than they are in reality. We first review bases for concluding that

scientists are vulnerable to *motivated research.* Next, we argue that it is in the best interest of truth-seeking for scientists to acknowledge these tendencies in themselves and vigilantly and proactively defend against them. We also suggest some concrete correctives.

A Primer on Motivated Reasoning

Reasoning—the ways in which we approach and avoid information, evaluate information, and construct our attitudes and beliefs about information—is motivated (Kunda, 1990). Sometimes it is motivated by desires to reach the most accurate conclusion. This is the scientific ideal. Unfortunately, however, reasoning can also be motivated by desires to reach particular conclusions rather than truth. This can undercut scientific validity.

Imagine a trial in which a defendant was accused of robbing a locally-owned mini mart and there were numerous pieces of evidence to evaluate, including a slightly blurry surveillance video, an eyewitness who claims the robber was of similar height and physique as the defendant, and a suspiciously timed bank deposit from the defendant shortly after the robbery took place. The prosecution attorney would be motivated to view this as clear and conclusive evidence of the defendant's guilt, the defense attorney would be motivated to view this as ambiguous and circumstantial evidence, and the judge and the jury would be motivated to make the most accurate evaluation of the defendant's likely guilt. Although humans prefer to see themselves as the judge—carefully weighing evidence and coming to conclusions most consistent with the data, humans often reason more like the lawyers, evaluating evidence in ways that allow them to reach conclusions most beneficial to themselves (Ditto, Liu, et al., 2019; Ditto, Clark et al., 2019; Haidt, 2001).

Humans likely evolved to reason this way because accuracy is not always the most important goal for reproductive success (Clark, Liu, Winegard, & Ditto, 2019). Sometimes it is

more beneficial to persuade others of one's own greatness, to demonstrate commitment and value to one's social group, to avoid a possibly correct but risky or costly conclusion, to protect one's own reputation or the reputation of one's kin, to secure a mate, or to deceive an enemy than to be correct. In the social sciences, the consequences that flow from many research findings are so difficult to evaluate, that inaccuracies can go undetected for decades. Popularity (of a scientific finding) can produce citations, grants, awards, and, therefore, career success. By the time invalidities of highly popular findings are discovered, the scientists producing them will have had wonderful careers. Thus, the current academic system is plausibly described as incentivizing popularity more than accuracy.

In science, motivated reasoning, or rather, *motivated research*, happens when extraneous concerns beyond accuracy influence how scientists familiarize themselves with extant data, reach hypotheses, collect and analyze observations, come to conclusions, and report those conclusions to other scientists and the public. Researchers do not merely forward their own conclusions however; they are also the gatekeepers (the editors, the peer reviewers, the hiring committee members, the peer commentators, etc.) for their peers' research, and thus motivated research can also happen when concerns beyond accuracy influence how scientists accept, elevate, reject, and suppress the work of their peers or the very peers themselves. The replication crisis has focused largely on how scholars advance erroneous conclusions by producing unreplicable results, but scholars may also obstruct accurate conclusions or useful information, which is problematic for advancing knowledge in the social sciences.

The social sciences supply especially fertile ground for motivated research. Ambiguous, noisy, and difficult information environments increase the likelihood of motivated reasoning (Kopko, Bryner, Budziak, Devine, & Nawara, 2011; Munro, Lasane, & Leary, 2010;

Munro, Weih, & Tsai, 2010). Accuracy motivations decrease because one cannot know with much certainty which conclusions are accurate, and thus other motivations take their place (Clark & Winegard, 2020). Science, and perhaps especially social science, generally deals with these ambiguous, noisy, and difficult information environments. Most if not all social phenomena cannot be studied in a vacuum. There is rarely if ever one clear best methodological strategy for testing a social science question, and even when scientists discover seemingly robust and replicable data patterns, there are often numerous ways of interpreting those patterns. Meehl (1990, p. 196) captured this state of affairs beautifully: "...theories in 'soft areas' of psychology have a tendency to go through periods of initial enthusiasm leading to large amounts of empirical investigation with ambiguous over-all results."

For example, any time some negative parenting behavior correlates with some negative outcome for children, did the parenting behavior have any causal influence or is there simply a genetic confound (e.g., Maranges, Hasty, Maner, & Conway, 2020)? Any time scholars discover an association between negative stereotypes or implicit attitudes and negative outcomes for the groups those stereotypes or implicit attitudes are about, did the stereotypes or implicit attitudes have any causal force on those negative outcomes, or did the negative stereotypes and implicit attitudes exist because people are reasonably skilled at detecting existing patterns in the world (e.g., Hehman, Flake, & Calanchini, 2018; Payne et al, 2017; Reber, 1989)? Even best practices in social science require scholars to make numerous at least somewhat arbitrary decisions at each step of the research process, from generating hypotheses to drawing conclusions. These characteristics of the social sciences make it very difficult for an accuracy-motivated social scientist to reach correct conclusions and simultaneously make it very easy for a social scientist motivated by extraneous concerns to reach the conclusion they desire (Duarte et al., 2015;

Simmons, Nelson, & Simonsohn, 2011). Consequently, social sciences as a discipline are vulnerable to motivated research practices.

Beyond the ambiguous information environment problem, there is even more reason to believe that motivated reasoning creates unique obstacles for the social sciences. The investigators and the objects of investigation are one in the same thing—humans—and humans *care* about human things. It likely makes little difference to the average human whether flying squirrels are fluorescent or whether there is a maximum speed of light, but average humans *might* care if middle-aged men are sexually attracted to 15-year-old females, if altruism is "selfish," and if grandparents evolved to love their daughters' kids more than their sons' kids. It is likely impossible to eliminate human desires from an understanding of humankind, thus social scientists likely have more extraneous motivations influencing their work than scientists who deal with amoebae, polymers, quarks, or any non-human objects.

Moreover, *morality* is frequently tangled up in the social sciences, and *moral concerns* are powerful motivators of reasoning (Clark et al., 2019; Tetlock, Kristel, Elson, Green, & Lerner, 2000). Sometimes accurate conclusions in the social sciences might cause concerns about morally undesirable implications, and people and scholars may then wish to avoid, ignore, disparage, or censor this kind of information, even when it could plausibly be correct (Campbell & Kay, 2014; Clark, Winegard, & Farkas, 2020; Stewart-Williams, Thomas, Blackburn, & Chan, 2019; von Hippel & Buss, 2017; Winegard, Clark, Hasty, & Baumeister, 2018). For just one recent example, a paper by AlShebli, Makovi, and Rahwan (2020) collected a very large sample of mentor and protégé pairs in scientific collaborations and found evidence that female protégés with higher proportions of female mentors were less impactful later in their careers. After widespread outrage among the scientific community, on November 19th, 2020, the editors of the

journal released a statement, “Readers are alerted that this paper is subject to criticisms that are being considered by the editors. Those criticisms were targeted to the authors’ interpretation of their data that gender plays a role in the success of mentoring relationships between junior and senior researchers, *in a way that undermines the role of female mentors and mentees*…” (emphasis added). Although there are plenty of legitimate criticisms of this paper (as there are of probably every published article in the social sciences), the investigation by editors of the journal occurred because of concerns about *undermining* the role of female mentors and mentees. Thus, this investigation is explicitly *morally* motivated. And these moral concerns might cause suppression of a real pattern and exploration into the causes of this pattern.

We are not saying that moral concerns are *never* a legitimate reason to suppress research findings (that is a difficult debate). But, in many cases outrage mobs of academics bear a striking resemblance to a mob stirring up a moral panic, as they cause the suppression of data in the absence of evidence of harm or thoughtful consideration of alternatives (Stevens, Jussim & Honeycutt, in press). Moreover, scholars often assert themselves and their comrades as the authorities on such matters. Thus while their intentions may be noble, such suppression is often ochlocratic and advances the interests of a subset of outraged scholars to the detriment of knowledge accruement. Occasionally, empirical reality will lead scholars to arrive at conclusions that trigger our moral alarms, and because scientists are humans and evolved to minimize certain harms, occasionally they will wish to suppress accurate information by suppressing their own findings (Ziggerell, 2018) or creating obstacles for their peers’ findings (Stevens et al., in press).

