Journal of Experimental Social Psychology 79 (2018) 188–199

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Is research in social psychology politically biased? Systematic empirical tests and a forecasting survey to address the controversy

Orly Eitan[a,1], Domenico Viganola[b,1], Yoel Inbar[c,1], Anna Dreber[b,d], Magnus Johannesson[b], Thomas Pfeiffer[e], Stefan Thau[a], Eric Luis Uhlmann[a,*,1]

[a] INSEAD, Singapore
[b] Stockholm School of Economics, Sweden
[c] University of Toronto, Canada
[d] University of Innsbruck, Austria
[e] Massey University, New Zealand

ARTICLE INFO

Handling editor: Aarti Iyer

Keywords:
Politics
Liberalism
Conservatism
Meta-science
Political bias
Liberal bias

ABSTRACT

The present investigation provides the first systematic empirical tests for the role of politics in academic research. In a large sample of scientific abstracts from the field of social psychology, we find both *evaluative differences*, such that conservatives are described more negatively than liberals, and *explanatory differences*, such that conservatism is more likely to be the focus of explanation than liberalism. In light of the ongoing debate about politicized science, a forecasting survey permitted scientists to state a priori empirical predictions about the results, and then change their beliefs in light of the evidence. Participating scientists accurately predicted the direction of both the evaluative and explanatory differences, but at the same time significantly overestimated both effect sizes. Scientists also updated their broader beliefs about political bias in response to the empirical results, providing a model for addressing divisive scientific controversies across fields.

He who knows only his side of the case, knows little of that.
— John Stuart Mill, *On Liberty*

Are scientific investigations of politically charged topics affected by the values of the scientists themselves? This question has been the subject of considerable debate in the social sciences, including psychology, for quite some time (Duarte et al., 2015; Eagly, 2014; Haidt, 2011; Redding, 2001; Tetlock, 1994). However, the empirical evidence that can be brought to bear on the issue is mainly indirect.

Scientists, including psychological scientists, overwhelmingly fall on the socially liberal end of the liberal-conservative dimension (Cardiff & Klein, 2005; Gross & Simmons, 2006; Inbar & Lammers, 2012; Klein & Stern, 2005; McClintock, Spaulding, & Turner, 1965; Rothman & Lichter, 2008). This does not necessarily mean that scientists engage in motivated reasoning (Ditto & Lopez, 1992; Kunda, 1990; Lord, Ross, & Lepper, 1979; Sherman & Cohen, 2002) when choosing topics and methodologies, analyzing data, or interpreting research results. However, the political demographics of academia do present a risk of intellectual homogeneity and consequent ideological bias. Scientific safeguards designed to guard against error and bias may not attenuate the natural motivated reasoning shown by all human beings enough to prevent biased research.

Critics have highlighted specific research programs putatively compromised by liberal politics (Al-Gharbi, in press; Arkes & Tetlock, 2004; Sniderman & Tetlock, 1986; Tetlock, 1994; Tetlock & Mitchell, 2009). However, these charges are typically denied by the original authors (e.g., Banaji, Nosek, & Greenwald, 2004; Sears, 1994; Tarman & Sears, 2005) and even if true do not necessarily show systematic liberal bias throughout any particular academic field. Outside of specific cases of potentially left-leaning academic research programs, the modal scientific investigation into politically charged topics may not be slanted toward any particular worldview.

In a survey of social psychologists, Inbar and Lammers (2012) found that many respondents reported a willingness to discriminate against conservative colleagues in grant and article reviews, symposium invitations, and hiring decisions. In addition, conservative social psychologists reported a work climate more hostile to their political beliefs than their liberal counterparts did. Further evidence suggests that academic reviewers evaluate findings that conflict with their own political orientation more negatively (Abramowitz, Gomes, & Abramowitz, 1975). These mechanisms could indirectly distort the

* Corresponding author.
E-mail address: eric.uhlmann@insead.edu (E.L. Uhlmann).
[1] First three and last author contributed equally.

https://doi.org/10.1016/j.jesp.2018.06.004
Received 23 January 2018; Received in revised form 8 June 2018; Accepted 14 June 2018
Available online 10 August 2018

scientific literature by reducing the population of conservative scientists or making it more difficult to carry out, publish, and disseminate research with results that challenge liberal political beliefs. Yet it remains unclear whether the output of the scientific process — the research itself — is affected by liberal values, and if so to what extent.

The present investigation leveraged a large sample of scientific abstracts from the field of social psychology to carry out empirical tests for two distinct effects of researcher politics. In Studies 1a and 1b, relying on thousands of independent raters with a wide range of political beliefs, we assess whether conservatives and conservative ideas are systematically characterized more negatively than liberals and liberal ideas. Of course, such *evaluative differences* do not necessarily reflect bias on the part of the scientists, since conservatism could be associated with objectively more negative characteristics than liberalism. Perhaps it is the case, as comedian Rob Corddry once joked, that “The facts have a well-known liberal bias” (Krugman, 2014). However, capturing differences in evaluation represents a necessary first step to establishing a political slant to psychological sciences.

In Studies 1a and 1b, we also test for a subtler effect in terms of what ideological positions are implicitly regarded as normative and non-normative. Prior research shows that groups implicitly seen as deviant from the norm are more likely to be the focus of explanation (Miller, Taylor, & Buck, 1991). For instance, gender differences tend to be explained in terms of women, not men (Miller et al., 1991), and differences in behavior between heterosexuals and homosexuals in terms of the deviance of homosexuals (Hegarty & Pratto, 2001). We therefore tested for *explanatory differences* — whether conservatives and conservative ideas are the targets of explanation more so than liberals and liberal ideas (Brandt & Spälti, 2018). Although by no means an exhaustive test of all the ways in which political values may play a role in the scientific process (Brandt & Proulx, 2015; Brandt & Spälti, 2018; Jussim, Crawford, Anglin, Stevens, & Duarte, 2016), these represent meaningful initial tests. In the General Discussion, we outline additional lines of inquiry regarding political bias in science that might be pursued in future investigations.

It is possible that politicized research is largely in the eye of the beholder, with conservative readers of scientific work perceiving a liberal bias not seen by liberal readers. This would be analogous to the well-known hostile media bias, in which opposing camps on a controversial issue both perceive neutral media reports as slanted in favor of the other side (Vallone, Ross, & Lepper, 1985). We therefore carefully took into account the political attitudes of those evaluating the scientific work for bias.

