Perspectives on Psychological Science
1–19

DOI: 10.1177/1745691620924463
www.psychologicalscience.org/PPS

Is the Political Slant of Psychology Research Related to Scientific Replicability?

Diego A. Reinero[1], Julian A. Wills[1], William J. Brady[1,2], Peter Mende-Siedlecki[3], Jarret T. Crawford[4], and Jay J. Van Bavel[1,5]

[1]Department of Psychology, New York University; [2]Department of Psychology, Yale University; [3]Department of Psychological and Brain Sciences, University of Delaware; [4]Department of Psychology, The College of New Jersey; and [5]Center for Neural Science, New York University

Abstract

Social science researchers are predominantly liberal, and critics have argued this representation may reduce the robustness of research by embedding liberal values into the research process. In an adversarial collaboration, we examined whether the political slant of research findings in psychology is associated with lower rates of scientific replicability. We analyzed 194 original psychology articles reporting studies that had been subject to a later replication attempt (N = 1,331,413 participants across replications) by having psychology doctoral students (Study 1) and an online sample of U.S. residents (Study 2) from across the political spectrum code the political slant (liberal vs. conservative) of the original research abstracts. The methods and analyses were preregistered. In both studies, the liberal or conservative slant of the original research was not associated with whether the results were successfully replicated. The results remained consistent regardless of the ideology of the coder. Political slant was unrelated to both subsequent citation patterns and the original study's effect size and not consistently related to the original study's sample size. However, we found modest evidence that research with greater political slant—whether liberal or conservative—was less replicable, whereas statistical robustness consistently predicted replication success. We discuss the implications for social science, politics, and replicability.

Keywords

replication, politics, ideology, liberal, conservative, bias

There is a growing debate about the political composition of faculty members at postsecondary institutions and its effect on research and teaching. Numerous studies have suggested that many academic fields are predominantly composed of Democrats or liberals[1] (Eagan et al., 2014; Gross, 2013; Gross & Simmons, 2014; Hamilton & Hargens, 1993; Ladd & Lipset, 1975; Lazarsfeld & Thielens, 1977). In a recent report of 40 leading American universities, researchers found that faculty who were registered Democrats outnumbered faculty who were registered Republicans across five kinds of departments (Langbert, Quain, & Klein, 2016). The imbalance was smallest in economics (4.5:1), larger in psychology (17.4:1), and largest in history (33.5:1).[2] Recent data on psychologists' self-reported political ideology (e.g., Duarte et al., 2015; Inbar & Lammers, 2012; Skitka, 2012; Von Hippel & Buss, 2017) suggest that roughly 85% to 90% of the field is liberal.[3] This has led to speculation that the large number of liberals in many academic fields might influence research and teaching.

Researchers have argued that political homogeneity among academics undermines the validity of some social psychological research (Buss & von Hippel, 2018; Crawford & Jussim, 2018; Duarte et al., 2015; Eagly, 1995; Redding, 2001) and jeopardizes the objectivity that science strives to achieve (Crawford, 2017; Jussim, Crawford, Anglin, & Stevens, 2015). According to this perspective, a homogeneous group without enough dissenting minorities can lead to groupthink (Crano, 2012; Fiske, Harris, & Cuddy, 2004; Janis, 1972). For example, the sociologist Musa al Gharbi (2018) asserted

Corresponding Author:
Jay J. Van Bavel, New York University, 6 Washington Pl., New York, NY 10003
E-mail: jay.vanbavel@nyu.edu

that "ideologically-driven errors likely permeate a good deal of social research" (p. 496), and psychologist Jonathan Haidt (2016) considered "the rapid loss of political diversity, over the last 20 years, to be the second-greatest existential threat to the field of social psychology, after the 'replication crisis'" (para. 14). This perspective may be echoed by members of the general public as well, who believe research in the social sciences is partially geared toward obtaining evidence consistent with researchers' ideologies (Hannikainen, 2019).

These growing concerns led at least one commentator to speculate that this political imbalance may have contributed to the low rates of replicability in psychology (Brooks, 2015) and propelled the Dutch government to pass a motion recommending that the Royal Netherlands Academy of Arts and Sciences study whether political bias affects research outcomes (Brugh, 2017a, 2017b). However, no work has formally tested the relationship between the political slant of research and its scientific robustness. To address this gap, in the current article, we examine the relationship between political ideology and the replicability of psychology research in a sample of 194 original psychology articles reporting studies that had been subject to a later replication attempt (with a total sample of 1,331,413 participants across replications).

The specific concern expressed by some critics is that a discipline composed overwhelmingly by scientists who are liberal might result in one-sided questions or mischaracterizations of other political viewpoints and that these scholars might be more lenient when reviewing liberal-leaning research (or stricter with conservative-leaning research). If the research or review process was selectively compromised, it could allow the publication of liberal-leaning claims based on flimsy evidence—even if they are unlikely to hold up to scientific replication. This could be viewed as a form of *liberal bias*.

There is extensive evidence that political identities can engage motivated cognition (Kahan, 2013; Van Bavel & Pereira, 2018). Research on politicized topics such as climate change (Funk & Kennedy, 2016), gun violence, vaccinations (Kahan, Braman, Cohen, Gastil, & Slovic, 2010), and health care reform (Nyhan, Reifler, & Ubel, 2013) has suggested that public belief in these data diverge along partisan lines. In addition, a recent meta-analysis found that both liberals and conservatives engage in motivated reasoning (Ditto et al., 2018; but see Baron & Jost, 2018).

Such motivated political cognition might influence various stages of the scientific process (see Duarte et al., 2015), from the study design and data analysis to editorial decisions and citation patterns. Indeed, there is evidence that peer review may be susceptible to the social preferences of reviewers and editors. For instance, single-blind reviewing confers a significant advantage to manuscripts with well-known authors and authors from high-prestige institutions relative to double-blind review (Tomkins, Zhang, & Heavlin, 2017). In addition, male STEM (science, technology, engineering, and mathematics) faculty rate research less favorably when it finds evidence of a gender bias against women in STEM (whereas women rate research less favorably when it *does not* find evidence of gender bias against women in STEM; Handley, Brown, Moss-Racusin, & Smith, 2015). These findings raise the possibility that scientists may express similar forms of bias toward manuscripts that do not align with their own political worldview.

Given the political base rates of academia, peer reviewers are likely to be liberal, and biases have been documented among social scientists' interpretations and evaluations of research (MacCoun, 1998). For instance, a review of 68 articles containing empirical evidence on journal peer review concluded that the peer-review system was unfair and discouraged innovation—a conclusion supported by evidence that "findings that conflict with current beliefs are often judged to have defects" (Armstrong, 1997, p. 63).[4] A more recent analysis of 306 politically relevant abstracts from the Society for Personality and Social Psychology revealed that liberals are characterized slightly more positively than conservatives and that conservatives are more often the target of explanation than liberals (Eitan et al., 2018).

More broadly, if scientists' personal political identities or beliefs cannot be sufficiently divorced from their own research, they might (a) solely form hypotheses that align with an ideologically congruent narrative (e.g., a liberal professor solely studying the inaccurate and pernicious effects of social stereotypes), (b) embed their personal ideological values into how they measure variables or broader constructs (e.g., leading survey questions or scales lacking construct validity), (c) look for ideologically congruent results (e.g., *p*-hacking results until a pattern emerges that supports their worldview), (d) interpret and report results in an ideologically congruent manner (e.g., framing their findings under ideologically congruent theories and using value-laden language in the abstract and manuscript), or (d) try to publish results that are ideologically congruent and place results that are ideologically incongruent in a file drawer. If the many liberals in academia (particularly those in the social sciences in which theories are more politically relevant) all pursued the above listed practices, it would result in a heavily skewed distribution of research topics, a distortion of many topics through one-sided framing, and ultimately, results that may not be reliable or replicable.

Indeed, the robustness and replicability of results could be affected by such political biases in several ways. For example, political biases could lead scientists to analyze data in certain ways to support their political worldview or exclude failed studies. Thus, when other scientists attempt to replicate the result, the result may not be replicable. Alternatively, political biases could lead scientists to convey a narrow effect (bounded by the precise stimuli, survey questions, or sample used) to others as a broader phenomenon or otherwise lack theory specification (Muthukrishna & Henrich, 2019). When other scientists seek to replicate that work with a different sample or method, the results may not be replicable. Alternatively, peer reviewers may be more lenient with tenuous effects (that may not be true effects) or statistical rigor (e.g., smaller sample size) if the findings support their own personal political ideology. When other scientists seek to replicate that work, the results would not be robust or replicable.

On the other hand, if scientists dogmatically followed contemporary ideological beliefs and ignored logic and empirical evidence, they would not have discovered that the world is round, advanced the theory of evolution, or invented modern medicine. Although scientists are not immune to human heuristics such as confirmation bias (Nickerson, 1998), they tend to be both more open-minded (Lounsbury et al., 2012) and require more empirical consistency (Hogan & Maglienti, 2001) than nonscientists. Thus, scientists are more willing than others to consider new, convincing data even if the data counter a dominant theory. Perhaps more importantly, the norms of science attenuate the biases of individual scientists by institutionalizing vigorous debate and criticism (Merton, 1973). For instance, the peer-review process is well designed to diminish groupthink because reviews are normally conducted in parallel by anonymous reviewers at arm's length from the authors (Van Bavel, Reinero, Harris, Robertson, & Pärnamets, 2020).

The current research was designed to address the role of political ideology on the quality of scientific articles. Empirical evidence on this topic is scant, and there is reason to believe that scholars may overestimate the role of political bias in the research literature. For instance, the size of political bias in the recent study cited above of abstracts from the Society for Personality and Social Psychology was not only small (Cohen, 1988) but also significantly smaller than predicted by a separate set of raters (Eitan et al., 2018). Thus, even when the peer-review process was limited to reading and evaluating abstracts, there was much weaker evidence of political bias than expected. Moreover, these abstracts had undergone only minimal peer review because the underlying features of scientific robustness (e.g., statistical power, effect size) are rarely available in conference abstracts. Thus, the attributes of scientists and the process of journal peer review may mitigate against the potential for political ideology to significantly influence research and publication decisions given that these individual differences and structural factors place a higher value on truth.