Similarly, occasionally, scientists will discover false patterns that are morally desirable, or real patterns but then explain these with false but morally desirable explanations. Such erroneous patterns or erroneous explanations may persist in the psychological canon for years or

decades because they are morally desirable to scholars and thus few scholars will wish to challenge them (Jussim, Krosnick, Stevens, & Anglin, 2019). To give a couple of examples, the ideas that stereotype threat could explain certain group disparities (e.g., Jussim, Crawford, Anglin, Stevens, & Duarte, 2016) or that implicit bias could explain subtle but impactful prejudices against certain groups (Forscher et al., 2019) are arguably some of the most prominent social psychological findings of all time, yet the effects are weak to non-existent and there is little if any evidence of their importance in the real world (e.g., Clark & Winegard, 2020). It seems likely that these effects were overblown and proper scrutiny was decades delayed because the findings were morally and thus socially desirable by scientists. Many scholars would want to forward such results themselves and few would want to challenge them.

Because the social sciences deal most directly with problems and questions with significance to humans, social scientific conclusions are vulnerable to morally motivated data suppression and morally motivated data elevation. Being a purely accuracy-driven social scientist will occasionally require an unnatural detachment from normal human concerns and motivations.

Human Motivations

We discuss four human motivations that likely influence the ways in which scholars conduct their research. We also discuss how those motivations can produce severely biased scientific research literatures.

Status desires. Humans desire status and behave in ways that increase their chances of attaining status in social groups (e.g., Anderson & Kilduff, 2009). Scientists desire to attain status within their discipline—to be respected and admired by their peers—but also, given the relative status of scientists in society (Pew, 2020), they likely desire to use their roles as

scientists to gain high status among society at large. Becoming a social scientist requires relatively high investment in education and a relatively high workload to attain a tenure-track position at a research institution, and the job actually pays relatively little compared to other degrees that require similar amounts of education and time investment (e.g., medical doctors). Therefore, it is plausible that social scientists are *more* motivated by desires for status than even the average highly educated person. Thus it seems plausible that the social sciences likely attract the kinds of people who are *especially* incentivized by status attainment and especially likely to engage in research behaviors that would allow them to attain status.

Perhaps the chief way people attain status is by creating the appearance of providing benefits for others (e.g., Durkee, Lukaszewski, & Buss, 2020). Although actually providing such benefits is one route to creating this appearance, it is not necessary. One can engage in virtue signaling or moral grandstanding without actually doing much else and this can sometimes be very effective at persuading others that one is a force for moral good (Grubbs, Warmke, Tosi, James, & Campbell, 2019). Providing benefits to others is also not sufficient to increase status (e.g., if it is done in a matter where few notice).

Therefore, regardless of whether anyone actually benefits, creating the appearance of providing benefits is highly incentivized. This would create a motivation to produce information that can be perceived as *new* or *novel* (Baumeister, Maranges, & Vohs, 2018) or to produce "Wow Effects" (Jussim et al., 2016) with seemingly broad implications. It may take years or even decades to do the hard work to evaluate whether the findings are replicable and generalizable, and then to test them in the real world; and, at the end of that process, the entire enterprise may be found to be worthless (findings unreplicable) or trivial (replicable but only with effect sizes so small no one cares). There are few incentives to wait 15 years for such a

payoff; people have jobs, tenure, promotions, and grants to obtain; bestselling books to write, workshop fees to collect, and consulting fees to garner. Put differently, the incentives all line up to create the impression that one has benefited society *now*, not 15 years from now.

Many of the most overblown findings in the social sciences fit this analysis exquisitely well (e.g., stereotype threat, implicit bias, growth mindset, various kinds of priming). We now know these findings were overblown and, in some cases, seem to be entirely invalid. Stereotype threat, priming, and growth mindset have all been subject to a series of pre-registered failures to replicate and/or findings that the effects are plausibly viewed as trivially small (Doyen, Klein, Pichon, & Cleeremans, 2012; Finnegan & Corker, 2016; Flore, Mulder & Wicherts, 2018; Pashler, Rohrer & Harris, 2013; Sisk, Burgoyne, Sun, Butler & Macnamara). After almost 20 years of "implicit bias" and the Implicit Association Test (IAT) in particular being wildly overstated and oversold, in recent years, critical reviews have described the construct as "delusive," identified a slew of psychometric problems with the IAT, and shown that its predictive validity and ability to explain racial gaps is limited at best and possibly nonexistent (Blanton, Jaccard, Strauts, Mitchell & Tetlock, 2015; Corneille & Hutter, 2020; Forscher et al, 2019; Jussim, Carrem, Goldberg, Honeycutt & Stevens, in press; Oswald, Mitchell, Blanton, Jaccard & Tetlock, 2013; Schimmack, 2019).

Although this is not the place to review all the debunking of the last 5-10 years, one example should suffice. Blanton and colleagues (2009) characterized a slew of studies making strong claims about racial discrimination produced by implicit bias as measured by the IAT as actually providing weak evidence. In a response, (Jost et al, 2009) published a paper titled "The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should

ignore." In a recent review, we did a deep dive into those ten studies and found something quite startling: those ten studies supposedly refuting the "weak evidence" charge provided almost no evidence of racial discrimination (Jussim et al., in press). Put differently, there was little or no racial bias to be explained (whether by IAT scores or anything else). Indeed, most of the studies did not even address racial discrimination *at all.*

Despite the extraordinary enthusiasm for these "discoveries" (as evidenced by the massive number of papers that use the terms and measures and by the eminence and awards given to their promoters and acolytes), the fullness of time (combined with the eventual emergence of vigorous scientific skepticism) has shown them to be far less than they were cracked up to be. This may help explain why diversity and implicit bias trainings based on these (nearly) nonexistent or poorly understood measures and phenomena are rarely demonstrably effective (Paluck, Porat, Clark & Green, 2020). Thus, these phenomena are all exquisite examples of how scholarship can create the impression that *AMAZING! WORLD-CHANGING!* phenomena have been discovered that will benefit humanity, without actually providing any noticeable benefit to humanity, and at great human cost in wasted effort, grant dollars, and time spent in useless trainings.

Nonetheless, selling *AMAZING! WORLD-CHANGING!* findings to an unsuspecting public and insufficiently critical scientists has been highly rewarded with status, promotions, grants, and consulting fees. And, to be clear, although scientists love to point to others (such as the media) as the culprits in overselling their findings, it is usually the scientists themselves who bear primary responsibility (Mitchell, 2018; Sumner et al, 2016). Regardless, scholars are heavily incentivized to create the appearance that their findings lead to simple and easy-to-implement interventions that will change the world. Unfortunately, many social problems persist

in affluent societies precisely *because* they are extremely difficult or perhaps even impossible to fix, and so the demand for such interventions inevitably creates a low quality supply. Unlike behavioral genetics or personality psychology, social psychology delivers simple environmental manipulations that ostensibly can create desirable changes in human behavior. The desire for effects that create potential for interventions and behavior change may even explain why social psychology is such an attractive discipline to normal people (McPhetres, 2019), despite its many flaws and embarrassments over the past decade (e.g., Nosek et al., 2015).

Scholars can also achieve media and public attention by generating findings with significance to current events and hot topics and so are likely motivated to study such topics, and to forward results quickly when they do. In a society where science often does and should move quite slowly and hot topics often change rather rapidly, this could cause scholars to draw hasty conclusions in order to be timely in their research. Moreover, quick movement to publicize *AMAZING! WORLD-CHANGING!* findings (see Mitchell, 2018 for a review of the wild overselling of implicit bias after the publication of the first IAT paper, in 1998) makes it difficult for other scholars to check such findings before they reach the broader public.