Both political and scientific debates can prove intractable, in part because contrary evidence can be discounted using post hoc motivated reasoning (Kuhn, 1962; Lakatos, 1970; Tetlock, 2005). One innovative means to render strongly held beliefs vulnerable to disproof is to use a forecasting survey to elicit prior beliefs (Dreber et al., 2015; Tetlock, Mellers, Rohrbaugh, & Chen, 2014; Wolfers & Zitzewitz, 2004). Parties to a debate are asked to make a priori predictions about future events and once the objective outcomes are revealed, those involved have the opportunity to update their beliefs (or not) in light of the new evidence. Although forecasting surveys have been used to predict future geopolitical events (Tetlock, 2005; Tetlock & Mellers, 2014), to our knowledge they have not previously been leveraged to address a scientific controversy. In Study 2, scholars with a range of positions on the role of politics in science were asked to make a priori predictions regarding the extent to which evaluative differences and explanatory differences would be found in the research reports. These same scientists were subsequently presented with the obtained effect sizes and provided the opportunity to update their positions on both the specific empirical questions at hand and broader controversy regarding the role played by scientists' political values in their research.

In the spirit of open science, and to reduce any bias on our part as much as possible (e.g., the “bias to find bias”; Krueger & Funder, 2004), the analyses for the project were pre-registered (Van't Veer & Giner-Sorolla, 2016; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012) and the data publicly posted online (Simonsohn, 2013; Wicherts & Bakker, 2012) to facilitate re-analyses and open debate and discussion, see https://osf.io/zhf98/, https://osf.io/vtyg4/, and https://osf.io/jh47m/. The complete study materials are further provided in Supplements 1, 2, and 3, and deviations from the pre-analysis plan described in Supplement 4.

1. Studies 1a and 1b

The primary goal of these investigations was to empirically estimate evaluative and explanatory differences with regards to conservatism and liberalism in abstracts from scientific research reports. Politically relevant conference abstracts were selected first using keywords (e.g., *liberal*, *conservative*), and then rated for political relevance by a large crowd of independent coders (Study 1a). Abstracts that touched on clearly political topics were then systematically assessed for evaluative and explanatory differences by thousands of independent raters who themselves varied greatly in their political values (Study 1b). This approach allowed us to parse the extent to which the political overtones of scientific research appear to be attributable to the report itself (i.e., the abstract is consistently rated as casting conservatives in a negative light, regardless of who is doing the rating) as opposed to in the eye of the beholder (e.g., conservatives see the research report as biased against their group, whereas liberals perceive it as evenhanded).

1.1. Study 1a: methods

1.1.1. Initial selection of abstracts

To carry out the project, we took advantage of the fact that the programs for the Society for Personality and Social Psychology (SPSP) annual conference (one of the main academic conferences in social psychology) are available online dating from the 2003 meeting. We collected programs for 10 years (2003–2013) and searched the listed abstracts (which could describe either poster presentations or talks) for the search terms *liberal*, *conservative*, *democrat*, *republican*, *politics*, *political*, *conservatism*, and *liberalism* to cull a subset of the abstracts that might reflect research investigations examining the psychology of political beliefs and behavior. This process led to the initial selection of 846 abstracts. We deliberately chose a broad set of search terms to avoid missing any potentially relevant abstracts. This meant that some abstracts matching these terms might not be politically relevant, since terms like “conservative” are also used in non-political contexts (e.g., “a conservative test of the hypothesis”).

1.1.2. Participants

We then recruited 934 U.S. based workers from Amazon's Mechanical Turk to rate each abstract for its political relevance on a simple dichotomous scale (Brown & Allison, 2014) (Supplement 1). We chose to use Mechanical Turk workers because they are more demographically diverse than typical undergraduate samples (Buhrmester, Kwang, & Gosling, 2011; Paolacci & Chandler, 2014; Paolacci, Chandler, & Ipeirotis, 2010). Overall, raters were 51% female. They were slightly left of center, as indicated by means below the scale midpoint of 4 (*1 = very liberal*, *4 = moderate*, *7 = very conservative*) for overall political orientation ($M = 3.52$, $SD = 1.78$), social issues ($M = 3.07$, $SD = 1.78$), economic issues ($M = 3.77$, $SD = 1.83$), as well as a mean below the scale midpoint of 3.5 (*1 = strongly support Democrats*, *6 = strongly support Republicans*) for political party preference ($M = 2.92$, $SD = 1.59$).

1.1.3. Materials and design

Raters were asked: “Is the research about how political liberals and conservatives think, about differences between political liberals and conservatives, about differences in opinion on a political issue about which liberals and conservatives typically have different opinions, or about voting or other political behavior?" (*Yes/No*). Each abstract was rated by an average of 25.1 raters, with a range of 18 to 56 raters per abstract. Raters reported demographic information including their gender, political orientation overall and separately for social and economic issues, and political party preference.

1.2. Results & discussion

Only abstracts rated as politically relevant by at least 60% of respondents ($N = 306$) were retained for further analysis. We pre-registered this cutoff for political relevance ratings because we believed that it maximized our sample size of abstracts, and therefore the statistical power of our analyses, while not including any non-relevant abstracts. Examining the abstracts that barely met the 60% cut-off showed that they were clearly relevant to politics. The four abstracts that were coded as politically-relevant by 60.0–60.9% of raters were titled: "Dehumanizing the poor: Attitudes and implications" (the abstract further indicates the reported "effect is strongest in those participants from high-income backgrounds who are self-reported conservatives"), "Ideology, method, and the seemingly intractable conflict between personality and social psychological worldviews" (contrasts ideologies "linked to conservatism, right-wing authoritarianism…" and those that are "intrinsically liberal"), "Intolerance of intolerance: The limits of liberal worldviews," and "Laypeople's perceived political orientation of psychologists moderate their judgments of psychologists' responsibility attribution to wrongdoers." As robustness checks, we repeated our primary analyses below (see Study 1b) with alternative cutoffs of 70%, 80%, and 90% for relevance ratings.

For the most part rater ideology did not affect how likely raters were to classify abstracts as politically relevant ($ps > .15$ for overall orientation, economic issues, and party preference in separate models). The sole exception was ideology on social issues, where greater conservatism was associated with a higher likelihood of rating abstracts as politically relevant, OR = 1.03, 95% CI [1.01, 1.05], $Z = 2.72$, $p = .007$. It is relevant to note that this result was in a model with social ideology as the only predictor.

1.3. Study 1b: methods

1.3.1. Participants

In Study 1b, each of the final set of 306 politically relevant abstracts from Study 1a was separately assessed for evaluative differences and explanatory differences by two independent groups of U.S. based Mechanical Turk raters (total raters = 2560; see Supplement 2). On average, raters placed themselves somewhat to the liberal side of the scale midpoint of 4 (*1 = very liberal*, *4 = moderate*, *7 = very conservative*): *overall* $M = 3.59$, $SD = 1.76$; *social* $M = 3.10$, $SD = 1.82$; *economic* $M = 3.88$, $SD = 1.79$. Forty-nine percent were male and 51% female. Their average yearly income was \$55,000 (SD = \$49,000); 96% were born in the United States and 4% in other countries.