In two studies, we examined whether the political slant of research (i.e., whether research conclusions are more consistent with a liberal or conservative worldview) was associated with less replicable or statistically robust (i.e., effect size and sample size of the original research) published psychology research. We also examined whether liberal findings are cited and discussed more often than conservative findings (Jussim, Crawford, Anglin, Stevens, & Duarte, 2016). We defined *left* (liberal) and *right* (conservative) according to contemporary American politics. The key difference between the two studies is the coders we used to determine the political slant of the original research: In Study 1, we used a politically balanced sample of six psychology doctoral students (including pairs of self-identified liberals, moderates, and conservatives), and in Study 2, we used a larger, politically diverse sample of American residents (using an Amazon Mechanical Turk [MTurk] sample). To mitigate the possible influence of our own views, we formed an "adversarial collaboration" (as advocated for by Kahneman, 2003; Tetlock & Mellers, 2011) using two sets of authors who were simultaneously testing the same question with different theoretical commitments (similar to the adversarial collaboration of Mellers, Hertwig, & Kahneman, 2001). Furthermore, both sets of authors independently preregistered their methods and analyses (as advocated for by Nosek & Lindsay, 2018; Shrout & Rodgers, 2018). Specifically, although we did not design our respective studies together, we realized that we were simultaneously testing the same question and that our methods happened to be fairly similar. Thus, we joined forces during the data-collection phase, which enabled us to collaborate on our data collection, analysis, interpretation, and writing of the results. Despite the fact that each study was run by a different set of authors, their results were strikingly similar, and so we present them together to show the robust nature of our findings across different types of coders (expert vs. lay coders) and scientific methods. We focus on inferences that are consistent across both studies.

Method

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All analyses were preregistered (Study 1 preregistration: https://osf.io/nh9gj/; Study 2 preregistration: https://osf.io/5ke68/) unless otherwise explicitly stated as exploratory. We adhered to all of our preregistered analyses and report any deviations from our preregistrations (see the Deviations From Preregistration section in the Supplemental Material available online). All data analyses were performed using the R software environment (Version 3.5.1; R Core Team, 2018), predominantly using *tidyverse* (Wickham, 2017) for data wrangling and producing figures, and *lme4* (Bates, Mächler, Bolker, & Walker, 2015) for mixed-model analyses. The full reproducible code can be found on OSF (https://osf.io/zftxe/).

In both studies, coders rated the political slant of 194 original psychology articles[5] that were also subject to a replication attempt.[6] This data set included 479 replication attempts[7] from eight different publicly available repositories, involving 1,331,413 participants (for further information, see preregistrations); all data (https://osf.io/pc9xd/) and code (https://osf.io/zftxe/) are available on OSF. We did not do a formal power analysis to determine sample size because we sought to collect the entire population of psychology replications from large-scale replication projects. Doing so mitigated researcher degrees of freedom and allowed us to maximize available power.[8] Several replication attempts were part of large-scale efforts that sought to replicate some of the most influential original findings in psychology (e.g., Association for Psychological Science's Registered Replication Reports [APS RRR]). Other replication efforts explicitly sought to minimize selection biases and maximize generalizability of the accumulated evidence (e.g., Reproducibility Project: Psychology) through choosing articles that ranged in topic and subdiscipline, time period, differing levels of certainty and existing impact, classic and contemporary effects, and publication outlets (e.g., Many Labs). The eight repositories were as follows: APS RRRs (Simons, Holcombe, & Spellman, 2014), Curate Science (LeBel & Battista, 2014), Many Labs 1 (Klein et al., 2014), Many Labs 2 (Klein et al., 2018), Many Labs 3 (Ebersole et al., 2016), Pre-Publication Independent Replications (Schweinsberg et al., 2016), Reproducibility Project: Psychology (Open Science Collaboration [OSC], 2015), and a special issue of *Social Psychology* (Epstude & Meerholz, 2014). To our knowledge, our article provides the largest analysis of replications in the social sciences.

Participants

Study 1. In the summer of 2016, we sent out a recruitment survey to the Society for Personality and Social Psychology mailing list calling for doctoral coders to rate psychology abstracts on the political orientation of their study conclusions. We recruited psychology doctoral coders because they would have experience reading and comprehending published psychological research while lacking the in-depth knowledge of most of the original research, which was published well before they started their doctoral degrees. We asked respondents to self-report age, sex, current doctoral year, and ideology (5-point scale: 1 = *very liberal*, 3 = *moderate*, 5 = *very conservative*). We received 340 responses and randomly selected[9] six social psychology doctoral student coders so that we had an equal number of very liberal, moderate, and very conservative coders who were maximally balanced on age, sex, and year of doctoral training to minimize differences (for doctoral coder demographics, see Table S1 in the Supplemental Material). This sample of coders allowed us to determine whether our conclusions would generalize to coders across the political spectrum.

Study 2. Although our doctoral-coder sample was selected to provide political balance to minimize inadvertent bias, those restrictions left us with a small sample of coders. In Study 2, we recruited a much larger sample of U.S. residents to serve as coders. Specifically, we recruited 511 online MTurk workers (47% male, mean age = 37 years; for MTurk coder demographics, see Fig. S1 in the Supplemental Material). The number of online coders was informally determined such that every abstract had at least a dozen ratings. This sample of coders allowed us to determine whether our conclusions would generalize to a lay audience of nonexperts.

Materials and procedure

Study 1. To strengthen coding reliability, the selected doctoral coders completed two practice rounds in which they rated the political slant of 12 abstracts (four in the first round and eight in the second round). The abstracts were selected to represent liberal and conservative findings (as well as moderate and nonpolitically relevant findings), and coders received feedback after each round regarding their accuracy (for details, see Practice Round Process for Study 1 in the Supplemental Material). Practice abstracts were selected from the same journals and time period as the test abstracts, although no practice abstract overlapped with a test abstract.

Before completing any ratings, coders were given example definitions of liberalism and conservatism along with flattering and unflattering profiles of liberals and conservatives (Tetlock & Mitchell, 1993) to provide a reminder about the common divergences between these ideologies and relative anchors for the different ends of the political-slant scale. In Study 1, political slant was rated on a 5-point scale (1 = *very left leaning*, 2 = *slightly left leaning*, 3 = *politically relevant but no lean*, 4 = *slightly right leaning*, 5 = *very right leaning*). We also included an option to code the abstract as not politically relevant ("does not apply") to allow greater sensitivity of our measure and avoid conflating a moderate abstract with one that was not politically relevant. Ratings from a coder that were along the 5-point scale were considered politically relevant, whereas a rating of "does not apply" was considered not politically relevant.

After successfully completing the two practice rounds, coders then read the abstracts of the 194 original psychology articles in our database and rated the political slant of the study's research conclusion (surveys were chunked into three waves to avoid rating fatigue). All original abstracts were reformatted to plain text and standardized to avoid incidentally providing clues to our coders as to which journal they came from. For each abstract, doctoral coders also rated the subdiscipline of the abstract by choosing from among five options: personality, social, developmental, cognitive, or perception.[10] The doctoral coders also rated the contextual sensitivity of an abstract using the 5-point scale from Van Bavel, Mende-Siedlecki, Brady, and Reinero (2016; 1 = *context is not at all likely to affect the results*, 3 = *context is somewhat likely to affect the results*, 5 = *context is very likely to affect the results*) and rated how robust the results seemed on a 5-point scale (1 = *not at all robust*, 2 = *slightly robust*, 3 = *moderately robust*, 4 = *very robust*, 5 = *extremely robust*). The order in which context sensitivity and robustness were rated was randomized. We then recorded whether the coder was familiar with the results of prior work on contextual sensitivity and replicability and gathered demographics.

We averaged the political-slant-scale ratings to form an average measure of political slant for each abstract. Although it was not part of our preregistered analysis plan, we also analyzed our data using random-effects models that included a random intercept and slope for each rater and thus incorporated the individual ratings of each coder. Following our preregistered rule, if at least four of the six coders (i.e., the majority of coders) rated the abstract as not politically relevant, that abstract was coded as not politically relevant and was not included in the primary analysis. We also averaged all six coders' ratings of contextual sensitivity and robustness to produce respective mean scores for each abstract.

Study 2. In Study 2, the online coders viewed a brief prompt and then rated a random selection of 10 abstracts from the same set of 194 anonymized abstracts from Study 1. Just as in Study 1, we calculated the political slant of each abstract by averaging the ratings from the online coders (and using individual coder ratings for random-effects models). In contrast to Study 1, the online coders did not undergo practice rounds and were not shown the example definitions of liberalism and conservatism or the flattering and unflattering liberal and conservative portraits. Instead, they were required to use their own sense of political ideology to guide their ratings. In addition, the political-slant scale was on a 7-point scale (1 = *consistent with a conservative worldview*, 4 = *unrelated to conservative or liberal worldviews*, 7 = *consistent with a liberal worldview*).[11] Unlike Study 1, Study 2 was unable to distinguish between an abstract lacking political relevance and one that was merely politically moderate. Moreover, the online coders did not rate each abstract on subdiscipline, contextual sensitivity, or robustness.

After rating the abstracts, the online coders completed an eight-item political-knowledge measure consisting of items typically used in the American National Election Studies (range = 0–8; higher scores indicated greater knowledge). This was followed by a measure of political engagement (i.e., interest in politics, importance of politics, summation of sources of political news and information; normalized score range = 0–1; higher scores indicate greater engagement; Malka, Soto, Inzlicht, & Lelkes, 2014). We then recorded age, gender, ethnicity, political ideology (1 = *very liberal*, 7 = *very conservative*), party identification, and educational attainment. These methodological differences between Study 1 and Study 2 were a result of the independent nature of our adversarial collaboration and allowed us to examine the generalizability and replicability of our findings under slightly different operationalizations determined by authors with different theoretical commitments.

Results

Agreement in political-slant rating

Study 1. The doctoral coders showed strong interrater reliability with respect to whether an abstract was politically relevant, ICC(3,6) = .84 (ICC = intraclass correlation; Shrout & Fleiss, 1979). When we examined the 52% (101/194) of abstracts deemed politically relevant, agreement of the precise political slant was lower, ICC(3,6) = .64, although still acceptable (Cicchetti, 1994; Koo & Li, 2016). This suggests that determinations of political relevance were easier to determine but that political-slant ratings were more challenging to ascertain—even among expert coders with prior training. Note that just 4% of the time, the doctoral coders were in a "majority disagreement" on the liberal–conservative direction of the political slant (these were cases in which at least a third of the coders said it was liberal and at least a third of the coders simultaneously said it was conservative, that is, when the majority of coders were split; for details, see Political Slant Agreement and Disagreement in the Supplemental Material). This suggests that although relative agreement was sometimes a challenge, very few abstracts resulted in the majority of doctoral coders producing ratings that classified an abstract's slant on opposite sides of the political spectrum.

Study 2. Agreement in political-slant rating was similar using the online coders' ratings. Because the 511 online coders rated a random selection of 10 abstracts from the database, each abstract received an average of 26 political-slant ratings in Study 2. Using Spearman-Brown's formula,[12] we found that the online coders showed similar levels of agreement for political slant, ICC(1,26) = .57. The online coders were in a majority disagreement just 7% of the time (operationalized the same way as in Study 1; for details, see Political-Slant Agreement and Disagreement in the Supplemental Material).