Of course, status motives will also lead scholars to pursue *accuracy* in their work, for two reasons. First, more accurate information is more useful to other people, and thus accuracy is a direct route to status attainment, and second, being *inaccurate* (if detected) can be costly. Having one's own research fail to replicate, or worse, being caught for outright research fraud are huge blows to status, and so scientists should be motivated to both appear accurate and be accurate. Given new developments in Open Science, it has become easier for other scholars to detect errors and other suspicious research practices in their peers' work, and so the current cohort of

scholars should be *more* motivated to be accurate (or at least avoid certain types of errors) than the cohort existing a decade or more ago.

Open Science practices have made certain types of QRPs more difficult to get away with. For example, pre-registration makes it more difficult to HARK (Kerr, 1998) and cherrypick variables, conditions, and even entire studies. However, many papers still report studies that are not pre-registered leaving the door wide open to such practices. Furthermore, if studies provide narratively or theoretically "inconvenient" findings, they can still be file drawered. When acting as a reviewer, it is easy enough to suppress others' inconvenient findings or arguments -- simply come up with scientifically-plausible justifications for declaring the work to be sub-par.

Ostracization avoidance. Just as people wish to gain status within their social groups, they wish to avoid being ostracized (Ouwerkerk, Kerr, Gallucci, & Van Lange, 2015). People tend to punish those who violate group norms or generate costs to the social group (Bowles & Gintis, 2004). Scholars are likely motivated to avoid these punishments, and, therefore, avoid violating group norms.

The extreme politically liberal homogeneity among social scientists (Duarte et al., 2015; Langbert, 2018) renders the entrenchment of liberal norms -- such as support for parties, policies, candidates, and causes on the left, hostility to those on the right, and equalitarianism (the asssumption that, but for discrimination, all demographic groups would have identical outcomes) -- virtually inevitable (Clark & Winegard, 2020; Honeycutt & Jussim, 2020; Prentice, 2012). Thus, for either or both of two reasons, scientists should be motivated to avoid advancing scientific findings that challenge a liberal political agenda: (1) They share that agenda and do not wish to oppose it or (2) They correctly discern these norms and believe (probably correctly) that work challenging those norms will be more difficult to publish and fund than work that advances

those norms. For example, some research has found that liberals are described more positively than conservatives in social scientific research (Eitan et al., 2018), that conservative social scientists fear ostracization and that other social scientists openly report that they would discriminate against conservatives (Honeycutt & Freberg, 2017; Inbar & Lammers, 2012), and that more liberal ideology predicts working at more prestigious universities, even after controlling for academic productivity, suggesting that ideological conformity helps one advance in their career (Rothman, Lichter, & Nevitte, 2005).

Another recent paper that sought to explore the relationship between ideological slant of research and replicability identified almost no papers at all in their analysis that violated liberal values, suggesting that such papers rarely come into existence (Reinero et al., 2020). Similarly, Zigerell (2017) discovered 17 unpublished experiments with nationally representative samples finding either no anti-Black bias among White respondents and/or anti-White bias among Black respondents. Although we may never know exactly why those studies were never published, one possibility is that they would risk violating liberal equalitarian norms and would, therefore, either be seen as not worth publishing, or not worth the (expected extraordinary) effort, and concomitant risk of being ostracized, to do so.

Arguably, these dynamics—political skew, bias and intolerance towards individuals or ideas that conflict with mainstream liberal views—have a direct connection to censorious behaviors (Honeycutt & Jussim, under review). This connection isn't inevitable—bias doesn't automatically produce direct or indirect censorship. But when academic fields such as the social sciences become so heavily skewed, excluding ideas or data that conflict with the norms and worldview of the majority becomes an increasing threat to the validity of the scientific literature. This is not to say that scholarship can never be rejected—papers are routinely rejected for flaws

and weaknesses that have nothing to do with political content or motivations. But ideologically motivated rejection can often be dressed up as legitimate critique, often manifested in selective calls for rigor, illusions of bad science, or claims of harm and danger (Honeycutt & Jussim, under review). One obvious casualty is the suppression of otherwise legitimate scholarship (Stevens, Jussim, and Honeycutt, 2020).

Scholars are likely motivated to reject information that could be perceived as opposing a politically liberal agenda both in their own research and in their peers' research. And they are likely motivated to frame their findings in ways that misleadingly portray liberals in a favorable light when their findings could just as easily or more easily be framed in ways that portray conservatives in a favorable light. For example, Lilienfeld (2015) critiqued the framing and description of conservatives having a "negativity bias" or "motivated closed-mindedness" when the findings on sensitivity to threat could have just as easily been framed as a liberal "motivated blindness to danger." More recently, a paper by Baltiansky, Jost, and Craig (2020) chose to highlight that high system-justifiers (correlated with more conservative ideology) found jokes targeting low status groups to be funnier than low system-justifiers in their abstract, portraying conservatives as being insensitive toward low status groups. However, high system-justifiers found jokes targeting low and high status groups similarly funny, whereas low system-justifiers found jokes targeting low status groups to be less funny than those targeting high status groups (Pursur & Harper, 2020). One could interpret such findings as showing that conservatives treated low and high status groups with equal consideration, whereas liberals were particularly condescending toward low status groups by suggesting they need protection from jokes. Similarly, a recent paper by Brady, Wills, Burkart, Jost, and Van Bavel (2019), highlighted that "conservative elites (on Twitter) gained greater diffusion when using moral-emotional language

compared to liberal elites," portraying conservatives as vulnerable to emotional appeals. However, this effect was mainly driven by *joy-related* content, which was misleadingly labeled "moral emotion expression related to religion and patriotism" in the abstract.

Scholars likely know that to frame results in ways that portray conservatives more favorably than liberals would make the results more difficult to publish. So, the easier route to attaining publications (and status) and avoiding ostracization, is to create misleading characterizations of findings. Thus scholars who wish to avoid ostracization among overwhelmingly liberal social scientists will engage in motivated research to generate findings and frames palatable to their liberal peers. Academia operates as a social-reputational system, whereby one's success is highly contingent upon the favorable evaluations and references of others at all career stages: to obtain admission to graduate school, publish in peer-reviewed academic journals, obtain grants, get a job, or obtain tenure/promotions. As such, there are strong incentives for doing work and staking out positions that will garner social approval from peers, and often strong disincentives surrounding the expression of ideas that colleagues reject or vehemently disagree with (Honeycutt & Jussim, under review).

Social scientists are even more homogenous in a domain other than politics—every last one of them is a social scientist. Thus, social scientists should be motivated to avoid harming social scientists and the social scientific enterprise. The types of scholars who critique the field, for example, by suggesting the field is politically biased, or by accusing the field of shoddy methods and unreliable findings, are likely to be revered by some and loathed by others. In an effort to protect the field and their own reputations, some scholars (likely, especially older and more established scholars with more to lose) might seek to create obstacles for scholars who forward data and arguments that challenge the field. Many scholars would avoid criticizing the

field, the field's theories, and the field's prominent scholars, even if they believe such criticisms are warranted, because it can be costly to them by virtue of incurring the hostility of colleagues on whom their success depends (via peer review). By writing this chapter in which we suggest that social scientists engage in *motivated research*, we risk making enemies who will dismiss us, have a lower opinion of us, or subtly punish us with ostracization.

Self-enhancement. People are motivated to self-enhance—or to perceive and portray themselves more positively than reality would suggest (Sedikides & Gregg, 2008). Of course, social science is rarely directly about the self, but it is often indirectly about the self by being about "people like me" (sometimes referred to, only half-jokingly, as "mesearch).