1.3.2. Materials and design

Two thousand four hundred raters assessed 20 abstracts each and 160 raters assessed 6 abstracts each (these were the remaining 6 abstracts after dividing 300 by 20). Thus, 48,960 ratings [$(2400 \times 20) + (160 \times 6)$] were recorded in all. Titles and abstracts for papers were presented without information about authors and publication year. To assess explanatory differences, raters were asked "To what extent does this research attempt to explain political liberalism or liberal ideas?" and "To what extent does this research attempt to explain political conservatism or conservative ideas?" (*1 = not at all*, *7 = a great deal*). To assess evaluative differences, items asked "How does this research characterize political liberals or liberal ideas?" and "How does this research characterize political conservatives or conservative ideas?" (*1 = extremely negatively*, *4 = neutral*, *7 = extremely positively*). The order in which the conservatism and liberalism items appeared was counterbalanced between subjects.

This study therefore used a 2 (explanatory differences ratings vs. evaluative differences ratings) × 2 (order of items: political conservatism ratings first or political liberalism ratings first) between-subjects design. All abstracts were rated for both evaluative and explanatory differences; thus raters rather than abstracts were randomly assigned to these conditions. We chose this design for three reasons: 1) there were too many abstracts for a single rater to rate all abstracts without fatigue or boredom (and therefore each rater only saw a subset of abstracts); 2) we did not want to assume that, for example, research that characterized liberals positively necessarily described conservatives negatively (and therefore participants were asked to complete conservatism and liberalism ratings on separate scales); 3) we did not want to suggest to participants that evaluative and explanatory differences might be related (and therefore any given participant rated either evaluative or explanatory differences, but not both). Demographics of the raters were further assessed, including gender, age, education, and place of birth. We also assessed raters' political orientation overall and separately for social and economic issues.

1.4. Results & discussion

1.4.1. Sensitivity analysis

A sensitivity analysis indicated the study had close to 100% power to detect the minimal effect size of interest with regard to evaluative and explanatory differences, specifically $d = 0.10$. Because the analyses below use cross-classified mixed-effects models, for which power formulas are not available, we estimated power via Monte Carlo simulation using the "simr" package (version 1.0.3; Green & MacLeod, 2016) in R 3.3.2 (R Core Team, 2016). This technique starts with a fitted model. It then repeatedly draws new values for the response variable from a distribution, refits the model, and tests the statistical significance of the parameter(s) of interest. Note that unlike a power analysis formula, which can answer a question like "In this design, what size of an effect can I detect with a power of 0.8?", a simulation answers the question "How often (i.e., with what power) can I detect an effect of size x?" For our power analysis, we chose a minimum effect size of interest: $r = 0.05$, which is equivalent to $d = 0.10$. For the pre-registered analyses for overall evaluative and explanatory differences, and the simple effects of self-rated ideology on perceived evaluative and explanatory differences, we had essentially 100% power to detect an effect of this size, 95% CI [99.26, 100] in 500 simulations.

1.5. Evaluative and explanatory differences

We predicted that in scientific abstracts in social psychology, conservatives and conservatism would be systematically characterized more negatively than liberals and liberalism. Recall that each abstract was rated by multiple raters, who each rated a different set of abstracts. Furthermore, each rater assessed each abstract separately for how positively or negatively it characterized liberals and conservatives. We therefore used the "lme4" package (version 1.1-12; Bates, Maechler, Bolker, & Walker, 2015) in R 3.3.2 (R Core Team, 2016) to fit a mixed-effects model with random intercepts for raters and abstracts and a fixed effect for whether the rating target was liberals or conservatives (dummy-coded as 1 = liberals, 0 = conservatives). A significantly positive coefficient for the dummy variable would indicate that, on average, abstracts were seen as characterizing liberals more positively than conservatives. Indeed, the coefficient was positive and significant, $t(45{,}720) = 37.81$, $p < .001$ (degrees of freedom are Satterthwaite-approximated). The coefficient point estimate (0.38) is directly interpretable as the average difference (in scale points) in the rated positivity of abstracts when describing liberals vs. conservatives. Using the formula developed by Lai and Kwok (2014; Eq. 9), this was a medium-sized effect, with $d = 0.33$ (Table 1).

We conducted a parallel analysis for explanatory differences — that

Table 1
Abstract ratings for evaluative and explanatory differences. Difference scores are at the abstract level (i.e., after averaging liberal/conservative ratings by abstract).

	Evaluative rating			Explanatory rating		
	Liberal	Conservative	Difference	Liberal	Conservative	Difference
Mean	4.17	3.78	0.38	3.74	4.12	−0.39
SD	0.39	0.46	0.70	0.90	0.80	1.19
Min, max	3.04, 5.63	2.44, 5.54	−2.50, 2.79	1.77, 6.17	2.15, 5.83	−4.01, 3.35

Note. $N = 306$.

is, whether research attempts to explain conservatives more than liberals. In this case, a significantly negative coefficient for the dummy variable would indicate that conservatives are targets of explanation more than liberals. Indeed, the coefficient was negative and significant, $t(44{,}180) = -25.90$, $p < .001$, $d = -0.21$ (using the equation from Lai & Kwok, 2014). Again, the coefficient point estimate (−0.39) is the average difference (in scale points) in the extent to which abstracts were seen as attempting to explain liberals vs. conservatives.

Table 1 shows abstract-level difference scores for evaluative and explanatory ratings. Eighty-six (28%) of the abstracts were rated as describing conservatives more positively than liberals; 220 (72%) were rated as describing liberals more positively than conservatives. One hundred seventy (56%) of the abstracts were rated as focused more on explaining conservatives; 130 (42%) were rated as focused more on explaining liberals (six abstracts were rated as explaining both equally). Fig. 1 plots how all abstracts were rated.

1.5.1. Robustness to inclusion criteria

The analyses above relied on our pre-registered inclusion criterion for abstracts, specifically that 60% or more of pre-test raters indicated that the abstract was politically relevant. As an exploratory supplementary analysis, we also tested for evaluative and explanatory differences using three more restrictive criteria: 70% agreement ($N = 231$ abstracts), 80% agreement ($N = 149$ abstracts), and 90% agreement ($N = 56$ abstracts). Under all criteria but the last, results were very similar to those reported above. Using the 70% criterion, for evaluative differences $d = 0.33$; for explanatory differences $d = -0.22$. Using the 80% criterion, for evaluative differences $d = 0.36$; for explanatory differences $d = -0.17$. Using the 90% criterion, for evaluative differences $d = 0.36$; for explanatory differences $d = -0.01$. The divergent result for explanatory differences in this case is likely due to the greatly reduced sample size ($N = 56$) when the 90% cutoff is used. We further correlated the percentage of respondents rating an abstract as politically relevant with average rated evaluative and explanatory differences for that abstract. These analyses revealed negligible correlations, $r_{eval}(304) = 0.04$, 95% CI [−0.07, 0.15], $p = .47$, $r_{expl}(304) = 0.09$, 95% CI [−0.02, 0.20], $p = .12$. The results reported throughout the remainder of the paper therefore use the originally planned 60% cutoff.