Distribution of political-slant ratings

Given the large proportion of liberal researchers in psychology, we first examined whether the political content of abstracts matched the political distribution of scientists from the field. Some estimates suggest that the liberal skew of psychologists themselves has an effect size (Pearson's r) ranging from .63 (Cardiff & Klein, 2005) to .89 (Langbert, Quain, & Klein, 2016; an effect size of 0 would imply an equal number of liberals and conservatives); social psychologists specifically produced an effect size (Pearson's r) of .87 (Inbar & Lammers, 2012; Von Hippel & Buss, 2017). However, both the doctoral-coder (Fig. 1) and online-coder[13] (Fig. 2) distributions of average political-slant ratings were fairly normal, with very few abstracts on the political extremes. The distributions of the abstracts were modestly shifted toward the political left, although the mean score was close to the midpoint: doctoral coders, M = 2.78 (midpoint of 3), $t(100) = -3.58$, $p < .001$, Pearson's $r = .34$ (20% of politically relevant abstracts were rated as liberal leaning, 4% were rated as conservative leaning); online coders, M = 3.85 (midpoint of 4), $t(193) = -5.51$, $p < .001$, Pearson's $r = .37$ (3% of all abstracts were rated as liberal leaning, 2% were rated as conservative leaning).[14] Thus, trained doctoral coders and untrained lay online coders rated psychology abstracts in a similar manner,[15] and the mean ideology of published articles, although slightly left of center, appeared to be quite different from the political makeup of scientists in the field.

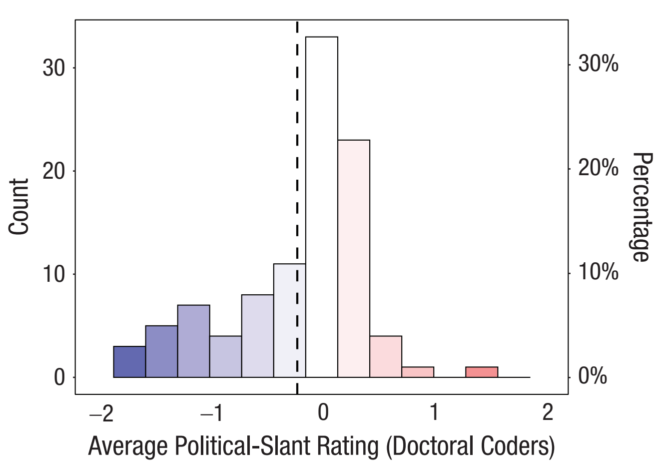

Fig. 1. Distribution of average political-slant ratings from Study 1's doctoral coders (centered on the midpoint of the scale). Scores of −2 reflect a very left-leaning abstract; +2 reflects a very right-leaning abstract. The dotted vertical line represents the mean. Skewness = −0.82, kurtosis = 3.44.

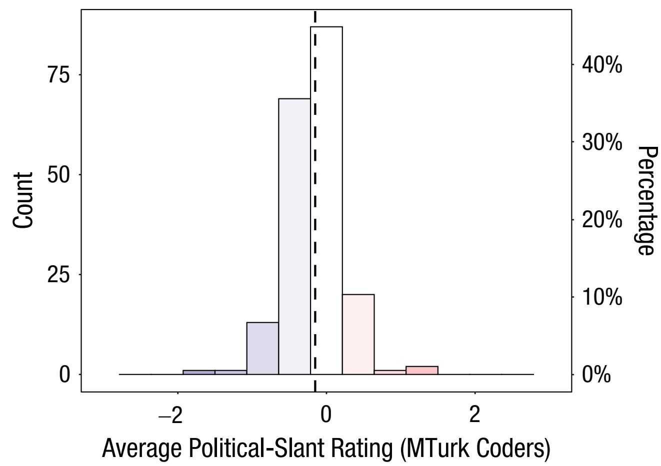

Fig. 2. Distribution of average political-slant ratings from Study 2's online coders (centered on the midpoint of the scale). Scores on the left (−3) reflect an abstract consistent with a liberal worldview, scores at 0 reflect an abstract that is unrelated to a conservative or liberal worldview, and scores on the right (+3) reflect an abstract consistent with a conservative worldview. Tbe dotted vertical line represents the mean. Skewness = 0.26, kurtosis = 5.25.

Political slant and replicability

Study 1. We next examined the relationship between the political slant of psychology results and the likelihood that the results were successfully replicated in subsequent research (as per our preregistration and similar to previous replication projects, replication was defined as a binary evaluation of whether research had been replicated). We conducted a mixed-model logistic regression for the politically relevant abstracts in which we estimated a random intercept and slope for each doctoral coder.[16] We found no evidence that political slant was associated with replicability, odds ratio (OR) = 1.03, SE = 0.10, p = .781, 95% confidence interval (CI) = [0.85, 1.24].[17] In addition, we performed a Bayesian analysis using Bayes factors (BF) based on the Bayesian information criterion. The analysis compared the null model with the fixed-slopes model and found a BF_{01} of 0.00002375277, which suggests that the null was 42,000 times more likely (Wagenmakers, 2007). Moreover, this null result remained consistent when we statistically adjusted for covariates previously shown to be related to replicability (e.g., effect

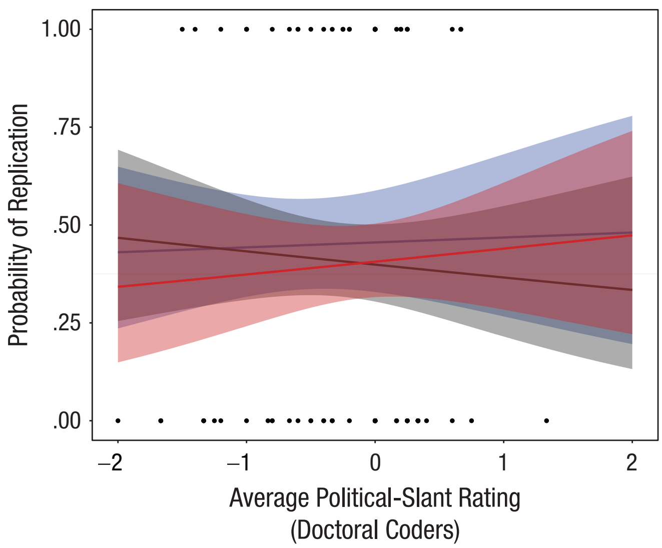

Fig. 3. Political slant predicting replicability, grouped by ideology of the coder. Binary logistic regression models showed that political slant was not significantly associated with replicability, regardless of the ideology of the doctoral coders (blue = liberals, gray = moderates, red = conservatives). For ease of interpretation, the unweighted fixed-effects model is shown here. The dots represent the political-slant rating of the studies that were replicated (top) and the studies that were not replicated (bottom). Shaded areas represent 95% confidence intervals. −2 = very left-leaning, 0 = moderate, 2 = very right-leaning.

size of the original research; see Model S1 in the Supplemental Material). Indeed, less than 1% of the variance in replicability could be explained by the political slant of the original research alone (Nakagawa-Schielzeth-Johnson [N-S-J] pseudo R^2 = .02%). Furthermore, the null association remained consistent regardless of the doctoral coder's ideology, all interaction ps > .342 (see Fig. 3 and Models S1a and S1b in the Supplemental Material). Indeed, we explored the zero-order correlations (Spearman) for each coder and found all to be nonsignificant: moderate female, $r = -.06$, $p = .543$; moderate male, $r = -.00$, $p = .970$; very conservative female, $r = -.02$, $p = .876$; very conservative male, $r = .07$, $p = .537$; very liberal female, $r = .11$, $p = .533$; very liberal male, $r = .04$, $p = .736$.

The overall results were also robust to various other models (e.g., fixed-effects models that used the average slant score for each abstract as well as exploratory weighted models that gave more weight to abstracts that had more ratings—a coder's rating of "does not apply" was not used to calculate the average slant score, so some abstracts have fewer ratings—or had more agreement among the ratings; see Models S2–S6 in the Supplemental Material). In addition, the null association remained consistent when we focused solely on social or personality psychology abstracts,[18] $OR = 1.04$, $SE = 0.10$, $p = .667$, 95% CI = [0.86, 1.26]. As a further robustness check, we ran the same fixed- and random-effects models including all abstracts—not just the politically relevant subset—and found similar results (see Models S7–S12 in the Supplemental Material). Thus, testing numerous models, we found no support for a liberal bias (or conservative bias) with respect to a specific political slant and replicability.

Study 2. We performed the same mixed-model logistic regression as in Study 1, this time using the online coders' ratings. Replicating our results from Study 1, we did not find evidence that political slant was associated with replicability, $OR = 0.98$, $SE = 0.02$, $p = .432$, 95% CI = [0.94, 1.03]. This result remained null when adjusting for covariates previously shown to be related to replicability (e.g., effect size of the original research; see Model S21 in the Supplemental Material). Less than 2% of the variance in replicability could be explained by the political slant of the original research alone (N-S-J pseudo R^2 = 1.95%). Moreover, the null association remained consistent regardless of the online coder's ideology, all interaction ps > .464 (see Fig. 4; see also Models S21a and S21b in the Supplemental Material).

The overall results were also robust to various other models (the same model variations as done for Study 1, although giving more weight to abstracts that had more ratings revealed a significant effect when we did not adjust for covariates; see Models S22–S26 in the Supplemental Material). In addition, the null association did not change when we focused only on social or personality

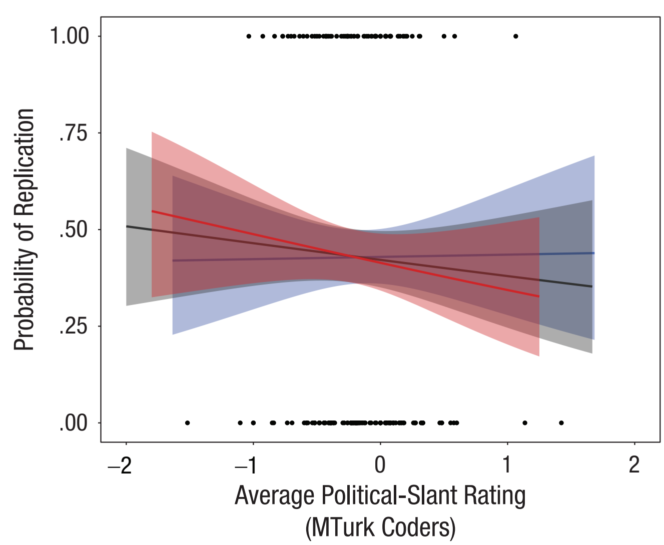

Fig. 4. Political slant predicting replicability, grouped by ideology of the coder. Binary logistic regression models showed that political slant was not significantly associated with replicability regardless of the ideology of the online coders (blue = liberals, gray = moderates, red = conservatives). For ease of interpretation, the unweighted fixed-effects model is shown here. The dots represent the political-slant rating of the studies that were replicated (top) and the studies that were not replicated (bottom). Shaded areas represent 95% confidence intervals. −3 = consistent with a liberal worldview, 0 = unrelated to conservative or liberal worldview, 3 = consistent with conservative worldview.

psychology abstracts,[19] $OR = 0.96$, $SE = 0.03$, $p = .161$, 95% CI = [0.92, 1.01]. Thus, testing numerous models, we replicated our null results from Study 1 and found no evidence of a liberal bias (or conservative bias) with respect to a specific political slant and replicability.