Social scientists likely have some tendencies to avoid advancing data and theories that portray their own social groups unfavorably or to create obstacles for others who do. This will not always be the case because there may be competing motives for why people might want to perceive different groups in different ways (e.g., men might be more strongly motivated to portray women in a positive light than to portray men in a positive light for ideological reasons, protective reasons, or desires to earn female approval), but absent competing motives, scholars are likely motivated to reject findings that portray their own groups in a negative light. This is one reason to support numerous kinds of diversity among scientists, because these preferences cancel out in the broader literature when numerous scientists have competing motives. These self-enhancement tendencies are more likely to create systematic biases in the field if most social scientists fall within one category (i.e., men, heterosexual, liberal, etc.).

Error management. When faced with complicated information and a noisy environment in which truth cannot be confidently discerned, people have a tendency to favor less costly errors over costlier ones. A classic example found that men have a tendency to overestimate a woman's

sexual interest in them because it is costlier to miss out on a mating opportunity than to make an unwanted sexual advance, whereas women have a tendency to underestimate a man's commitment to her because it is costlier for her to risk pregnancy from a man who will abandon her after sex than to miss out on a sexual opportunity from a man who might commit to and support her (Haselton & Buss, 2000).

This is not a motive separate from the others (desires to gain status, avoid ostracization, and self-enhance), but rather one constantly interacting with the other motives. Imagine, for example, that you have run two studies that found interesting and novel pattern X. You decide to run one more study to really solidify your set of studies before submitting for publication, and you fail to replicate your first two studies even though this third study was very similar to the first two. This is an ambiguous piece of new information—you do not know *why* the third study failed to replicate. Maybe the effect is not real. Or maybe, it was because you ran this third study late at night or toward the end of the semester or because the first two studies used up all the conscientious subjects in the subject pool. In the first case, you miss out on a publication and have wasted time and perhaps money running studies that will never be published. In the latter cases, you can—with justification to yourself—file drawer your third "flawed" study and move forward with just the two. (In such a situation, the right thing to do would be to run a fourth study to test which of your hypotheses about your own findings is correct, but some scholars would not want to risk confirming that the first two studies were flukes.)

Motivated research in practice

Thus far we have explained why the social sciences create an environment ripe for motivated research and why scholars will often have preferences for certain kinds of conclusions

over others—occasionally, though not always—to the detriment of accuracy. But how might motivated research work in practice?

Selective exposure and selective avoidance. At the information recruitment stage, people have a tendency to seek out information that confirms their desired conclusions and avoid information that challenges their desired conclusions (DeMarree, Clark, Wheeler, Briñol, & Petty, 2017; Frimer, Skitka, & Motyl, 2017; Stroud, 2010). These are referred to as selective exposure and selective avoidance, respectively. Although such patterns are often explored in media consumption among everyday people (Stroud, 2008), scholars likely engage in selective exposure in selecting which articles to read. But people also engage in selective exposure by creating social information environments that are likely to deliver information that confirms their desired beliefs, by surrounding themselves with other people who share their cherished beliefs (McPherson, Smith-Lovin, & Cook, 2001) both in person (Motyl, Iyer, Oishi, Trawalter, & Nosek, 2014) and on social media (Bakshy, Messing, Adanic, 2015). In academia, scholars likely "follow" the scholarly and social media outputs of particular scholars whose research and research interests support their own research agenda. Further, one novel source of selective exposure among academics lies within their ability to *create* information that supports particular conclusions. By selecting certain materials and methods that they believe are most likely to confirm their hypotheses and avoiding the use of materials and methods they have less confidence in, they can often generate their own confirmatory information.

Consequently, scholars will be more aware of information that supports their preferred hypotheses than information that challenges it, making their hypotheses appear more plausible and correct than perhaps a more balanced understanding of the literature would predict. Such tendencies would be particularly problematic for review papers, as scholars likely overrepresent

information consistent with their theory and underrepresent contradictory information. These same tendencies can happen with editors and reviewers, who may have imbalanced information about the phenomenon under investigation. If the reviewers have the same blind spots as the authors, they will be unable to point these out. Given the aforementioned and discussed ideological lopsidedness of social science disciplines, blindspots are likely more common than many in the field are willing to concede.

Selective exposure and avoidance can therefore create biased citation patterns, which can continue to perpetuate biased understandings of different domains of research (for recent examples, see Honeycutt & Jussim, 2020). If scholars have preferences for certain conclusions, scholars will be more aware of those findings and thus more likely to cite those findings, and then those highly-cited articles become accepted as the authority on the particular issue. Discordant findings, in contrast, are ignored and eventually forgotten. Ideally, the findings in these highly cited articles are valid, and the relevant knowledge improves theory and applications. But, if biased citation patterns result in the canonization of invalid findings, this can produce a reign of error (Honeycutt & Jussim, 2020) whereby socially desirable, but nonetheless flawed work is propped up to reflect the field's general knowledge. This, in turn, creates dynamics and crises of confidence such as those that have stemmed from psychology's replication crisis. Under these dynamics, biased citation patterns can also contribute to ignoring valid findings, which produces a loss in understanding and deprivation of relevant knowledge. Science strives to be self-correcting, but if invalid findings are canonized and continue to be highly cited, and valid findings (e.g., failed replications) go ignored, self-correction does not occur.

If scholars can acknowledge these tendencies in themselves, they should be motivated to overcome them. A biased awareness of extant data will make it harder to generate hypotheses that are likely to be confirmed by data collection. Exposing oneself to unpalatable information will help scholars identify dead-end hypotheses before they sink time and money into testing them.

Motivated skepticism and credulity. Once people are exposed to information (whether they sought it out or could not avoid it), they engage in motivated skepticism and credulity, or the tendencies to be highly skeptical and critical of discordant information and relatively credulous and uncritical of concordant information (e.g., Ditto & Lopez, 1992; Taber & Lodge, 2006). For example, people are more critical of the methods of a scientific study when the results come to an inconvenient conclusion than the *same methods* when the results come to a preferred conclusion (Lord, Ross, & Lepper, 1979). This can also be conceptualized as a selective call for rigor, whereby one rejects work they do not like on supposedly scientific grounds, but then fail to apply those same standards to work they do like or agree with (Honeycutt & Jussim, under review). People also make more mistakes with both numeric and logical reasoning when conclusions are discordant (Gampa, Wojcik, Motyl, Nosek, & Ditto, 2019; Kahan, Peters, Dawson & Slovic, 2017). Among scholars, this likely happens both in evaluations of one's own findings, as well as in evaluations of others' findings in peer review, acceptance into conferences, awards, decisions to cite, and decisions to hire.

Running experiments on the peer review process can be difficult with tightly controlled methods, but there have been a couple of examinations, which have found that reviewers tended to judge research as higher quality when the findings supported their prior beliefs and theoretical orientations than when the findings challenged their prior beliefs and theoretical orientations

(Koehler, 1993; Mahoney, 1977). This suggests scholars may evaluate research more leniently when findings support their own research agendas. Some research has identified how personal values interfere with the human subjects review process (Ceci et al., 1985). Similarly, research suggests that ideological and moral concerns influence scholars' evaluations of research (Abramowitz, Gomes, & Abramowitz, 1975) and perhaps even their understanding of empirical reality. For example, von Hippel and Buss (2017) found that social psychology professors were more likely to believe that women could have evolved to be more verbally talented than men than that men could have evolved to be more mathematically talented than women. To our knowledge, there is no legitimate scientific reason to believe that one evolved gender difference is more plausible than the other, which suggests their beliefs may be partially motivated by ideological or moral concerns. Moreover, some scholars even openly admit that they would discriminate against conservative research and conservative scholars (Honeycutt & Freberg, 2017; Inbar & Lammers, 2012; Peters et al., 2020).