1.5.2. Relationship between evaluative and explanatory differences

To assess whether perceived evaluative and explanatory differences are correlated at the abstract level, we calculated average ratings for explaining and evaluating liberals and conservatives. This gave us four aggregate scores per abstract. Computing correlations between these aggregates separately for ratings of liberals and conservatives revealed that abstracts seen as explaining conservatives more were also seen as evaluating them significantly more negatively, $r(304) = -0.22$, 95% CI [−0.33, −0.11], $p < .001$. In contrast, abstracts seen as explaining liberals more were seen as evaluating them directionally, but non-significantly, more positively, $r(304) = 0.08$, 95% CI [−0.03, 0.19], $p = .17$.

1.5.3. Trends over time

Finally, we examined whether evaluative and explanatory differences in abstracts decrease over time, increase, or stay the same, over the 10 year timespan of our sample. In order to answer this question, we fit two linear mixed models (one for evaluative differences and one for exploratory differences) in which we added the main effect of abstract publication year (mean-centered) and its interaction with target ratings. This interaction is the key test of the hypothesis. Note that these models also included random intercepts for rater and abstract. For evaluative differences, the interaction between target ratings and publication year was significant, $t(45{,}720) = -2.96$, $p = .003$. Decomposing this interaction showed that as year increased, abstracts were rated as evaluating both conservatives and liberals more positively, but neither trend over

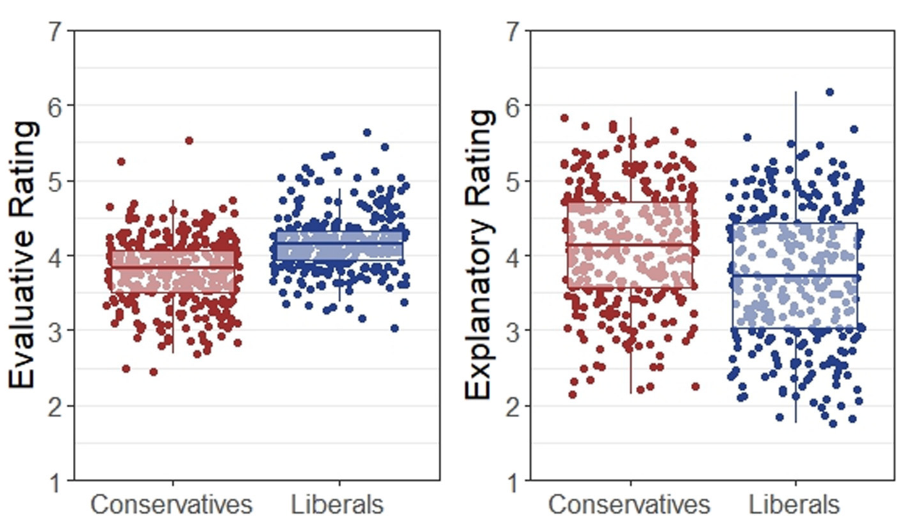

Fig. 1. Ratings of how positively abstracts characterize liberals and conservatives (left panel) and of how much abstracts explain liberals and conservatives (right panel).

time was significant, $t(304) = 1.78$, $p = .076$ and $t(304) = 0.66$, $p = .510$, respectively. For explanatory differences, the interaction between target ratings and publication year was significant, $t(44{,}180) = 3.48$, $p < .001$. Decomposing this interaction showed that as year increased, abstracts were rated as explaining conservatives less, $t(303) = -2.31$, $p = .021$, with no change for explaining liberals, $t(304) = -1.11$, $p = .267$.

1.5.4. Role of rater ideology

We next tested for interactions between raters' self-reported ideology on social issues and perceptions of bias in the scientific abstracts (we pre-registered self-rated ideology on social issues as our primary operationalization of ideology because we reasoned that these sorts of "culture war" issues are where bias in research seems most likely). For both perceived evaluative and explanatory differences, we found significant interactions between rater ideology and perceived differences, evaluative differences $t(45{,}720) = -17.98$, $p < .001$; explanatory differences $t(44{,}180) = 3.47$, $p < .001$. Contrary to predictions, however, these interactions were not driven by conservatives perceiving greater bias. Rather, greater self-rated social conservatism was associated with seeing the abstracts as describing conservatives more positively, $t(1218) = 9.972$, $p < .001$, and liberals more negatively $t(1196) = -2.31$, $p = .021$. For explanatory differences, greater self-rated social conservatism was associated with rating abstracts as less focused on explaining conservatives, $t(1191) = -2.71$, $p = .007$, but was not related to the extent to which abstracts were seen to explain liberals, $t(1186) = -0.71$, $p = .476$. This unpredicted pattern may be the result of ideology-linked characteristics being seen as more positive by those sharing that ideology (for example, conservatives may see patriotism as positive and sexual permissiveness as negative, whereas the opposite may be true for liberals). Exploratory follow-up analyses showed, however, that even when restricting the analysis to socially conservative raters (i.e., self-reported ideology > 4), abstracts were seen as evaluating liberals more positively, $t(10{,}240) = 6.81$, $p < .001$, and as explaining them less than conservatives, $t(9850) = -10.50$, $p < .001$.

We carried out three robustness checks by fitting models using alternative measures of ideology: a) overall self-rated liberalism/conservatism; b) a composite of social liberalism/conservatism, economic liberalism/conservatism, and overall liberalism/conservatism; c) political party identification (each model included only that ideology measure). Using each of these alternative measures revealed very similar moderation by rater ideology. For evaluative differences, we found a significant interaction between ideology and target dummy for overall ideology, $t(43{,}560) = -16.67$, $p < .001$, the ideology composite, $t(43{,}560) = -17.23$, $p < .001$, and party identification, $t(36{,}680) = -14.50$, $p < .001$. For explanatory differences, we found a significant interaction between ideology and target dummy for overall ideology, $t(42{,}220) = 4.50$, $p < .001$, the ideology composite, $t(42{,}220) = 4.24$, $p < .001$, and party identification, $t(35{,}670) = 3.47$, $p < .001$.

Further analyses directly compared the variance in ratings explained by differences between abstracts to the variance explained by rater politics. We compare three nested models: Model A had random intercepts for raters and abstracts, but no random slopes. Model B added random slopes for the target dummy with respect to abstract (i.e., it allowed the size of the effect of target to vary randomly across abstracts). Model C added fixed effects of rater political orientation and its interaction with the target dummy. For evaluative differences, we found an R^2 of 0.126 for Model A, which increased to 0.212 when adding random slopes by abstract (Model B) and to 0.217 when adding rater political orientation (Model C). Likewise, for explanatory differences the Model A R^2 was 0.288, which increased to 0.386 for Model B. The model C R^2 increased by < 0.001. Thus, overall a substantial amount of variance in rater perceptions of evaluative and explanatory differences is explained by variability between abstracts, such that modeling random slopes for abstracts increased R^2 by about 10 percentage points. In contrast, modeling rater political ideology increased model R^2 by less than one percentage point. Comparing relatively, differences between abstracts explained 17 to 245 times the variance that rater politics did.