Political slant and statistical robustness

Whereas replicability is a cornerstone of the scientific method and an overall measure of the robustness of research findings, replication success is due to numerous factors. Therefore, we examined the relationship between political slant and several objective measures of the robustness of the original research. For example, if liberal reviewers and editors are more prone to overlook statistical red flags (e.g., tiny sample sizes or weak effects) when a research finding accords with their own personal political ideology (or conversely, increase the standards of evidence required when reviewing research that clashes with their own ideology), then liberal-leaning research in the literature should be associated with smaller sample sizes or weaker effects. Thus, in several exploratory analyses, we examined whether political slant was associated with objective indices of statistical robustness, such as the sample size and effect size of the original research.

Using the doctoral coders' ratings, political slant was not significantly associated with the effect sizes of the original research (Spearman's $r = -.07$, $p = .461$), and this null result was replicated using the online coders' ratings (Spearman's $r = .00$, $p = .976$). Using the doctoral coders' ratings, we found that political slant was not significantly associated with the sample size of the original research (Spearman's $r = .17$, $p = .091$); if anything, this pattern flipped when using the online coders' ratings (Spearman's $r = -.14$, $p = .055$). Because of the opposite findings, which did not achieve statistical significance, there is no clear evidence for a relationship between political slant and sample size. In sum, there is no obvious relationship between these measures of statistical robustness and political slant. Across a diverse range of abstracts and coding performed by both trained doctoral-level students and untrained lay online coders, we did not find evidence that replicability or statistical robustness of psychological science is significantly associated with the political slant of the research.

The lack of association between political slant and our various measures of statistical robustness can also be seen in Figure 5 (doctoral coders) and Figure 6 (online coders). Although most original findings were statistically significant (density plot along top edge),

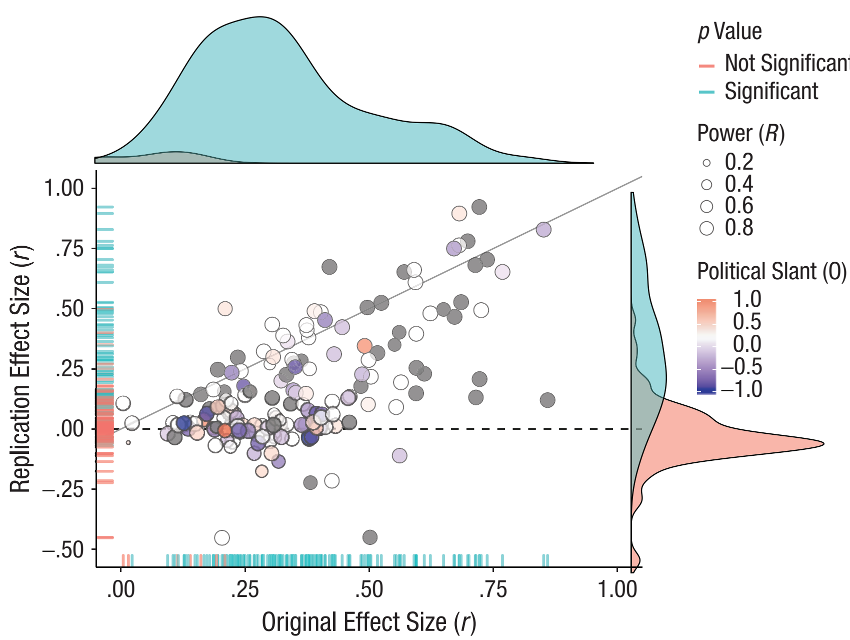

Fig. 5. Original-study effect size compared with replication effect size (correlation coefficients) according to doctoral coders in Study 1. The diagonal line represents a replication effect size equal to the original effect size. The dotted line represents a replication effect size of 0. The points below the dotted line are effects in the opposite direction of the original. The density plots are separated by significant (aqua green) and nonsignificant (light red) effects. The normalized political slant of data points ranges from very left-leaning (−1, blue), to moderate (0, white), to very right-leaning (1, red); gray data points reflect abstracts that are not politically relevant.

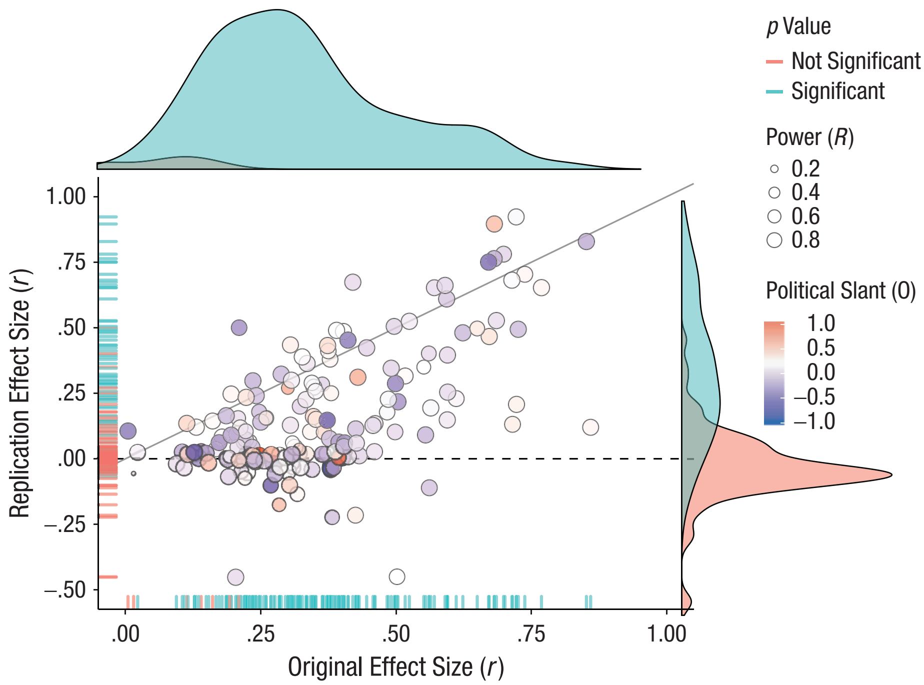

Fig. 6. Original-study effect size compared with replication effect size (correlation coefficients) according to online coders in Study 2. The diagonal line represents a replication effect size equal to the original effect size. The dotted line represents a replication effect size of 0. The points below the dotted line are effects in the opposite direction of the original. The density plots are separated by significant (aqua green) and nonsignificant (light red) effects. The normalized political slant of data points ranges from consistent with a liberal worldview (−1, blue), to unrelated to conservative or liberal worldviews (0, white), to consistent with a conservative worldview (1, red).

most replication attempts were not (density plot along right edge), which yielded an overall replication rate of 42%. Correspondingly, replication effect sizes tended to be smaller than the original effect sizes (data points falling below the diagonal line), although the two were significantly positively correlated, Pearson's $r(190) = .56$, $p < .001$, which indicates that the size of the original findings largely predicted the size of the replication findings. Political slant, however, shows no clear relationship to these measures of statistical robustness.

Political slant and postpublication impact

Our data did not provide evidence that political slant is associated with replicability or statistical robustness. However, some critics have argued that liberal findings are cited and discussed more often than conservative findings (Jussim et al., 2016). Therefore, we conducted exploratory analyses to determine whether political slant predicted citation counts and Altmetric scores[20] (a measure of how widely an article is discussed online, e.g., in public-policy documents, mainstream media, blogs, Facebook, Twitter, and Wikipedia) of the original research. Using the doctoral coders' ratings, we found that political slant was not associated with citation counts (Spearman's $r = -.06$, $p = .413$; Fig. 7) or Altmetric scores (Spearman's $r = -.005$, $p = .948$; Fig. 8). These null results replicated using the online coders' ratings for both citations (Spearman $r = -.07$, $p = .314$; Fig. 9) and Altmetric scores (Spearman's $r = -.04$, $p = .541$; Fig. 10). Instead, a few seminal articles (e.g., Tversky's research on decision-making) received the bulk of citations, and a few popular findings (e.g., Rand's research on intuitive cooperation) received the most online attention,[21] and this appears to be unrelated to the political content of the research.

Political extremity and replicability

Although the specific political slant (i.e., liberal vs. conservative) of psychology research was not related to replicability, it is possible that research with more of a political slant, regardless of whether the research is slanted toward liberal or conservative, is less replicable. To test this possibility, we midpoint-centered and computed the absolute value of each coder's rating for every abstract (to create a measure of *political extremity*) and performed a mixed-model logistic regression that estimated a random intercept and slope for each coder.

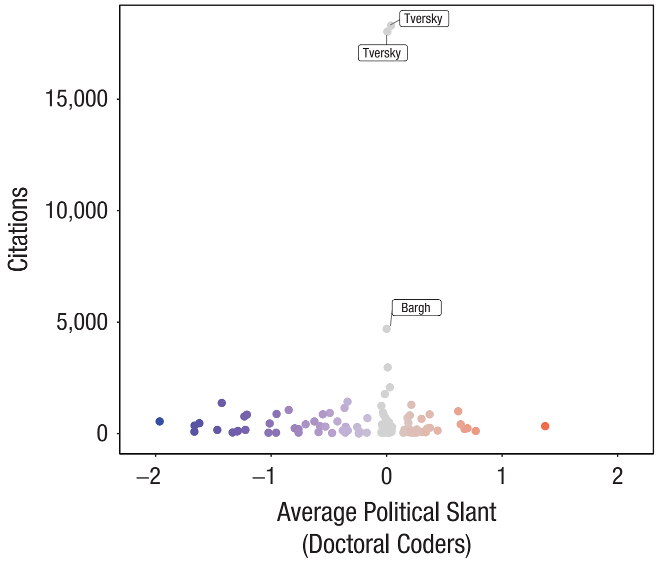

Fig. 7. Average political-slant ratings from Study 1's doctoral coders (centered on the midpoint of the scale) and citation counts of the original research. Scores of −2 reflect a very left-leaning abstract; +2 reflects a very right-leaning abstract. The ideology of the finding was unrelated to citations.

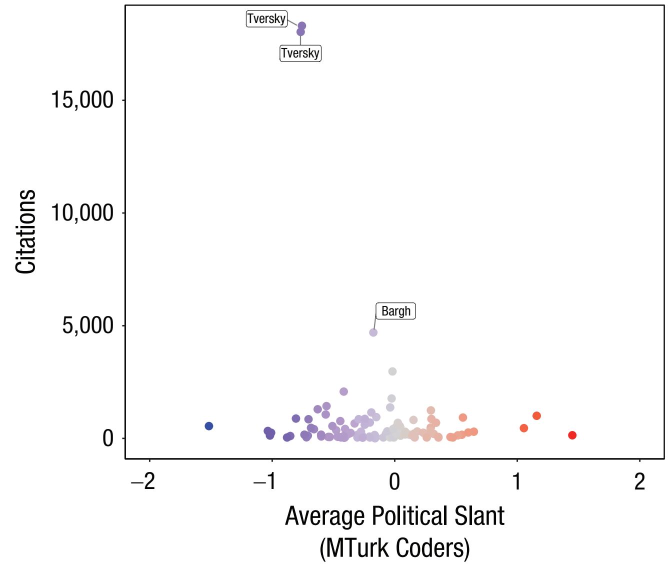

Fig. 9. Average political-slant ratings from Study 2's online coders (centered on the midpoint of the scale) and citation counts of the original research. Scores on the left (−3) reflect an abstract consistent with a liberal worldview, scores at 0 reflect an abstract that is unrelated to a conservative or liberal worldview, and scores on the right (+3) reflect an abstract consistent with a conservative worldview. The ideology of the finding was unrelated to citations.