Other extraneous concerns influence the reviewing process as well. For example, conference submissions from more prestigious scholars and institutions are evaluated more favorably in single blind than double blind reviews, which suggests that either scholars are using a heuristic about prestige and quality or that perhaps scholars are hesitant to give negative evaluations to people and institutions with high status (Tomkins, Zhang, & Heavlin, 2017). Such biases, sometimes also referred to as an eminence obsession (Vazire, 2017), are likely quite common in reviews of peers and research because, as noted above, there is a great deal of noise and ambiguity in evaluating the quality of work. Some scholars have pointed out that the inter-rater reliability of peer review is barely above chance (Lee, Sugimoto, Zhang, & Cronin, 2012). On the one hand, this suggests the possibility that editors are selecting diverse reviewers with

different strengths and perspectives, which in many cases could help cancel out systematic biases. On the other hand, it is a reminder that scientific evaluations—even among experts—are not perfectly objective, and that features of the reviewers influence the perceived quality of science, not merely the science itself.

One report found that reviewer agreement on funding applications was higher for low scoring applications than for top scoring applications (Gallo, Sullivan, & Glisson, 2016). Differentiating between a handful of top candidates is likely more subjective—all the top candidates are high quality, so there is no clear "accurate" or "best" decision, and thus extraneous concerns of the individual scholars have greater influence on their evaluations. Given how frequently scholars are differentiating between high quality content for limited outcomes (journal and conference acceptances, awards, hiring), many of these important decisions that determine scholars' success depend on the idiosyncratic motivations of the reviewers and committee members (so long as applicants reach a certain quality threshold to be considered in the first place). Of course, scholars understand this, and that is why such decisions are usually made among multiple people. This strategy will be more useful when the panel of decision-makers have diverse motivations and preferences, for example, different theoretical and ideological orientations.

Some scholars have contended that these biased information processes are more likely to occur among "experts" or the cognitively sophisticated (e.g., Kahan, 2015; Kahan et al., 2012). People who are more cognitively sophisticated or more knowledgeable would be better able to justify their own biased reasoning processes to themselves and to other people, and thus could get away with more bias than less sophisticated people. Other scholars have challenged this hypothesis, finding that greater cognitive sophistication is instead associated with converging

toward accuracy (McPhetres & Pennycook, 2019). Future research will shed more light on these patterns. It may be that expertise and cognitive sophistication simultaneously increase motivated reasoning and ability to detect accurate patterns (and perhaps motivation to detect accurate patterns), and so in some cases scholars will be more biased than the average person and in other cases, less. There also could be individual differences in whether people tend to "use" their cognitive sophistication more to approach accuracy or to advance their own interests. At minimum, there is no strong evidence that experts and those high in cognitive sophistication are immune to biases.

The Protective Powers of Science

Although scientists themselves are but mere mortals, the *institution* of science can mitigate against scientists' human fallibilities. Peer review requires scholars to convince two to five other scholars who do not (necessarily) share the same motives of the scientist and thus who are not particularly motivated to enhance the importance or quality of whatever manuscript they are reviewing. Sometimes these peers are actually competitors (there is only so much journal space), and so in some cases, reviewers might be strongly motivated to find flaws, which requires authors to be particularly impressive (although, this could also incentivize p-hacking to generate impressive results).

The (mostly) shared mission of seeking true and accurate information incentivizes truth and accuracy-seeking in scientists. All else equal, scholars would prefer to forward *true* impressive results rather than *false* impressive results, because both contribute to status, but the latter creates reputational risk of being discovered as a phony. Scholars likely feel some shame and embarrassment when their own theories fail to hold up and their findings fail to replicate, and much more shame and humiliation when they are caught in outright fraud. Science has

created a culture in which the social response to indicators of dishonest research practices likely disincentivize the most obvious transparent forms of data manipulation and fraud. However, it remains unclear whether that culture has disincentivized more subtle influences and tactics (e.g., using positions of power, such as organization leadership roles and journal editorships, to benefit one's own and one's friends' research and careers).

The Open Science movement has also done a lot to mitigate motivated research, primarily through incentivizing transparency. It is now much more difficult to get away with certain dishonest or questionable research practices. Preregistering studies binds scholars to distinguishing transparently their initial hypotheses from post-hoc fishing expeditions and to their methods and analysis plan, and requires them to indicate when they deviate. Making data publicly available is a big step toward transparency, and likely increases accountability for data tampering. The new "replication movement" has created an atmosphere where all scholars must consider the possibility or probability that some other scholar will try to replicate their findings. This might render scholars more hesitant to publish papers they have little confidence in, because the status and esteem reward could be short-lived and the long-term consequences of work failing to replicate or being labeled a fraud could be far costlier. However, it may be years before other scholars attempt to replicate one's work, let alone publish it, so that the short-term rewards of publishing may still overwhelm the costs of others failing to replicate, which might not be felt for a very long time. By that time, the original researcher may be a full professor with a large grant portfolio, lots of graduate assistants, and a New York Times bestselling book.

But We Can Do Better

Science has an impressive history of generating accurate information over long stretches of time (i.e., converging upon truth), but most of this progress was made by scientists being

completely wrong for long periods of time (young Earth, bleeding to cure illnesses, spontaneous generation of life, all of which were believed for centuries). Some norms of scientific practice in psychology are improving and we hope replication rates in the future will confirm that these new procedures are effective at minimizing researcher degrees of freedom to pursue preferred results and effective in generating a more reliable body of information. But, there are many problems these norms either do not help at all or help only very little.

File drawering. Open science practices do very little to minimize file drawering. Depending on preregistration platform, even preregistered studies can be file drawered without notice. One solution would be to ask scholars to declare in their papers that they have no file drawer studies or any other studies that tested the same hypothesis tested in the paper. Of course, scholars could still lie, but requiring an explicit, public, and published declaration of the lie turns the act of omission into an act of commission, which could create additional psychological barriers. If it is discovered that there were other studies, this act can be considered active fraud rather than a more ambiguous questionable practice. This also increases the likelihood that at least one co-author on a multi-authored paper would object to the outright lie.

There are also selfish reasons not to file drawer your own studies. When scholars file drawer, they inflate their own effect sizes. If and when there are replication attempts, and the findings either fail to replicate or have smaller effect sizes, this will raise suspicion. The more surprising findings are, the more likely it is that other scholars will attempt to replicate the findings, and so by exaggerating the size of one's own effect, scholars likely increase the odds that they will be caught and viewed with suspicion by peers.

Updated replication tracking. A more laborious, but perhaps beneficial, strategy would be for journals to include replication sections on their journal pages for each article where

scholars can link their successful or failed replications of the published study and code their own replication study as failed, successful, or ambiguous/semi-successful. Published studies could then have a live "replication score" attached to them that is easily visible to other scholars who read those published articles. This would help scholars know whether they should take a particular study seriously when theorizing, developing hypotheses, and deciding whether to conduct further replication attempts.

A replication tracking system within the journals might, over the long run, influence the reputation of a journal, and, therefore, incentivize editors to publish robust science rather than flashy science. Such a system might also disincentivize authors from publishing science they have little confidence in because their publication could end up being flagged with a low replication score, which would be embarrassing. This would also provide a greater incentive to those scholars who do fail to replicate a particular study to write up their failed replications. Their replications would be more visible to other scholars interested in the particular effect (rather than buried on some other website) and thus increase the chances that they will at least receive citations for their work (if not publications). This, in turn, would also make it much easier for scholars who wish to conduct meta-analyses to detect successful and failed replications.

Review papers. Review papers are often highly cited and help solidify many broader theories and ideas in the social sciences that are then used by other scholars to generate new hypotheses. Therefore, more than sets of experiments, it is important that scholars get review papers right so they do not waste the time and resources of their peers. Yet Open Science practices do little to help review papers be more accurate and portray the full range of relevant data (rather than a biased subset).

One corrective for review papers would be for editors and reviewers to require explicit and clearly labeled sections containing a mini review of findings that are inconsistent with the present theory or hypothesis. If the scholars know of no research that contradicts their hypothesis, they could be required to say this explicitly in their paper. This should incentivize them to do a dedicated search of inconsistent findings so other scholars do not accuse them of being unfamiliar with the topic even after writing a review.