In sum, robust evaluative and explanatory differences emerged in the scientific abstracts that were more attributable to the scientific reports than to the politics of the perceiver. In other words, abstracts were consistently rated as evaluating conservatives more negatively than liberals, and also as explaining conservatives more so than liberals, regardless of whether a political liberal or conservative did the rating. Contrary to predictions, conservatives were slightly less (rather than more) likely to see scientific research reports as biased against their group and beliefs. Overall, rater political orientation explained a tiny fraction of the variance in ratings, whereas the lion's share of the variability was explained by the abstracts themselves. Explanatory and evaluative differences were positively but only modestly correlated, suggesting they represent related but distinct facets of politics in science. Finally, some initial evidence emerged that conference abstracts in social psychology have become progressively more politically neutral over time, perhaps due to recent publicity and debates on the topic of politicized science. However, these trends over time were not statistically robust, and future investigations should rely on a larger sample of abstracts extending over a longer time period than the ten-year span of our sample.

2. Study 2

Studies 1a and 1b carefully culled politically relevant abstracts from the field of social psychology and systematically assessed them for the tendencies to treat conservatives as the target of explanation and cast conservatives in a negative light, relative to liberals. Adding to prior work assessing demographic imbalances in political orientation across scientific fields (e.g., Duarte et al., 2015; McClintock et al., 1965; Rothman & Lichter, 2008), surveying academics on their willingness to discriminate in selection decisions against individuals whose political beliefs differ from theirs (Inbar & Lammers, 2012), and critiquing specific research programs for alleged liberal bias (e.g., Jussim et al., 2016; Tetlock, 1994), these investigations provide the first empirical estimates of the degree of political bias throughout a field of scientific inquiry. We examine political bias in the research reports themselves, and do so systematically across many articles rather than focusing on specific (and potentially non-representative) papers. As the controversy regarding the extent to which political values influence scientific research is ongoing and divisive (e.g., see Duarte et al., 2015, and the associated commentaries), it is of interest what researchers believe about political bias in science, whether or not these beliefs map on to the available empirical estimates, and whether learning about new evidence can change opinions and help address this debate.

Study 2 therefore examined whether independent scientists provided with the design and materials could predict the key outcomes of Study 1b (i.e., the effect sizes for evaluative and explanatory differences in the abstracts), and whether this new evidence had any effect on their more general beliefs about politics in scientific research. Recent forecasting surveys indicate that academics can, when presented with the description of a research study, make surprisingly accurate predictions about the direction, significance, and effect size of the finding (Camerer et al., 2016; DellaVigna & Pope, in press, 2016; Dreber et al., 2015). However, these previous investigations did not examine the role of scientists' political values in their forecasts about politicized topics. Political partisanship and strongly held prior beliefs may, or may not, distort scientists' perceptions of and predictions about politics in science. Finally, although it is possible that academics rationally update their beliefs in light of new empirical evidence, they might alternatively employ motivated reasoning (Ditto & Lopez, 1992; Kunda, 1990) and counterfactual thinking (Tetlock, 2005) to dismiss the findings.

2.1. Methods

2.1.1. Participants

Using social media (Twitter and Facebook), we recruited participants for a survey examining whether respondents could predict the direction and effect sizes for evaluative and explanatory differences (Supplements 3 and 5). Three hundred and nine respondents began and 202 completed the forecasting survey. We received informal feedback over email from two respondents who discontinued the survey that they found the instructions and effect size estimation task unclear, which together with the length of the survey may have contributed to the completion rate of 65%. In addition, 4 respondents who completed all the items and 1 respondent who partially completed them elected to withdraw from the study when offered the opportunity at the end of the survey, resulting in a final sample of 198 participants.

The final sample of forecasters was 40% female, with an average age of 35.90 ($SD = 10.83$). Sixty-eight percent of the participants were native English speakers; among the non-native speakers, participants had on average > 25 years of experience with the English language. In terms of ethnicity, 164 participants self-categorized as White/Caucasian, 8 as Asian, 6 as mixed ethnicities, 3 as Latino/Hispanic, 2 as Black, 2 as Middle Eastern, and 13 did not provide a response. One hundred and nineteen were born in the United States and 70 in other countries (9 did not specify their place of birth), and forecasters currently resided in 21 countries including the U.S., Canada, United Kingdom, Germany, the Netherlands, and Sweden, among others.

The majority of the forecasters (185) came from academia and the social sciences (132 from psychology or sociology departments; 26 from economics, management, organizational behavior, and marketing departments; 7 from political science and public policy departments; 3 from education departments; 2 from departments of medicine; 5 from other departments including anthropology, communication, decision science, linguistics, and film and media studies; 8 did not specify any academic department, 11 were not in academia, and 4 did not specify their job). In terms of job rank, 21 forecasters were full professors, 28 were associate professors, 31 were assistant professors, 28 were postdoctoral researchers, 10 were non tenure-track lecturers, 52 were graduate students, 3 were masters students, 3 were lab managers, and 2 were research assistants.

Consistent with past surveys of the political values of academics (Inbar & Lammers, 2012), 140 self-categorized as liberal on social issues, 45 as moderate, and 11 as conservative. One hundred self-categorized as liberal on economic issues, 69 as moderate, and 27 as conservative. With regard to U.S. political party identification, 137 participants supported the Democratic party, 14 supported the Republican party, 21 supported other parties (10 Libertarian, 5 Green Party, 3 Democratic Socialist, 1 centrist, 1 working families party, 1 stated to be "farther left" than the democratic party), 3 stated that their preferences depend on the candidates, 1 stated themselves to be an anarchist, and the remaining 22 participants reported no preferred political party.

2.1.2. Materials and design

In the first part of the forecasting survey, participants were told they would attempt to predict the results of an empirical study assessing two potential political effects in scientific research abstracts: *evaluative differences* (who is explained more negatively, conservatives or liberals?) and *explanatory differences* (who is the focus of explanation more, regardless of valence, conservatives or liberals?). They were informed of the data source (SPSP conference abstracts 2003–2013) and selection criteria for political relevance, told about the study design and raters, and provided with the full text of the items used for all ratings. They were further provided with a brief explanation of effect size statistics and a link to further information online, together with definitions of explanatory and evaluative differences. Forecasters were then asked "What do you predict will be the effect size for evaluative [explanatory] differences?" and gave numeric free responses for their estimates.