Specifically, we centered each abstract's rating on the midpoint of the scale (i.e., for Study 1, which used a 1–5 scale, we subtracted 3 from each abstract's rating). Thus, a rating of 3 (a moderate abstract) would become a 0, a rating of 1 (a liberal abstract) would become −2, and a rating of 5 (a conservative abstract) would become +2. Then we took the absolute value of these midpoint-centered ratings. This gave us a measure of how politically extreme an abstract was overall regardless of whether it was leaning in the liberal or conservative direction.

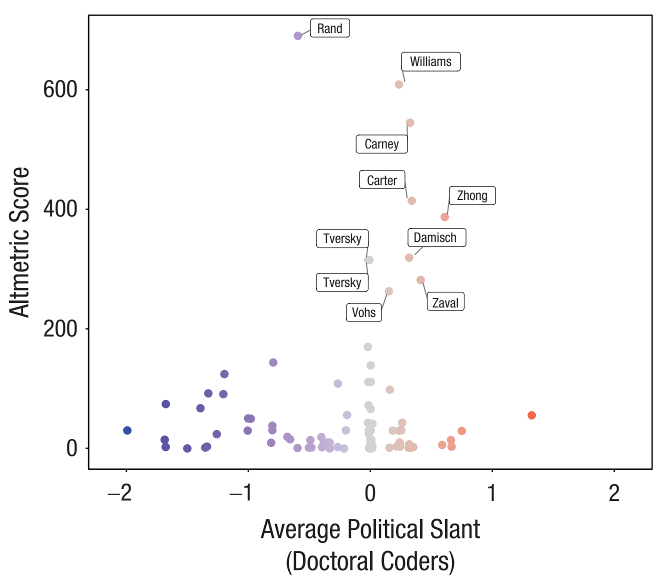

Fig. 8. Average political-slant ratings from Study 1's doctoral coders (centered on the midpoint of the scale) and Altmetric scores of the original research. Scores of −2 reflect a very left-leaning abstract; +2 reflects a very right-leaning abstract. The ideology of the finding was unrelated to the Altmetric score.

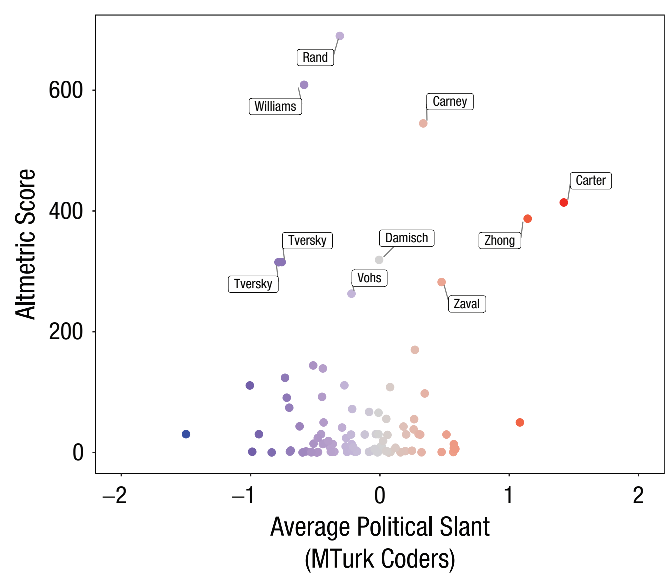

Fig. 10. Average political-slant ratings from Study 2's online coders (centered on the midpoint of the scale) and Altmetric scores of the original research. Scores on the left (−3) reflect an abstract consistent with a liberal worldview, scores at 0 reflect an abstract that is unrelated to a conservative or liberal worldview, and scores on the right (+3) reflect an abstract consistent with a conservative worldview. The ideology of the finding was unrelated to Altmetric score.

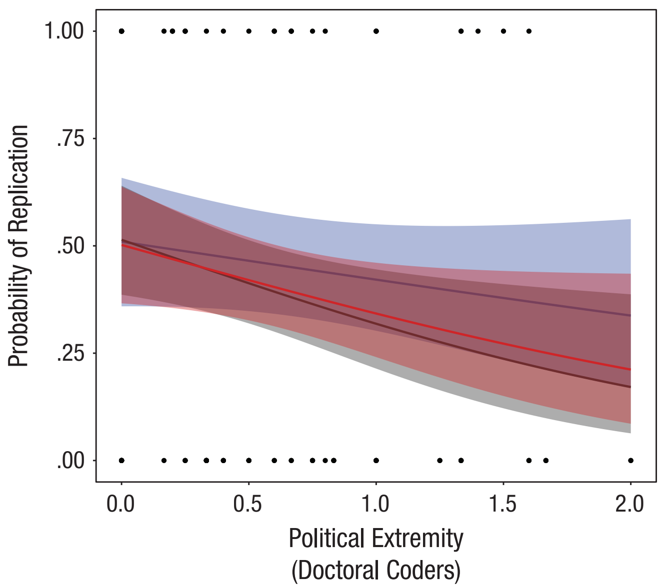

Fig. 11. Political extremity predicting replicability, grouped by ideology of the coder. Binary logistic regression models show that political extremity was significantly associated with replicability, although only among moderates and conservative doctoral coders (blue = liberals, gray = moderates, red = conservatives). For ease of interpretation, the unweighted fixed-effects model is shown here. The dots represent the political-slant rating of the studies that were replicated (top) and the studies that were not replicated (bottom). Shaded areas represent 95% confidence intervals.

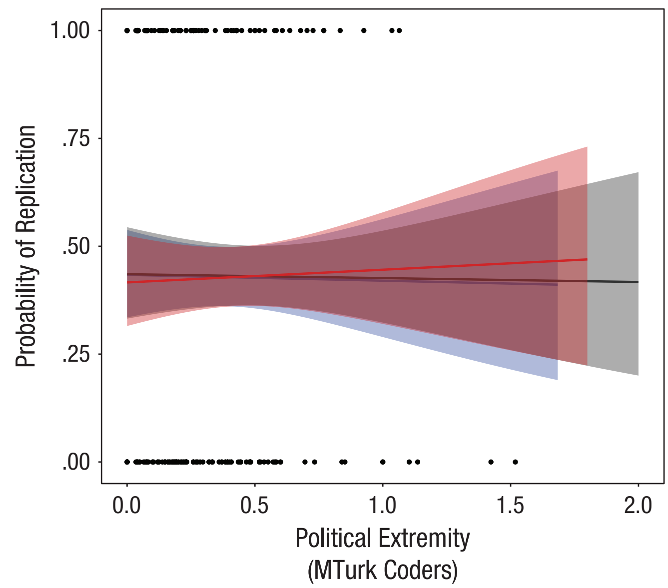

Fig. 12. Political extremity predicting replicability, grouped by ideology of the coder. Binary logistic regression models show that political extremity was not consistently significantly associated with replicability, regardless of the ideology of the doctoral coders (blue = liberals, gray = moderates, red = conservatives). For ease of interpretation, the unweighted fixed-effects model is shown here. The dots represent the political-slant rating of the studies that were replicated (top) and the studies that were not replicated (bottom). Shaded areas represent 95% confidence intervals.

Study 1. We found a statistically significant association such that abstracts describing more politically extreme research were less likely to be replicable, $OR = 0.66$, $SE = 0.12$, $p < .001$, 95% CI = [0.52, 0.85]. This result also held when adjusting for covariates previously shown to be related to replicability (e.g., effect size of the original research; see Model S13 in the Supplemental Material). Focusing solely on social or personality abstracts continued to show evidence of an association: $OR = 0.64$, $SE = 0.13$, $p < .001$, 95% CI = [0.50, 0.82].

These results were also robust to various model specifications (the same model variations as done when testing political slant), although our originally preregistered unweighted fixed-effects model subset on politically relevant abstracts was not quite significant (see Fig. 11; see also Models S14–S20 in the Supplemental Material). However, all other fixed-effects, random-effects, and exploratory weighted models revealed a significant effect both when political extremity was a sole predictor and when adjusting for covariates. Taken together, these data suggest that research with greater political slant—whether liberal or conservative—is associated with reduced replicability.

Study 2. These findings were partially replicated using the online coders' ratings. We found a statistically significant effect such that abstracts describing more politically extreme research were less likely to be replicable, $OR = 0.94$, $SE = 0.03$, $p = .036$, 95% CI = [0.88, 1.00]. This result was not quite significant after adjusting for covariates previously shown to be related to replicability (e.g., effect size of the original research; see Model S27 in the Supplemental Material). Moreover, focusing solely on social or personality abstracts did not show evidence of an effect: $OR = 1.01$, $SE = 0.04$, $p = .811$, 95% CI = [0.94, 1.08].

In addition, as in Study 1, the unweighted fixed-effects model did not show a significant effect (see Fig. 12). However, other exploratory models revealed a significant effect (weighted fixed effect and both unweighted and weighted random effect; see Models S28–S30 in the Supplemental Material) when political extremity was a sole predictor and across the weighted models when adjusting for covariates. Taken together, although the political extremity effect appeared robust according to the doctoral coders' ratings, it was inconsistent according to the online coders' ratings, and thus we urge caution in drawing strong conclusions given the mixed results.

Political extremity and statistical robustness

Mimicking our analyses of political slant, we also conducted exploratory analyses to examine whether political extremity was associated with statistical robustness of the original research. Using the doctoral coders' ratings, we found that political extremity was not associated with the effect size of the original research (Spearman's $r = -.05$, $p = .624$), and this null result was replicated using the online coders' ratings (Spearman's $r = -.07$, $p = .357$).

Using the doctoral coders' ratings, we found that political extremity was significantly negatively associated with the sample size of the original research (Spearman's $r = -.34$, $p < .001$), which suggests that research with greater political slant is associated with smaller sample sizes. However, this result was in the opposite direction when using the online coders' ratings (Spearman's $r = .17$, $p = .015$), which suggests that research with greater political slant is associated with larger sample sizes. Thus, political extremity was not consistently related to the effect size or sample size of the original research.