Such papers could also be required to include statements of falsification. If they present a new theory, it will be important to know not only what it explains or what it predicts or the conditions under which such predictions apply, it will be crucial to know what would falsify the theory's predictions. If stereotypes are declared to be "the default basis of person perception" (Fiske & Neuberg, 1990), how would we know if this was wrong? Would it be falsified by evidence showing powerful individuating information effects? Weak biasing effects? Easily eliminated biasing effects? Even better, scholars could be required to identify the most severe test of the hypothesis—that is, the test that would be *most likely* to detect the falsity of the theory or finding under investigation (O'Donohue, 2021). Theories that generate non-falsifiable predictions are plausibly considered non-scientific, so that one means of elevating the scientific credibility and validity of psychological science would be to articulate explicit statements of what it would take to falsify a theory.

Evaluations of the *quality* of the evidence, and not just the presence/absence or even size of some effect of phenomenon would also be valuable, as is common in Cochrane reviews (the gold standard in medical research). Do studies have large or small N's? Are they experimental or non-experimental? Are they pre-registered or not? If so, did they follow the pre-registration closely or not? The reviews could also use the *new forensics* (p-curves, R-Index, etc.) to evaluate

the quality of the evidence they reviewed (Bartoš & Schimmack, 2020; Simonsohn, 2015; Simonsohn, Nelson, & Simmons, 2014a, 2014b). Evaluations of the quality of the evidence might reduce the wild overclaiming that has characterized so many conclusions in social psychology for decades (Jussim et al, 2016).

Academic reviewing, gatekeeping, and data suppression. Scholars have little ability to criticize the gatekeepers in academia. Calling attention to the flaws of reviewers or editors risks alienation, making it more, not less, difficult to get one's work published and funded. Similarly, accusing a hiring committee, conference committee, or award committee of bias would violate norms in the field; and generally, it is difficult or impossible to know or prove when another scholar is supporting or opposing a particular finding or scholar for non-accuracy reasons. Consequently, there are almost no ways to hold the gatekeepers of academia accountable. Although accountability to reviewers constitutes a check on author biases, there is no comparable check on reviewers or editors' biases.

Open peer review is, however, one way to mitigate some of those biases. Reviews, with or without reviewer identifying information can be publicly posted. Therefore, the entire scientific community at least has the opportunity to evaluate for itself whether a set of reviews are themselves valid and whether their evaluations of a paper have been fair. As public information, it might even become possible for authors to criticize reviews without fear of retaliation.

One growing trend in academia are mob demands to retract papers that have already passed peer review. Unless fraud or statistical errors that change the conclusions are detected, these are data suppression attempts, usually in the service of some moral or political goal (Stevens et al, in press). Attempts to suppress research by mob can be plausibly interpretable as

inability to refute the work -- because if the work could be refuted, the solution would be to publish the supposed refutation and allow readers to judge for themselves which is stronger. Our view is that *building* the discussion, rather than erasing it, is far more likely to advance science. Nonetheless, many journals and editors may feel extreme pressure to give in to such demands because they fear their own reputation or the reputation of the journal. And numerous recent examples exist attesting to this trend (described at length in Honeycutt & Jussim, under review, and Jussim, 2020).

To guard against mob retraction demands, journals should have explicit guidelines for when they will consider retractions. We recommend the Committee on Publication Ethics' guidelines (https://publicationethics.org/retraction-guidelines), which include unreliable findings resulting from major errors or fabrication of data, plagiarism, redundant publications, unauthorized material or data, copyright infringement, research that violated ethics, compromised peer review, or failure to disclose competing interests. Journals should adhere to their guidelines without exception, thus disincentivizing calls for retraction based on other concerns such as alternate explanations, concerns about possible moral implications of the data, or methodological weaknesses. With the rise of retraction-by-outrage-mob procedures, it would be especially useful for journals whose policy is to retract only in cases of fabrication or massive data errors to explicitly state in their instructions to authors that "under no conditions will we retract a paper that has passed peer review and been accepted for publication, or published, on grounds other than those articulated here, no matter how many people sign petitions or open letters or send us outraged emails to do so."

Journals, of course, could create their own guidelines, including, for example "public concerns about the moral implications." This would then signal that they are willing to retract

papers that are objected to by outrage mobs, even when they have passed peer review. This would permit scholars to make their own decisions about whether to publish in journals with retraction policies that violate their own standards for science. Just as transparency will improve the work of authors, it will also improve the work of journals and editors.

Use of strong theories. The idea that human minds and human behavior are the product of evolution and thus what humans think and do should generally promote reproductive success is one of the few broad theories within the social sciences that has withstood substantial criticism and has been very useful for generating countless other more specific hypotheses (Lewis, Al-Shawaf, Conroy-Beam, Asao, & Buss, 2017). If, for example, a particular finding seems inconsistent with natural and sexual selection in human cognition and behavior, skepticism would be warranted. Some may contend that any hypothesis or finding could be made consistent with an evolutionary account, but we doubt this is so. For example, Freud's Oedipus complex, or the idea that young boys would have sexual desire for their mothers and jealousy and hostility toward their fathers, makes almost no sense at all from an evolutionary perspective (e.g., Daly & Wilson, 1990). Using strong theories, those which we can have high confidence in, can help scholars generate better hypotheses, which is advantageous both for scientists and scientific progress.

Adversarial collaborations. Working with others with whom you disagree might be temporarily unpleasant, but it will make you a better scientist (see, e.g., Bateman, Kahneman, Munro, Starmer, & Sugden, 2005; Mellers, Hertwig, & Kahneman, 2001). Those who disagree with you have a different perspective, and possibly different motives and biases, that can help cancel out systematic error in your own work. Adversarial collaborations require scholars to articulate their hypothesis in a clear, testable way, understand their adversary's hypothesis as

their adversary understands it (and not as a caricature), identify actual points of disagreement (rather than imagined ones based on caricatures), and generate methods that could differentiate between the two hypotheses and feasibly falsify either hypothesis. These kinds of collaborations constrain researcher degrees of freedom because adversaries will not approve of methodological approaches that provide (even if unintentionally) rigged tests of hypotheses or that appear designed to confirm a scholar's hypothesis.

They also have greater potential to advance debates and change minds. Because scholars commit to a methodological approach before testing their competing hypotheses, this minimizes scholars' ability to post-hoc criticize methods, explain away unexpected results, and file drawer undesired results. Third parties should be more persuaded by results of adversarial collaborations knowing that a scholar who made opposite predictions stands behind the methods and findings.

Although adversarial collaborations might feel like an unnecessary constraint in the short-term, it will likely improve research in the long run (Ellemers, Fiske, Abele, Koch, & Yzerbyt, 2020). If your hypothesis is correct, it likely will win out in an adversarial collaboration. If it is incorrect, likely it will eventually be falsified regardless of whether you discover this on your own in an adversarial collaboration or whether other scholars discover this in failed replications or failed conceptual replications. Delaying the inevitable by refusing to participate in adversarial collaborations only risks wasting more time and money and lowering the ratio of science that *will* withstand the test of time.

Nonetheless, adversarial collaborations can also be quite difficult. Especially if adversaries have been openly hostile with one another, forging the cooperative bonds needed to work together on a project may be a bridge too far. Even without personal antipathy, however, bridging differences in assumptions, perspectives, and motives can be a formidable and effort-

intensive task. We all have only limited time and resources, and, sometimes, such a project may not be viewed as worth the effort.

On the other hand, we can also imagine a scientific world in which adversarial collaborations were incentivized, thereby rewarding researchers who succeed at bridging these divides. Given the higher confidence we can have in the findings resulting from adversarial collaborations, editors and reviewers should consider these a methodological strength, similar to preregistrations, large sample sizes, and meta-analyses. Given the self-discipline and commitment to rigor required to participate in adversarial collaborations, such efforts should be rewarded by other scholars when making hiring, funding, and award decisions. Participation in these collaborations indicates that a scientist is committed to truth-seeking rather than in advancing flashy results that may not hold up to higher scrutiny.