In addition to their estimates of the objective effect sizes, forecasters were asked for their subjective beliefs regarding differences in evaluation and explanation: "Do you think research in social psychology evaluates conservatives and liberals differently (in terms of characterizing them negatively or positively)?" (*1 = yes, conservatives are evaluated much more negatively* to *5 = yes, liberals are evaluated much more negatively*), and "Do you think research in social psychology seeks to explain conservatism and liberalism to different degrees?" (*1 = yes, conservatives are explained much more* to *5 = yes, liberals are explained much more*).

They were also surveyed regarding their more general beliefs about the extent to which politics shapes the conclusions drawn by scientists: "The personal political beliefs of social scientists do not ultimately influence the conclusions of their research," and "Methodological safeguards prevent the personal beliefs of researchers from unduly biasing their research" (*1 = strongly disagree*, *5 = strongly agree*). These two items formed a reliable index of beliefs about political bias on the part of researchers, $r(195) = 0.55$, 95% CI [0.45, 0.64], $p < .001$. Respondents were also asked whether they thought the field in general is politicized: "In your opinion, is social psychology as a field generally speaking politically neutral, generally biased against conservatives, or generally biased against liberals?" (*1 = strongly biased against liberals*, *5 = strongly biased against conservatives*).

Finally, four items asked *why* evaluative and explanatory differences in the abstracts might exist. To capture the sentiment that "reality has a liberal bias," one of each pair of items attributed evaluative/explanatory differences to bias on the part of scientists, and the second to objective attributes of conservatives and liberals. Regarding evaluative differences, respondents were asked "If evaluative differences in research on liberals and conservatives occur, this is because scientists' political beliefs bias their research conclusions," and "If evaluative differences in research on liberals and conservatives occur, this is because conservatives and liberals objectively differ in positive and negative characteristics" (reverse coded) (*1 = strongly disagree*, *5 = strongly agree*). These two items were only modestly correlated, $r(194) = 0.29$, 95% CI [0.16, 0.41], $p < .001$, and based on our pre-analysis plan were therefore analyzed separately. Regarding explanatory differences, participants were asked, "If liberals and conservatives are explained to different degrees in research, this is because scientists' political beliefs bias who they choose to focus on in their research," and "If liberals and conservatives are explained to different degrees in research, this is because conservatives and liberals really do differ on characteristics that are objectively in need of explanation" (reverse coded) (*1 = strongly disagree*, *5 = strongly agree*). These items were weakly correlated, $r(194) = 0.18$, 95% CI [0.04, 0.31], $p = .012$, hence they were likewise analyzed separately.

After forecasters reported their effect size predictions and beliefs about politics in science, the survey revealed the numeric effect sizes for explanatory and evaluative differences as captured in Study 1b. Participants were then re-surveyed regarding all of the beliefs above (excluding their numeric effect size predictions) and reported demographic characteristics including their academic discipline, job rank, and political orientation on both social and economic issues (choosing from the three categories *conservative*, *moderate*, or *liberal*).

2.2. Results & discussion

2.2.1. Sensitivity analysis

A sensitivity power analysis was performed using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009). This indicated that given our sample of participants, for our key hypothesis tests regarding participants' forecasts the minimum effect size that could be detected with 80% power is (in terms of Cohen's d) between 0.21 (for the tests of predicted direction of the bias) and 0.27 (for the tests of accuracy of forecasters in predicting the estimated effect sizes from Study 1b). This range is conventionally acknowledged as small. Likewise, for changes in

Table 2
Forecasts regarding evaluative and explanatory differences.

	Forecasted evaluative differences	Forecasted explanatory differences
Mean	0.45	−0.41
Median	0.50	−0.40
SD	0.33	0.31
Min, max	−0.80, 1.80	−1.40, 0.70
IQ range	0.24, 0.60	−0.50, −0.25
N	176	173

subjective beliefs in light of new evidence, the minimum effect size that could be detected is 0.20. Overall, the forecasting survey was adequately powered.

2.2.2. Forecasting accuracy

Forecasters accurately predicted the direction of both the evaluative differences (predicted effect size of 0.45 that conservatives would be explained more negatively) and explanatory differences (predicted effect size of −0.41 that conservatives would be explained more), with both predicted effect sizes significantly different from zero, $t(175) = 18.512$, $p < .001$, $t(172) = -17.66$, $p < .001$ (see Table 2). At the same time, they significantly overestimated both effect sizes (see Fig. 2).

We performed a z-test on the average predicted directional effect sizes being different from the estimated effect sizes from Study 1b, and we found that predicted effect size for evaluative differences (0.45) is significantly different from its estimated effect size (0.33), $z = 4.77$, $p < .001$. Similarly, we found that the predicted effect size for explanatory differences (−0.41) is significantly different from its estimated effect size (−0.21), $z = -8.37$, $p < .001$. Approximately 62% of forecasters overestimated evaluative differences, while only 38% underestimated them, with the parallel figures for explanatory differences 76% and 24%, respectively.

To ensure that our results with regard to the overestimation of bias were not driven by extreme views on the part of forecasters who are not academics, we repeated our key analyses selecting only forecasters currently working in academia. We again find that the predicted effect size for evaluative differences (0.46) is significantly larger than the obtained effect size (0.33), $z = 4.80$, $p < .001$. Similarly, we found that forecasters' predicted effect size for explanatory differences

Table 3
Forecaster characteristics and direction of predictions.

	Dependent variable: predicted differences	
	Evaluative differences	Explanatory differences
	(1)	(2)
Age	0.001	0.002
	(0.003)	(0.003)
Gender	−0.047	0.058
	(0.052)	(0.049)
Academic	−0.105	0.080
	(0.132)	(0.124)
Socially liberal	−0.049	−0.087
	(0.065)	(0.061)
Socially conservative	−0.016	−0.173
	(0.126)	(0.117)
Education bachelors' degree	−0.136	0.199
	(0.350)	(0.326)
Education masters' degree	−0.095	0.202
	(0.341)	(0.318)
Education professional degree	−0.164	0.070
	(0.387)	(0.361)
Education doctoral degree	−0.096	0.193
	(0.343)	(0.320)
Constant	0.574*	−0.612*
	(0.337)	(0.315)
Observations	175	172
R^2	0.018	0.037
F statistic	0.328 (df = 9; 165)	0.685 (df = 9; 162)

Note. Standard errors in parentheses. Omitted categories: 'moderate' for political ideology on social issues, 'high school' for education; no participants in the category 'some college but no degree.'
$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$.

(−0.42) is significantly larger than the obtained effect size (−0.21), $z = -8.56$, $p < .001$.