General Discussion

The current research examines a contentious issue: Can the political composition of a scholarly field undercut the scientific rigor of the research? To address this question, we analyzed a set of nearly 200 psychology studies and subsequent replication attempts. Although there are many more psychologists who are liberal, the results in the literature did not completely mirror the heavy political skew of psychologists. Whereas there were more findings consistent with a liberal worldview than a conservative worldview, the average ideology of research was fairly centrist, and the majority of research was either nonpolitical (48% according to doctoral-coder ratings) or politically relevant but without a clear political slant (74% among the politically relevant subset, according to doctoral-coder ratings). More importantly, liberal findings were just as likely to be replicable and, in exploratory analyses, were as statistically robust as conservative findings and as likely to be cited or mentioned in the media. These results remained consistent across both liberal, moderate, and conservative coders; expert and lay coders; and when numerous covariates known to account for replicability were added to our statistical models.

Instead, we found mixed evidence of a *political extremity* effect, such that research that was more politically slanted (regardless of liberal vs. conservative political slant) was between 34% (Study 1) and 6% (Study 2) less likely to be replicated. These results were stable across all model specifications that statistically adjusted for variables associated with statistical robustness in Study 1, but in Study 2, the effect of political extremity was reduced when statistically adjusting for variables related to statistical robustness. On one hand, the preliminary political extremity effect in our data accords with concerns about highly politicized research (Tetlock, 1994). On the other hand, the majority of our analyses suggest that statistical robustness is the consistent predictor of replicability rather than the political slant or extremity of a research topic. These results suggest that it is important to focus on study characteristics such as sample size and effect size to help improve replicability. Moreover, we urge caution in interpreting our political extremity effect given that most studies in our database were not ideologically extreme and that there may be a restriction of range.

Although we did not find evidence of a liberal bias in scientific replicability in these data, perceptions of political bias still exist both among laypeople (Hannikainen, 2019) and academics. For example, Eitan and colleagues (2018) found that academics (students and professors) believed that personal political beliefs slightly bias scientific research (Pearson's $r = .62$) and that social psychology is biased against conservatives (Pearson's $r = .83$). The fact that we find some evidence of a political extremity effect coupled with the fact that there tends to be more liberal-leaning research in psychology offers one possible explanation. If highly political research is less replicable but people are sampling only one side of the political spectrum because of a shifted distribution of published research, it would appear rational to arrive at the conclusion that liberal-leaning research is less robust. Of course, our findings suggest that such an asymmetric sampling may inadvertently miss the possibility that the root cause is in fact a symmetric political bias in scientific replicability.

Moreover, when scientists use the word *bias*, they often mean different things at different times. For example, bias may refer to the skewed political distribution of psychologists themselves or the possible tendency to study certain topics (although our distributions of political slant were fairly normal). Whereas Duarte and colleagues (2015) hypothesized reasons for the large number of liberals in the field (see also Haidt, 2011), our data suggest that the political skew of psychologists is not tightly coupled with the political skew of the literature itself, and future work should seek to disentangle these discrepancies.

In other instances, bias might refer to the systematic tendency to evaluate research differently on the basis of its political slant. Duarte and colleagues (2015) cited evidence of such a peer-review political bias from Abramowitz, Gomes, and Abramowitz (1975). Yet that study had methodological shortcomings,[22] and even the authors themselves admitted that "the amount of bias detected might be so slight as to be meaningless in the real world of publish or perish" (p. 193). In fact, similar tentative evidence of peer-review bias has even been found against liberal-leaning diversity research (King, Avery, Hebl, & Cortina, 2018) and gender-bias research

(Cislak, Formanowicz, & Saguy, 2018). This suggests that both liberal and conservative perspectives may experience subtle bias but that, overall, the peer-review process mitigates most egregious instances of political favoritism.[23]

Still, some data suggest that discrimination based on the political orientation of research may exist. For instance, a survey of 292 members of the Society for Personality and Social Psychology found that respondents self-reported a willingness to discriminate against a hypothetical grant application or manuscript, at least to some minimal extent (i.e., chose a scale point above "not at all"), if there was a feeling that it took a "politically conservative perspective" (Inbar & Lammers, 2012). However, our data suggest that tenuous liberal research does not systematically find its way into the published literature. Duarte and colleagues (2015) acknowledged that "the lack of political diversity is not a threat to the validity of specific studies in many and perhaps most areas of research in social psychology" (p. 2). It remains possible that their claims may apply to a very small subsection of psychology, if any, given that we found no evidence that research aligned with a majority viewpoint (liberalism) was less replicable or less statistically robust than research aligned with a minority viewpoint (conservatism).

Although our data, to our knowledge, provide the first test of whether the political slant of research is associated with scientific replicability, there are a number of limitations to our work. One important limitation is that our sample was not a random sample of the entire field of psychological research. Although the largest database we used was intentionally designed to sample a relatively representative group of high-impact psychology articles (OSC, 2015), our sample was nevertheless limited to studies for which replication data were readily available and thus was not completely representative of the entire field. Although selection biases could occur, there are at least two possible counterarguments mitigating this issue. First, given that psychologists have historically prioritized surprising results, it might be more likely that replicators would choose studies that surprised them or about which they were skeptical (e.g., studies that did not align with their own personal political ideology). With a predominantly liberal field, conservative findings would be most surprising. Second, many of the studies selected for replication were chosen because they represent some of the most influential findings in psychology (e.g., APS RRRs) or were specifically chosen to reflect a range of effects and contexts (e.g., Many Labs). Therefore, replicators made explicit efforts to identify a combination of important and representative research. Future research should examine whether these findings extend to other areas of psychology as well as other social sciences because larger and more representative samples will be more likely to produce generalizable knowledge. In addition, our analyses examining the association between political slant and statistical robustness or postpublication impact could also be performed on a much broader swath of the literature if future scholars are willing to code political slant for more studies. Such an analysis would be useful for future research.

A second important limitation is that measuring political slant is challenging. Labels such as "liberal" or "conservative" may be too broad to capture the nuanced ideologies and assorted political attitudes of people (Ditto et al., 2018). For example, we did not differentiate among the political slants for social, economic, or foreign-policy subcategories (Inbar & Lammers, 2012). In addition, we decided on using the binary political spectrum of American politics, which is quite common (Inglehart & Klingemann, 1976; Jost, Glaser, Kruglanski, & Sulloway, 2003) but is still debated among political scientists (Feldman & Johnston, 2014). Moreover, political contexts and relative ideologies rapidly shift, and what we refer to as liberal today may differ from its usage 50 years ago (e.g., many older liberals might claim that current mainstream liberals are quite moderate relative to the 1960s). Thus, our data speak to the current construction of American politics. In addition, the specific political slant of many studies was not clear-cut in many cases, as reflected by our lower political-slant reliability across coders. This suggests that debates about political bias may hinge on idiosyncratic definitions rather than a clear, shared definition of ideology that can be easily observed and coded. Future research should further clarify political slant and continue to pursue additional operationalizations of political slant to accumulate evidence.

It is unclear whether the personal political beliefs of scientists have a measurable influence on the replicability and robustness of the published literature. The peer-review process may be sufficient to weed out most political manuscripts that are not backed by sufficient scientific data, and scientists may be more motivated by scientific identities and norms when they are writing and reviewing manuscripts (Merton, 1973; Van Bavel et al., 2020). In fact, the identity of *scientist* is more likely to be salient during this process, which can help reduce motivated political cognition (Van Bavel & Pereira, 2018). These features can help mitigate political groupthink (Van Bavel et al., 2020). However, bias of various forms can still emerge, and it is unclear from our data when people can or do pursue value-free science (Longino, 1990; Richardson & Polyakova, 2012; Rykiel, 2001; Sears, 1994). Indeed, we are all shaped by our experience, and science cannot avoid at least some aspect of subjectivity. For example, it is possible that any ideological biases that affected the original research (e.g., measurement strategies) were simply carried over in a direct replication, which yields subjective bias on both ends of the scientific process. Thus, we believe that pursuing adversarial collaborations (Shi, Teplitskiy, Duede, & Evans, 2019) and performing "turnabout" tests, wherein a hypothesis is inverted to test a reverse claim, may be a helpful guard against confirmation bias and groupthink (Duarte et al., 2015; McGuire, 1997; Washburn & Skitka, 2018).

The current research also speaks to the quality and replicability of research more broadly. Sparked by difficulty in replicating findings in genetics (Hirschhorn, Lohmueller, Byrne, & Hirschhorn, 2002), pharmacology (Prinz, Schlange, & Asadullah, 2011), oncology (Begley & Ellis, 2012), biology (Reaves, Sinha, Rabinowitz, Kruglyak, & Redfield, 2012), psychology (OSC, 2015), and economics (Chang & Li, 2015), researchers have turned the microscope on themselves and started a dialogue about best research practices. Some of science's most well-known journals (e.g., *Nature* and *Science*) and funding agencies have called for more replications and implemented new procedures to enhance the robustness of published research (Baker, 2016; Bollen, Cacioppo, Kaplan, Krosnick, & Olds, 2015; McNutt, 2014; *Nature*, 2013, 2017). Many factors reduce replicability, including the publication of false positives (Cohen, 1992; Simmons, Nelson, & Simonsohn, 2011), publication bias (Ferguson & Heene, 2012), and low-fidelity replications (Gilbert, King, Pettigrew, & Wilson, 2016). The current research suggests that research that is more politically extreme may be another factor associated with reduced replication rates. However, this preliminary result applies equally to liberal and conservative findings and may be partially related to the contextual sensitivity of highly political findings (Crawford, Vodapalli, Stingel, & Ruscio, 2019; Van Bavel et al., 2016). Indeed, several politically relevant effects have fluctuated over time (e.g., flag priming, the "Obama effect"), and meta-analyses that have examined politically loaded topics have revealed a large degree of heterogeneity in effect sizes, possibly because findings change as a function of the broader political context (e.g., legislation, governmental actions; e.g., Tankard & Paluck, 2017). Future work would do well to include larger samples of replication studies as they become available to see if our findings are robust under conditions of increased statistical power.

Taken together, our results are a starting point for a richer conversation about the role and influence of politics in science. It seems that our intuitions about political bias may at times be imprecise (Eitan et al., 2018). Our findings provide clear evidence that statistical robustness (e.g., sample size and effect size) is a consistent predictor of replicability rather than the political slant or extremity of a research topic. Thus, it might be more fruitful to shift our focus from the politics of scientists to their research practices. We hope other researchers can build off of our work because these issues are critical for scientists' epistemological pursuits.

Transparency

Action Editor: Laura A. King
Editor: Laura A. King
Author Contributions
D. A. Reinero, J. A. Wills, W. J. Brady, P. Mende-Siedlecki, J. T. Crawford, and J. J. Van Bavel conceptualized the research question and designed the experiments. D. A. Reinero, J. A. Wills, W. J. Brady, P. Mende-Siedlecki, and J. J. Van Bavel wrote the preregistration for Study 1. J. T. Crawford wrote the preregistration for Study 2. D. A. Reinero created the replication database. D. A. Reinero and J. A. Wills collected the data for Study 1. J. T. Crawford collected the data for Study 2. D. A. Reinero, J. A. Wills, and W. J. Brady analyzed the data for Study 1. D. A. Reinero and J. T. Crawford analyzed the data for Study 2. D. A. Reinero drafted the manuscript, and W. J. Brady, P. Mende-Siedlecki, J. T. Crawford, and J. J. Van Bavel provided critical revision. All the authors approved the final manuscript for submission.
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
Funding
This work was supported by National Science Foundation Graduate Research Fellowship DGE 1342536 (awarded to D. A. Reinero) and National Science Foundation Grant 1349089 (awarded to J. J. Van Bavel).