Reward rigor. Grants and awards in the social sciences should prioritize scholars who produce robust effects—those that are reliable and replicable and stand up to severe scrutiny. Truth-telling and rigor should be prioritized over flash, drama, novelty, counter-intuitiveness, and supposedly easy solutions to complex problems. By providing resources to researchers that produce findings that are true, powerful, and robust, psychology will wander down far more scientifically productive paths than if it follows every bright shiny object that shows up flashing $p<.05$ and a compelling narrative. Of course, *sometimes*, even those findings *will* be flashy or dramatic. But flash and drama should not detract from the value of work, and might be valuable add-ons, *if* that work was conducted in such a manner as to lead to high confidence that it is true, powerful, and robust.

The Case for Accuracy Motivations

Scholars have much to gain by forwarding flashy, socially important, self-promotional, group-promotional, timely results. However, in the new era of Open Science, such gains could be short lived if findings are not also accurate—replicable, with correct interpretations and conclusions. Accepting that we ourselves are humans who are vulnerable to unconscious motivations that influence the ways we conduct science and the conclusions we come to should motivate us to place regulations on ourselves (e.g., refusing to file drawer our own studies, searching for information that challenges our beliefs and hypotheses, working with scholars with whom we disagree). Unfortunately, even when people are presented with reasonably compelling evidence that they might have biases that steer their judgments away from accuracy, they seem unable to recognize these tendencies in themselves (Pronin, Lin, & Ross, 2002). If you wish to be the exception to the rule, start not by denying that you are human and prone to biases and motivations, but instead by having a conscience that bravely admits this to yourself.

References

Abramowitz, S. I., Gomes, B., & Abramowitz, C. V. (1975). Publish or politic: Referee bias in manuscript review. *Journal of Applied Social Psychology, 5*, 187-200.

AlShebli, B., Makovi, K., & Rahwan, T. (2020). The association between early career informal mentorship in academic collaborations and junior author performance. *Nature Communications*, *11*, 1-8.

Anderson, C., & Kilduff, G. J. (2009). The pursuit of status in social groups. *Current Directions in Psychological Science*, *18*, 295-298.

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*, 1130-1132.

Baltiansky, D., Jost, J., & Craig, M. A. (2020). At whose expense? System justification and the appreciation of stereotypical humor targeting high vs. low status groups. *Humor: International Journal of Humor Research*.

Bartoš, F., & Schimmack, U. (2020). Z-Curve.2.0: Estimating Replication Rates and Discovery Rates. *PsyArXiv*.

Bateman, I., Kahneman, D., Munro, A., Starmer, C., & Sugden, R. (2005). Testing competing models of loss aversion: An adversarial collaboration. *Journal of Public Economics*, *89*(8), 1561-1580.

Baumeister, R. F., Maranges, H. M., & Vohs, K. D. (2018). Human self as information agent: Functioning in a social environment based on shared meanings. *Review of General Psychology, 22*, 36-47.

Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P. E. (2009). Strong claims and weak evidence Reassessing the predictive validity of the IAT. *Journal of Applied Psychology, 94,* 567-582.

Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theoretical Population Biology, 65,* 17-28.

Brady, W. J., Wills, J. A., Burkart, D., Jost, J. T., & Van Bavel, J. J. (2019). An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of Experimental Psychology: General, 148,* 1802-1813.

Campbell, T. H., & Kay, A. C. (2014). Solution aversion: On the relation between ideology and motivated disbelief. *Journal of Personality and Social Psychology, 107,* 809-824.

Ceci, S. J., Peters, D., & Plotkin, J. (1985). Human subjects review, personal values, and the regulation of social science research. *American Psychologist, 40,* 994–1002.

Clark, C. J., Liu, B. S., Winegard, B. M., & Ditto, P. H. (2019). Tribalism is human nature. *Current Directions in Psychological Science, 28,* 587-592.

Clark, C. J., & Winegard, B. M. (2020). Tribalism in war and peace: The nature and evolution of ideological epistemology and its significance for modern social science. *Psychological Inquiry, 31,* 1-22.

Clark, C. J., Winegard, B. M., & Farkas, D. (2020). A cross-cultural analysis of censorship on campuses. Unpublished manuscript.

Daly, M., & Wilson, M. (1990). Is parent-offspring conflict sex-linked? Freudian and Darwinian models. *Journal of Personality, 58,* 163-189.

DeMarree, K. G., Clark, C. J., Wheeler, S. C., Briñol, P., & Petty, R. E. (2017). On the pursuit of desired attitudes: Wanting a different attitude affects information processing and behavior. *Journal of Experimental Social Psychology, 70*, 129-142.

Ditto, P. H., Clark, C. J., Liu, B. S., Wojcik, S. P., Chen, E. E., Grady, R. H., ... & Zinger, J. F. (2019). Partisan bias and its discontents. *Perspectives on Psychological Science, 14*, 304-316.

Ditto, P. H., Liu, B. S., Clark, C. J., Wojcik, S. P., Chen, E. E., Grady, R. H., ... & Zinger, J. F. (2019). At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives. *Perspectives on Psychological Science, 14*, 273-291.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, *63*, 568-584.

Drummond, C. & Fischhoff, B. (2017). Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. *Proceedings of the National Academy of Sciences*, *114,* 9587-9592.

Duarte, J. L., Crawford, J. T., Stern, C., Haidt, J., Jussim, L., & Tetlock, P. E. (2015). Political diversity will improve social psychological science 1. *Behavioral and Brain Sciences*, *38*, 1-13.

Durkee, P. K., Lukaszewski, A. W., & Buss, D. M. (2020). Psychological foundations of human status allocation. *Proceedings of the National Academy of Sciences*, *117*, 21235-21241.

Eitan, O., Viganola, D., Inbar, Y., Dreber, A., Johannesson, M., Pfeiffer, T., ... & Uhlmann, E. L. (2018). Is research in social psychology politically biased? Systematic empirical tests and

a forecasting survey to address the controversy. *Journal of Experimental Social Psychology, 79*, 188-199.

Ellemers, N., Fiske, S. T., Abele, A. E., Koch, A., & Yzerbyt, V. (2020). Adversarial alignment enables competing models to engage in cooperative theory building toward cumulative science. Proceedings of the National Academy of Sciences, 117, 7561–7567.

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019). A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology, 117*, 522-559.

Frimer, J. A., Skitka, L. J., & Motyl, M. (2017). Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions. *Journal of Experimental Social Psychology*, *72*, 1-12.

Gallo, S. A., Sullivan, J. H., & Glisson, S. R. (2016). The influence of peer reviewer expertise on the evaluation of research funding applications. *PloS one*, *11*, e0165147.

Gampa, A., Wojcik, S. P., Motyl, M., Nosek, B. A., & Ditto, P. H. (2019). (Ideo) Logical Reasoning: Ideology Impairs Sound Reasoning. *Social Psychological and Personality Science*.

Grubbs, J. B., Warmke, B., Tosi, J., James, A. S., & Campbell, W. K. (2019). Moral grandstanding in public discourse: Status-seeking motives as a potential explanatory mechanism in predicting conflict. *PLOS ONE*, *14*, e0223749.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review, 108*, 814-834.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, *78*, 81-91.

Hehman, E., Flake, J. K., & Calanchini, J. (2018). Disproportionate use of lethal force in policing is associated with regional racial biases of residents. *Social Psychological and Personality Science, 9*, 393-401.

Honeycutt, N., & Freberg, L. (2017). The liberal and conservative experience across academic disciplines: An extension of Inbar and Lammers. *Social Psychological and Personality Science, 8*, 115–123.

Honeycutt, N. & Jussim, L. (under review). On the connection between bias and censorship in academia. J. Lambe & S. Stevens (Eds.), *Building Trust Through Freedom of Expression*.

Honeycutt, N., & Jussim, L. (2020). A model of political bias in social science research. *Psychological Inquiry, 31*, 73-85.