To make sure that the forecasting results are not driven by outliers, we focused only on predictions between the 5th and the 95th percentile of the distribution of forecasts. We again found that predicted effect size for evaluative differences (0.45) is significantly larger than the obtained effect size, $z = 6.71$, $p < .001$. Similarly, we found that the predicted effect size for explanatory differences (−0.41) is significantly larger than its estimated effect size, $z = -11.6$, $p < .001$. The results also

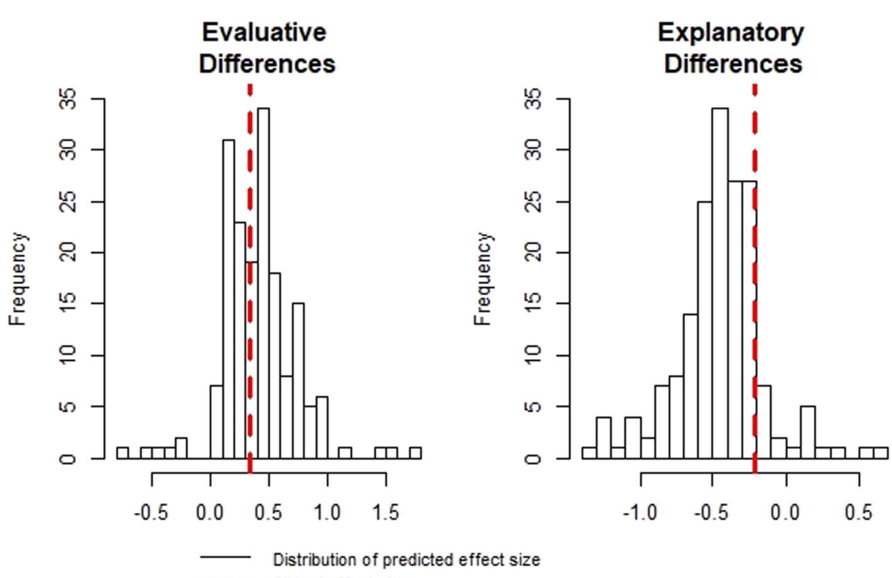

Fig. 2. Histograms of forecasted evaluative and explanatory differences.

Table 4
Forecaster characteristics and accuracy of predictions.

	Dependent variable: absolute prediction error	
	Evaluative differences	Explanatory differences
	(1)	(2)
Age	−0.0004	−0.0001
	(0.002)	(0.002)
Gender	−0.047	−0.057
	(0.038)	(0.039)
Academic	0.022	0.047
	(0.098)	(0.099)
Socially liberal	0.037	−0.013
	(0.048)	(0.048)
Socially conservative	0.158*	0.033
	(0.093)	(0.094)
Education bachelors' degree	−0.067	0.019
	(0.258)	(0.260)
Education masters' degree	−0.027	−0.108
	(0.252)	(0.253)
Education professional degree	−0.184	−0.018
	(0.286)	(0.288)
Education doctoral degree	−0.023	−0.060
	(0.253)	(0.255)
Constant	0.279	0.376
	(0.249)	(0.251)
Observations	175	172
R^2	0.037	0.037
F statistic	0.708 (df = 9; 165)	0.689 (df = 9; 162)

Note. Standard errors in parentheses. Omitted categories: 'moderate' for political ideology on social issues, 'high school' for education; no participants in the category 'some college but no degree'.
*$p < .1$; **$p < .05$; ***$p < .01$.

hold true when restricting the range of the two distributions of forecasts to between the 10th and the 90th percentiles.

2.2.3. Individual characteristics and accuracy

Personal characteristics of forecasters, such as their gender, age, education level, whether they were an academic or not, and their political ideology on social issues were not consistently associated with the direction and extremity of predictions, nor with forecasting accuracy (Tables 3 and 4). Of particular interest, for evaluative differences, being socially liberal was associated with a non-significant increase in prediction error of 0.037 with respect to the reference category "moderate," $t(165) = 0.77$, 95% CI [−0.06, 0.13], $p = .445$; being conservative was likewise associated with a non-significant increase in prediction error of 0.158 with respect to the reference category "moderate," $t(165) = 1.70$, 95% CI [−0.03, 0.34], $p = .091$. For explanatory differences, being liberal was associated with a non-significant reduction in prediction error of 0.013, $t(162) = -0.27$, 95% CI [−0.11, 0.08], $p = .792$; being conservative was associated with a non-significant increase in prediction error of 0.033, $t(162) = 0.35$, 95% CI [−0.15, 0.22], $p = .726$. These results were robust to alternative measures of political orientation (Supplement 7).

Finally, training in the psychological sciences did not appear to confer any meaningful advantage in terms of forecasting accuracy. Psychologists were no more accurate than non-psychologists at predicting the outcome of Study 1b: for evaluative differences, being a psychologist was associated with a non-significant decrease in prediction error of 0.047, $t(164) = -1.153$, 95% CI [−0.13, 0.03], $p = .251$; for explanatory differences being a psychologist was associated with a non-significant decrease in prediction error of 0.043, $t(161) = -1.03$, 95% CI [−0.13, 0.04], $p = .305$. See Supplement 7 for the full regression tables and for more detailed breakdowns of predictions by academic subdisciplines.

2.2.4. Belief updating

After exposure to the empirical results regarding evaluative and explanatory differences, forecasters updated their initial beliefs about the degree to which politics affects science in the direction suggested by the empirical findings (Table 5). They were less likely to believe that research in social psychology evaluates conservatives and liberals differently (*Ms* = 1.66 and 1.86, *SDs* = 0.56 and 0.46), $t(195) = -5.77$, 95% CI [−0.27, −0.13], $p < .001$, and explains conservatism to a greater degree than liberalism (*Ms* = 1.91 and 2.05, *SDs* = 0.80 and 0.60, respectively), $t(195) = -2.63$, 95% CI [−0.24, −0.03], $p = .009$, more likely to believe that the personal beliefs of scientists do not influence their conclusions (*Ms* = 2.29 and 2.42, *SDs* = 0.90 and 0.92, respectively), $t(194) = -3.23$, 95% CI [−0.20, −0.05], $p = .001$, and less likely to see the field as biased against conservatives (*Ms* = 4.15 and 4.06, *SDs* = 0.77 and 0.66, respectively), $t(195) = 2.04$, 95% CI [0.003, 0.20], $p = .043$.

Although beliefs about the *extent* of political bias in science changed in line with the empirical results, beliefs about the *underlying causes* of evaluative and explanatory differences did not appear to shift meaningfully. After learning about the results, forecasters were more likely to agree that evaluative differences are due to scientists' political beliefs biasing their conclusions (*Ms* = 3.07 and 3.27, *SDs* = 1.10 and 1.04), $t(195) = -4.07$, 95% CI [−0.29, −0.10], $p < .001$, and marginally more likely to agree that such differences are due to objective differences between conservatives and liberals (*Ms* = 3.14 and 3.22, *SDs* = 1.09 and 1.05), $t(194) = -1.79$, 95% CI [−0.16, 0.01], $p = .075$. This in an inconclusive pattern of results since these beliefs are to some extent in opposition (albeit not mutually exclusive), the first statement making attributions to politically biased perceptions and the second to an objective underlying reality. At the same time, there was no significant updating of beliefs regarding whether explanatory differences are due to scientists political beliefs (*Ms* = 3.95 and 3.90,

Table 5
Initial beliefs and updated beliefs.