ORCID iDs

Diego A. Reinero https://orcid.org/0000-0002-6124-2623
J. J. Van Bavel https://orcid.org/0000-0002-2520-0442

Acknowledgments

We thank Jonathan Haidt and members of the New York University Social Identity and Morality Lab for insightful comments on this manuscript (in particular, Philip Pärnamets for providing helpful statistical counsel). We are grateful to the many researchers who made data from several repositories publicly available, particularly teams from the Open Science Center. In addition, we thank Chris Hartgerink for sharing his data analysis scripts as well as Suraiya Allidina, Anna Balchunas, Matthew Leipzig, and Simone Van Taylor for their assistance in developing study materials and helping build the replication database. We also thank (in alphabetical order) Rhonda Balzarini, Mark Brandt, Robert Calin-Jageman, Evan Carter, Joseph Cesario, Katherine Corker, Katinka Dijkstra, Irena Domachowska, Brent Donnellan, Fredrick "Andy"

Eichler, Christopher French, Katie Garrison, Carolyn Gibson, Martin Hagger, Karl Healey, Kevin Holmes, Jay G. Hull, David Johnson, Florian Lange, Richard Lucas, John Lurquin, Florian Müller, Sanne Nauts, Leif Nelson, Eva Ranehill, Stuart Ritchie, Brent Roberts, Doug Rohrer, Daniel Simons, Collen Sinclair, Niels van de Ven, Peter Verkoeijen, Ivar Vermeulen, Roberto Weber, Deena Weisberg, Eric Wesselmann, and Iris Žeželj for their help in confirming key analyses. We note that the adversarial collaboration in the present article was such that one "side" was led by J. J. Van Bavel at New York University (including his current and former lab members D. A. Reinero, J. A. Wills, W. J. Brady, and P. Mende-Siedlecki) and the other "side" by J. T. Crawford at The College of New Jersey.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/1745691620924463

Notes

1. Although political party identification (Democrats) and ideology (liberal) are different and have diverged at various times in the past (Barber & McCarty, 2015), currently, the Democratic party trends liberal/left-leaning, and the Republican party trends conservative/right-leaning (Levendusky, 2009).
2. The other two departments were law (8.6:1) and journalism (20:1). The overall ratio across all universities and departments was 11.5 Democrats to 1 Republican. The sample size from the report was 7,243 faculty.
3. Exact percentages vary. For example, when focusing on economic issues, Inbar and Lammers's (2012) survey suggested less of an imbalance: 63.2% liberal and 17.9% conservative.
4. That said, scientific beliefs are instilled through the accumulation of evidence, and scientists may merely be acting as rational Bayesian agents when challenges to current beliefs are presented.
5. Abstracts were used as proxies for original articles. Coding abstracts in this way has previously been used as a valid way to code for content (e.g., Eitan et al., 2018; Handley et al., 2015; King et al., 2018; Van Bavel et al., 2016). In Study 1, all coders rated all 194 abstracts. In Study 2, each coder rated a random sample of 10 abstracts.
6. Two original articles from the Curate Science repository did not have replication analyses that sufficiently matched the original to compare the two. Those two replication attempts were excluded.
7. Some original articles had multiple labs attempting to replicate a single effect (e.g., APS RRR). Other original articles contained multiple effects that the replicator or replicators attempted to replicate. For such instances, aggregate or meta-analytic effects are presented when provided in the replication article. Otherwise, when theoretically reasonable, multiple replication attempts/effects corresponding to the same original article were averaged and weighted by sample size.
8. Indeed, our sample fully encapsulates the largest replication article in the history of the field (OSC, 2015).
9. Our random selection was limited to respondents who had self-identified as either "very liberal," "moderate," or "very conservative" to provide maximal balance across the political spectrum.
10. If the doctoral coders felt more than one subdiscipline applied, they were asked to select the two most relevant.
11. For ease of interpretation and to match Study 1, we reverse-scored the political-slant scale from Study 2 such that 1 = *consistent with a liberal worldview* and 7 = *consistent with a conservative worldview*.
12. The default calculation for ICC(1,k) is to assume that k = the total number of raters. That is appropriate if every rater gives a judgment on every abstract (as in Study 1), but this is not the case for Study 2. Thus, we used Spearman-Brown's formula: $(k \times r) / (1 + (k - 1) \times r)$, where k = the number of ratings per abstract and r = the single rater reliability derived from an ICC that assumes each rater gave a judgment on every abstract (see Shrout & Fleiss, 1979).
13. Because Study 2's scale did not allow for distinguishing abstracts that were politically relevant, the distribution of political-slant ratings from Study 2 includes all abstracts in the database (N = 194), whereas Study 1's distribution stems from the subset of articles deemed politically relevant on the basis of the doctoral coder's ratings (n = 101).
14. Each study's political-slant scale was divided into thirds to create a "liberal bin," a "moderate bin," and a "conservative bin."
15. That Study 1 and Study 2's distributions are similar is supported by the fact that the two sets of political-slant ratings are significantly positively correlated with each other, Spearman r = .29, p < .001 (across all abstracts from which computing a political-slant score was possible, i.e., when all six doctoral coders from Study 1 rated an abstract as does not apply, that abstract could not receive a political-slant score).
16. Our preregistration originally stated that we would use the average political-lean score for each abstract in a fixed-effects model; however, to avoid losing meaningful variance, we updated to run random-effects models. The results are consistent either way; see Models S2 through S6 in the Supplemental Material.
17. This model provides a warning of a singular fit, so we also performed a Bayesian model that provided nearly identical estimates, and thus we retain the reported mixed model above.
18. As per our preregistration, any abstract in which at least one doctoral coder had categorized the study as related to social or personality psychology was counted as such. These results hold even when changing the subdiscipline categorization to be based more on consensus such that any abstract receiving three or more votes for a given subdiscipline (i.e., at least half of the coders) was counted as such, OR = 1.07, SE = 0.10, p = .489, 95% CI = [0.88, 1.29].
19. The subdiscipline of each abstract was borrowed from the aforementioned rating system done by the doctoral coders in Study 1.
20. Scraping of citation counts (via Google Scholar) and Altmetric scores (via Altmetric) conducted as of November 21, 2018.
21. We also note that articles published before social media became popular (around 2004) likely suffer from not having had the opportunity to be shared widely online, although only ~22% of articles were before 2004.

22. Abramowitz et al. (1975) asked research psychologists to rate the suitability of a manuscript for publication. The methods and analyses were held identical for all reviewers, but the main finding was *presumably* ideologically congruent or incongruent for a left-leaning person. Most problematically, the ideology of the reviewers was *assigned* by Abramowitz et al. on the basis of the reviewer's past contributions to certain journals or membership with certain societies, which may be a noisy measure of reviewer ideology. In addition, most measures of bias were nonsignificant or marginal. Finally, this study was conducted nearly 45 years ago, so contexts may have now shifted.
23. Although subtle political bias on either side should not be ignored, there are heated and unresolved debates as to the source of possible forms of bias—and if it should even be considered a bias—and such a lengthy explication is beyond the scope of this article (Baron & Jost, 2018).

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.

al Gharbi, M. (2018). Race and the race for the White House: On social research in the age of Trump. *The American Sociologist*, *49*, 496–519.

Armstrong, J. S. (1997). Peer review for journals: Evidence on quality control, fairness, and innovation. *Science and Engineering Ethics*, *3*, 63–84. doi:10.1007/s11948-997-0017-3

Baker, M. (2016). *Dutch agency launches first grants programme dedicated to replication*. Retrieved from https://www.nature.com/news/dutch-agency-launches-first-grants-programme-dedicated-to-replication-1.20287

Barber, M., & McCarty, N. (2015). Causes and consequences of polarization. In N. Persily (Ed.), *Political negotiation: A handbook* (pp. 15–58). Cambridge, England: Cambridge University Press.

Baron, J., & Jost, J. T. (2018). False equivalence: Are liberals and conservatives in the US equally "biased?" *Perspectives on Psychological Science*, *14*, 292–303.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Begley, C. G., & Ellis, L. M. (2012). Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531–533.

Bollen, K., Cacioppo, J. T., Kaplan, R. M., Krosnick, J. A., & Olds, J. L. (2015). *Social, behavioral, and economic sciences perspectives on robust and reliable science*. Retrieved from https://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf

Brooks, A. (2015 October 31). Academia's rejection of diversity. *The New York Times*. Retrieved from https://www.nytimes.com/2015/10/31/opinion/academias-rejection-of-diversity.html

Brugh, M. (2017a). *Onderzoek naar linkse wetenschap 'zinnig'* [Research into left-wing science "makes sense"]. Retrieved from https://www.nrc.nl/nieuws/2017/02/08/onderzoek-naar-linkse-wetenschap-zinnig-6605485-a1545138

Brugh, M. (2017b). *Te linkse vragen in de wetenschap* [Questions that are too leftist in the sciences]. Retrieved from https://www.nrc.nl/nieuws/2017/02/20/te-linkse-vragen-in-de-wetenschap-6825072-a1546873

Buss, D. M., & von Hippel, W. (2018). Psychological barriers to evolutionary psychology: Ideological bias and coalitional adaptations. *Archives of Scientific Psychology*, *6*(1), 148–158.

Cardiff, C. F., & Klein, D. B. (2005). Faculty partisan affiliations in all disciplines: A voter-registration study. *Critical Review*, *17*, 237–255.

Chang, A., & Li, P. (2015). *Is economics research replicable? Sixty published papers from thirteen journals say "usually not"* (FEDS Working Paper No. 2015-083). Washington, DC: Board of Governors of the Federal Reserve System. doi:10.17016/FEDS.2015.083

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284–290.

Cislak, A., Formanowicz, M., & Saguy, T. (2018). Bias against research on gender bias. *Scientometrics*, *115*, 189–200.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Crano, W. D. (2012). *The rules of influence: Winning when you're in the minority*. New York, NY: St. Martin's Press.

Crawford, J. T. (2017). Are conservatives more sensitive to threat than liberals? It depends on how we define threat and conservatism. *Social Cognition*, *35*, 354–373.

Crawford, J. T., & Jussim, L. (2018). Introduction to the politics of social psychology. In J. T. Crawford & L. Jussim (Eds.), *The politics of social psychology* (pp. 1–5). New York, NY: Routledge.

Crawford, J. T., Vodapalli, S., Stingel, R. E., & Ruscio, J. (2019). Do status-legitimizing beliefs moderate effects of racial progress on perceptions of anti-White bias? A replication of Wilkins and Kaiser (2014). *Social Psychological and Personality Science*, *10*, 768–774. 1948550618792057.