Inbar, Y., & Lammers, J. (2012). Political diversity in social and personality psychology. *Perspectives on Psychological Science, 7*, 496-503.

Jussim, L. (2020). The threat to academic freedom … from academics. Retrieved from: https://medium.com/@leej12255/the-threat-to-academic-freedom-from-academics-4685b1705794

Jussim, L., Careem, A., Goldberg, Z., Honeycutt, N., & Stevens, S. (in press). IAT scores, racial gaps, and scientific gaps. In J. A. Krosnick, T. H. Stark, & A. L. Scott (Eds.), *The Future of Research on Implicit Bias*.

Jussim, L., Crawford, J. T., Anglin, S. M., Stevens, S. T., & Duarte, J. L. (2016). Interpretations and methods: Towards a more effectively self-correcting social psychology. *Journal of Experimental Social Psychology, 66*, 116-133.

Jussim, L., Krosnick, J. A., Stevens, S. T., & Anglin, S. M. (2019). A social psychological model of scientific practices: Explaining research practices and outlining the potential for successful reforms. *Psychologica Belgica*, *59*, 353-372.

Jussim, L., & Maoz, I. (2014). Desperately seeking the "wow effect": Data interpretation and scientific story-telling as issues of research integrity. *Presentation at the October 2014 meeting of the Society for Experimental Social Psychology*.

Kahan, D. M. (2012). Ideology, motivated reasoning, and cognitive reflection: An experimental study. *Judgment and Decision Making*, *8*, 407-424.

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, *1*, 54-86.

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, *2*, 732-735.

Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes*, *56*, 28-55.

Kopko, K. C., Bryner, S. M., Budziak, J., Devine, C. J., & Nawara, S. P. (2011). In the eye of the beholder? Motivated reasoning in disputed elections. *Political Behavior*, *33*, 271-290.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480-498.

Langbert, M. (2018). Homogenous: The Political Affiliations of Elite Liberal Arts College Faculty. *Academic Questions*, *31*, 186–197.

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, *64*, 2-17.

Lewis, D. M., Al-Shawaf, L., Conroy-Beam, D., Asao, K., & Buss, D. M. (2017). Evolutionary psychology: A how-to guide. *American Psychologist*, *72*, 353-373.

Lilienfeld, S. O. (2015). Lack of political diversity and the framing of findings in personality and clinical psychology. *Behavioral and Brain Sciences; New York*, *38*, n/a.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*, 2098-2109.

Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, *1*, 161-175.

Maranges, H. M., Hasty, C. R., Maner, J. K., & Conway, P. (2020). The behavioral ecology of moral dilemmas: Childhood unpredictability, but not harshness, predicts less deontological and utilitarian responding. *Journal of Personality and Social Psychology.*

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*, 415-444.

McPhetres, J. (2019). A Perspective on the Relevance and Public Reception of Psychological Science. *Collabra: Psychology*, *5*, 34.

McPhetres, J., & Pennycook, G. (2019). Science beliefs, political ideology, and cognitive sophistication. Unpublished manuscript.

Meehl, P. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*, 195-244.

Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science*, *12*, 269-275.

Motyl, M., Iyer, R., Oishi, S., Trawalter, S., & Nosek, B. A. (2014). How ideological migration geographically segregates groups. *Journal of Experimental Social Psychology*, *51*, 1-14.

Munro, G. D., Lasane, T. P., & Leary, S. P. (2010). Political partisan prejudice: Selective distortion and weighting of evaluative categories in college admissions applications. *Journal of Applied Social Psychology*, *40*, 2434-2462.

Munro, G. D., Weih, C., & Tsai, J. (2010). Motivated suspicion: Asymmetrical attributions of the behavior of political ingroup and outgroup members. *Basic and Applied Social Psychology*, *32*, 173-184.

Nosek, B. A., Aarts, A. A., Anderson, J. E., Kappes, H. B., & Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, aac4716-aac4716.

O'Donohue, W. (in press). Are psychologists appraising research properly?: Some Popperian notes regarding replication failures in psychology. *Journal of Theoretical and Philosophical Psychology*.

Ouwerkerk, J. W., Kerr, N. L., Gallucci, M., & Van Lange, P. A. M. (2005). *Avoiding the Social Death Penalty: Ostracism and Cooperation in Social Dilemmas.* In K. D. Williams, J. P. Forgas, & W. von Hippel (Eds.), *Sydney Symposium of Social Psychology series. The social outcast: Ostracism, social exclusion, rejection, and bullying* (p. 321–332). Psychology Press.

Payne, B. K., Vuletich, H. A. & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry, 28,* 233-248.

Peters, U., Honeycutt, N., Block, A. D., & Jussim, L. (2020). Ideological diversity, hostility, and discrimination in philosophy. *Philosophical Psychology*, 1–38.

Pew (2020). Science and scientists held in high esteem across global publics. Retrieved on November 25, 2020 from https://www.pewresearch.org/science/2020/09/29/scientists-are-among-the-most-trusted-groups-in-society-though-many-value-practical-experience-over-expertise/

Prentice, D. A. (2012). Liberal Norms and Their Discontents. *Perspectives on Psychological Science, 7*, 516–518.

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*, 369-381.

Pursur, H., & Harper, C. (2020). Low system justification drives ideological differences in joke perception: A critical commentary and re-analysis of Baltiansky et al. (2020). Unpublished manuscript.

Reber, A. S. (1989). Implicit learning and tacit knowledge. Journal of Experimental Psychology: General, 118, 219-235.

Reinero, D. A., Wills, J. A., Brady, W. J., Mende-Siedlecki, P., Crawford, J. T., & Van Bavel, J. J. (2020). Is the political slant of psychology research related to scientific replicability?. *Perspectives on Psychological Science*, *15*, 1310-1328.

Rothman, S., Lichter, S. R., & Nevitte, N. (2005). Politics and professional advancement among college faculty. *The Forum, 3.*

Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives on Psychological Science*, *3*, 102-116.

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.

Simonsohn, U. (2015). Small Telescopes: Detectability and the Evaluation of Replication Results. *Psychological Science, 26,* 559–569.

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

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). p-Curve and Effect Size: Correcting for Publication Bias Using Only Significant Results. *Perspectives on Psychological Science, 9,* 666–681.

Stevens, S. T., Jussim, L., & Honeycutt, N. (2020). Scholarship Suppression: Theoretical Perspectives and Emerging Trends. *Societies*, *10,* 82.

Stewart-Williams, S., Thomas, A., Blackburn, J. D., & Chan, C. Y. M. (2020). Reactions to male-favoring vs. female-favoring sex differences: A preregistered experiment. *British Journal of Psychology*.

Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, *60*, 556-576.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50,* 755-769.

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*, 853-870.

Tomkins, A., Zhang, M., & Heavlin, W. D. (2017). Reviewer bias in single-versus double-blind peer review. *Proceedings of the National Academy of Sciences*, *114*, 12708-12713.

Vazire, S. (2017). Our obsession with eminence warps research. *Nature News*, *574*(7661), 7.

von Hippel, W., & Buss, D. M. (2017). Do ideologically driven scientific agendas impede the understanding and acceptance of evolutionary principles in social psychology. In J. T. Crawford & L. Jussim. *Frontiers of Social Psychology Series: The Politics of Social Psychology* (pp. 7–25). New York, NY: Routledge.

Winegard, B. M., & Clark, C. J. (2020). Without Contraries is no Progression. *Psychological Inquiry, 31*, 94-101.

Winegard, B. M., Clark, C. J., Hasty, C., & Baumeister, R. F. (2018). Equalitarianism: A source of liberal bias. Unpublished manuscript.

Winegard, B., Winegard, B., & Geary, D. C. (2018). The status competition model of cultural production. *Evolutionary Psychological Science, 4*, 351-371.

Zigerell, L. J. (2018). Black and White discrimination in the United States: Evidence from an archive of survey experiment studies. Research and Politics, 5, 2053168017753862.