Ditto, P. H., Liu, B. S., Clark, C. J., Wojcik, S. P., Chen, E. E., Grady, R. H., . . . Zinger, J. F. (2018). At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives. *Perspectives on Psychological Science*, *14*, 271–291

Duarte, J. L., Crawford, J. T., Stern, C., Haidt, J., Jussim, L., & Tetlock, P. E. (2015). Political diversity will improve social psychological science. *Behavioral and Brain Sciences*, *38*, Article e130. doi:10.1017/S0140525X14000430

Eagan, K., Stolzenberg, E. B., Lozano, J. B., Aragon, M. C., Suchard, M. R., & Hurtado, S. (2014). *Undergraduate teaching faculty: The 2013–2014 HERI faculty survey*. Los Angeles, CA: Los Angeles Higher Education Research Institute, UCLA.

Eagly, A. H. (1995). The science and politics of comparing women and men. *American Psychologist*, *50*, 145–158.

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., . . . Nosek, B. A.

(2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, *67*, 68–82.

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.

Epstude, K., & Meerholz, W. (Eds.). (2014). [Special issue]. *Social Psychology*, *45*(3). Retrieved from https://econtent.hogrefe.com/toc/zsp/45/3

Feldman, S., & Johnston, C. (2014). Understanding the determinants of political ideology: Implications of structural complexity. *Political Psychology*, *35*, 337–358.

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, *7*, 555–561.

Fiske, S. T., Harris, L. T., & Cuddy, A. J. (2004). Why ordinary people torture enemy prisoners. *Science*, *306*, 1482–1483.

Funk, C., & Kennedy, B. (2016). *The politics of climate*. Pew Research Center. Retrieved from http://www.pewinternet.org/2016/10/04/the-politics-of-climate/

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on "Estimating the reproducibility of psychological science." *Science*, *351*, 1037–1037.

Gross, N. (2013). *Why are professors liberal and why do conservatives care?* Cambridge, MA: Harvard University Press.

Gross, N., & Simmons, S. (2014). *Professors and their politics*. Baltimore, MD: JHU Press.

Haidt, J. (2011). *The bright future of post-partisan social psychology*. Retrieved from https://www.edge.org/conversation/jonathan_haidt-the-bright-future-of-post-partisan-social-psychology

Haidt, J. (2016). *New study indicates existence of eight conservative social psychologists*. Retrieved from https://heterodoxacademy.org/new-study-indicates-existence-of-eight-conservative-social-psychologists/

Hamilton, R. F., & Hargens, L. L. (1993). The politics of the professors: Self-identifications, 1969–1984. *Social Forces*, *71*, 603–627.

Handley, I. M., Brown, E. R., Moss-Racusin, C. A., & Smith, J. L. (2015). Quality of evidence revealing subtle gender biases in science is in the eye of the beholder. *Proceedings of the National Academy of Sciences, USA*, *112*, 13201–13206.

Hannikainen, I. R. (2019). Ideology between the lines: Lay inferences about scientists' values and motives. *Social Psychological and Personality Science*, *10*, 832–841. doi:10.1177/1948550618790230

Hirschhorn, J. N., Lohmueller, K., Byrne, E., & Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genetics in Medicine*, *4*(2), 45–61.

Hogan, K., & Maglienti, M. (2001). Comparing the epistemological underpinnings of students' and scientists' reasoning about conclusions. *Journal of Research in Science Teaching*, *38*, 663–687.

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

Inglehart, R., & Klingemann, H.-D. (1976). Party identification, ideological preference and the left-right dimension among Western mass publics. In I. Budge, I. Crewe, & D. J. Farlie (Eds.), *Party identification and beyond: Representations of voting and party competition* (pp. 243–273). London, England: John Wiley & Sons.

Janis, I. L. (1972). *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes*. Oxford, England: Houghton Mifflin Company.

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, *129*, 339–375.

Jussim, L., Crawford, J. T., Anglin, S. M., & Stevens, S. T. (2015). Ideological bias in social psychological research. In J. P. Forgas, K. Fiedler, & W. D. Crano (Eds.), *Social psychology and politics* (pp. 91–109). New York, NY: Taylor & Francis.

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.

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, *8*, 407–424.

Kahan, D. M., Braman, D., Cohen, G. L., Gastil, J., & Slovic, P. (2010). Who fears the HPV vaccine, who doesn't, and why? An experimental study of the mechanisms of cultural cognition. *Law and Human Behavior*, *34*, 501–516.

Kahneman, D. (2003). Experiences of collaborative research. *American Psychologist*, *58*, 723–730.

King, E. B., Avery, D. R., Hebl, M. R., & Cortina, J. M. (2018). *Systematic subjectivity: How subtle biases infect the scholarship review process*. Los Angeles, CA: Sage.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., . . . Nosek, B.A. (2014). Investigating variation in replicability: A "Many Labs" replication project. *Social Psychology*, *45*, 142–152.

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., . . . Nosek, B. A. (2018). Many Labs 2: Investigating variation in replicability across sample and setting. *Advances in Methods and Practices in Psychological Science*, *1*, 443–490.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*, 155–163.

Ladd, E. C., Jr., & Lipset, S. M. (1975). *The divided academy: Professors and politics*. New York, NY: McGraw-Hill Book Company.

Langbert, M., Quain, A. J., & Klein, D. B. (2016). Faculty voter registration in economics, history, journalism, law, and psychology. *Econ Journal Watch*, *13*, 422–451.

Lazarsfeld, P. F., & Thielens, W. (1977). *The academic mind: Social scientists in a time of crisis*. Lowell, MA: Ayer Publishing.

LeBel, E., & Battista, C. (2014). Curate science. Available from http://www.curatescience.org/replications

Levendusky, M. (2009). *The partisan sort: How liberals became Democrats and conservatives became republicans*. Chicago, IL: University of Chicago Press.

Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton, NJ: Princeton University Press.

Lounsbury, J. W., Foster, N., Patel, H., Carmody, P., Gibson, L. W., & Stairs, D. R. (2012). An investigation of the personality traits of scientists versus nonscientists and their relationship with career satisfaction. *R&D Management*, *42*, 47–59.

MacCoun, R. J. (1998). Biases in the interpretation and use of research results. *Annual Review of Psychology*, *49*, 259–287.

Malka, A., Soto, C. J., Inzlicht, M., & Lelkes, Y. (2014). Do needs for security and certainty predict cultural and economic conservatism? A cross-national analysis. *Journal of Personality and Social Psychology*, *106*, 1031–1051.

McGuire, W. J. (1997). Creative hypothesis generating in psychology: Some useful heuristics. *Annual Review of Psychology*, *48*, 1–30.

McNutt, M. (2014). *Reproducibility*. Washington, DC: American Association for the Advancement of Science.

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

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago, IL: University of Chicago Press.

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behavior*, *3*, 221–229.

Nature. (2013 April 24). Announcement: Reducing our irreproducibility [Editorial]. *Nature*, *496* (398). doi:10.1038/496398a

Nature. (2017 March 15). Announcement: Transparency upgrade for *Nature* journals. *Nature*, *543*(288). doi:10.1038/543288b

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220.

Nosek, B. A., & Lindsay, D. S. (2018). Preregistration becoming the norm in psychological science. *APS Observer*, *31*(3), 19–21. Retrieved from https://www.psychologicalscience.org/observer/preregistration-becoming-the-norm-in-psychological-science

Nyhan, B., Reifler, J., & Ubel, P. A. (2013). The hazards of correcting myths about health care reform. *Medical Care*, *51*, 127–132.

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

Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: How much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, *10*, 712–712.

R Core Team. (2018). R: A language and environment for statistical computing (Version 3.5.1) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.

Reaves, M. L., Sinha, S., Rabinowitz, J. D., Kruglyak, L., & Redfield, R. J. (2012). Absence of detectable arsenate in DNA from arsenate-grown GFAJ-1 cells. *Science*, *337*(6093), 470–473.

Redding, R. E. (2001). Sociopolitical diversity in psychology: The case for pluralism. *American Psychologist*, *56*, 205–215.

Revelle, W. (2018). psych: Procedures for psychological, psychometric, and personality research [Computer software]. Retrieved from https://cran.r-project.org/package=psych

Richardson, E. T., & Polyakova, A. (2012). The illusion of scientific objectivity and the death of the investigator. *European Journal of Clinical Investigation*, *42*, 213–215.

Rykiel, E. J. (2001). Scientific objectivity, value systems, and policymaking. *AIBS Bulletin*, *51*, 433–436.

Schweinsberg, M., Madan, N., Vianello, M., Sommer, S. A., Jordan, J., Tierney, W., . . . Uhlmann, E. L. (2016). The pipeline project: Pre-publication independent replications of a single laboratory's research pipeline. *Journal of Experimental Social Psychology*, *66*, 55–67.

Sears, D. O. (1994). Ideological bias in political psychology: The view from scientific hell. *Political Psychology*, 547–556.

Shi, F., Teplitskiy, M., Duede, E., & Evans, J. (2019). The wisdom of polarized crowds. *Nature Human Behavior*, *3*, 329–336.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, *69*, 487–510.

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.

Simons, D. J., Holcombe, A. O., & Spellman, B. A. (2014). An introduction to registered replication reports at perspectives on psychological science. *Perspectives on Psychological Science*, *9*, 552–555.

Skitka, L. J. (2012). Multifaceted problems: Liberal bias and the need for scientific rigor in self-critical research. *Perspectives on Psychological Science*, *7*, 508–511.

Tankard, M. E., & Paluck, E. L. (2017). The effect of a Supreme Court decision regarding gay marriage on social norms and personal attitudes. *Psychological Science*, *28*, 1334–1344.

Tetlock, P. E. (1994). Political psychology or politicized psychology: Is the road to scientific hell paved with good moral intentions? *Political Psychology*, *15*, 509–529.

Tetlock, P. E., & Mellers, B. A. (2011). Intelligent management of intelligence agencies: Beyond accountability ping-pong. *American Psychologist*, *66*, 542–554.

Tetlock, P. E., & Mitchell, G. (1993). Liberal and conservative approaches to justice: Conflicting psychological portraits. In B. A. Mellers & J. Baron (Eds.), *Psychological perspectives on justice: Theory and applications* (pp. 234–255). Cambridge, England: Cambridge University Press.

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

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences, USA*, *113*, 6454–6459.

Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in Cognitive Sciences*, *22*, 213–224.

Van Bavel, J. J., Reinero, D. A., Harris, E., Robertson, C., & Pärnamets, P. (2020). Breaking groupthink: Why scientific identity and norms mitigate ideological epistemology. *Psychological Inquiry*, *31*, 66–72.

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 (Eds.), *The politics of social psychology* (pp. 7–25). New York, NY: Psychology Press.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*, 779–804.

Washburn, A. N., & Skitka, L. (2018). *Strategies for promoting strong inferences in political psychology research*. doi:10.31234/osf.io/c4wv5

Wickham, H. (2017). tidyverse: Easily install and load the 'Tidyverse' [Computer software]. Retrieved from https://cran.r-project.org/package=tidyverse