	Original beliefs	Updated beliefs		
	Mean (SD)	Mean (SD)	Difference (*p*-value)	*N*
Who's evaluated more negatively? (from 1 "conservatives" to 5 "liberals")	1.66 (0.56)	1.86 (0.46)	0.20 (3×10^{-8})	196
Who's explained more? (from 1 "conservatives" to 5 "liberals")	1.91 (0.80)	2.05 (0.60)	0.14 (0.009)	196
Political beliefs do not bias scientific research (from 1 "disagree" to 5 "agree")	2.29 (0.90)	2.42 (0.92)	0.13 (0.001)	195
Is social psychology politicized? (from 1 "biased against liberals" to 5 "biased against conservatives")	4.15 (0.77)	4.06 (0.66)	−0.10 (0.043)	196
Scientist's political beliefs bias research conclusions (from 1 "disagree" to 5 "agree")	3.07 (1.10)	3.27 (1.04)	0.19 (6.8×10^{-5})	196
Cons. and lib. objectively differ in characteristics (from 1 "disagree" to 5 "agree")	3.14 (1.09)	3.22 (1.05)	0.08 (0.075)	195
Political beliefs bias the focus of research (from 1 "disagree" to 5 "agree")	3.95 (0.90)	3.90 (0.89)	−0.05 (0.309)	196
Explanatory differences are due to objective differences (from 1 "disagree" to 5 "agree")	3.29 (1.16)	3.36 (1.08)	0.06 (0.115)	194

Note. All the beliefs were expressed on a scale ranging from 1 to 5. Differences are computed subtracting original beliefs from updated beliefs. *p*-values in parentheses refer to the paired *t*-test of equal means (null hypothesis: difference = 0).

SDs = 0.90 and 0.89), $t(195) = 1.02$, 95% CI [−0.05, 0.15], $p = .309$, or objective differences between conservatives and liberals (*Ms* = 3.29 and 3.36, *SDs* = 1.16 and 1.08), $t(193) = 1.58$, 95% CI [−0.02, 0.14], $p = .115$. This suggests reasonable limits to the extent to which beliefs changed in light of the evidence. Scientists appear to have adjusted their beliefs about the degree of political bias in science, which are logically linked to the obtained effect sizes for evaluative and explanatory differences, but not about the underlying reasons for such biases, which the findings from Study 1b do not speak to directly.

In sum, independent scientists presented with the study's methods and materials were able to accurately predict both the direction and general magnitude of the explanatory and evaluative differences captured in Study 1b. At the same time, they significantly overestimated both effect sizes, especially for explanatory differences. Notably, although predicted effect sizes were significantly larger than the obtained ones, the difference was not particularly dramatic. Further, although forecasters were asked to predict effect sizes on a continuous numeric scale, they may have thought about the decision in categorical terms such as “small-medium-large” using heuristic benchmarks that do not necessarily map onto the magnitude of even reliable social-psychological effects (Richard, Bond, & Stokes-Zoota, 2003). Regarding this concern, it is worth pointing out that neither educational attainment nor job rank moderated forecasting accuracy. It remains possible however that, despite the anonymous format of the survey, forecasters erred on the side of overestimating the degree of bias in order to avoid seeming to underestimate the problem of politically biased science. Future research should examine scientists' forecasts under conditions designed to further minimize social desirability concerns (John, Loewenstein, & Prelec, 2012). Contrary to our expectation that political orientation would moderate expectations about research outcomes, politically liberal and conservative academics did not differ in the accuracy of their forecasts. Finally, regardless of their own political orientation, scientists updated relevant beliefs about politics in science in light of the new evidence.

3. General discussion

The present investigation provides the first systematic empirical evidence that political values can affect the end products of the scientific process — the research reports themselves. In scientific abstracts from social psychology, conservatives and conservative ideas are described significantly more negatively than liberals and liberal ideas. At the same time, conservatives are more likely to be treated as a target of explanation than are liberals. In a forecasting survey, scientists accurately anticipated the direction and general magnitude of both effects, but significantly overestimated both, predicting that evaluative and explanatory differences in the research reports would be somewhat greater than they actually were. A comparison of their stated beliefs both before and after knowing the results of the empirical investigation indicates that scientists revised their views regarding the extent to which politics shapes science in light of the evidence.

Although moderate in size statistically by conventional standards (Cohen, 1988; Sawilowsky, 2009), it is important not to understate the implications of the evaluative and explanatory differences in the scientific research reports. The obtained effect sizes are comparable to those typically found in research on psychology (Richard et al., 2003). In terms of the consequential validity of our findings (Messick, 1995; Rosenthal, 1990), it is useful to place them in the context of effect sizes associated with discrimination against women and underrepresented racial minorities, which can have societally important effects even when modest in size statistically (Greenwald, Banaji, & Nosek, 2015), especially when bias accumulates across many decisions over time (Martell, Lane, & Emrich, 1996). The same point applies to biases against conservatives and conservative ideas within academia (Inbar & Lammers, 2012). Therefore, the present research suggests that collectively and individually, scientists should be aware of and seek to consciously correct for the tendency to focus on conservatives as a target of explanation and explain them in negative terms. An analogy can again be drawn to the literature on intergroup prejudice, which identifies awareness and motivation as the keys to breaking the link between psychological biases and discriminatory actions (Devine, Monteith, Zuwerink, & Elliot, 1991; Monteith, Lybarger, & Woodcock, 2009). At the same time, in considering the practical significance of the findings, researchers should keep in mind that these are overall estimates which may not generalize across all research contexts and topics.

Notably, several other predictions regarding politics and science failed to find evidentiary support. Our hypothesis that politically conservative raters would perceive a liberal bias in scientific abstracts not seen by liberal raters, was not born out by the data. Similarly, the political values of scientists failed to moderate their forecasts about the politics of research abstracts. Finally, in a supplementary study tracking publication outcomes, we found no evidence that conference abstracts that evaluated conservatives negatively relative to liberals, and treated conservatives as more in need of explanation than liberals, were any more likely to eventually appear in prestigious academic journals (Supplement 6).

3.1. Limitations and future directions

The present results represent initial empirical evidence of a political overtone to scientific research reports in social psychology, but are also open to interpretation. As noted earlier, evaluative differences in scientific reports could reflect liberal bias, or conservatism being objectively correlated with characteristics widely regarded as undesirable, such as racism and close-mindedness (the “reality has a liberal bias” hypothesis). Explanatory differences, too, might have more benign instigators than viewing conservatives as abnormal and in need of explanation.

Although we provide tests of one potential dimension of political bias — how liberals and conservatives are portrayed in research — this is merely one among many aspects of politicized research (Brandt & Proulx, 2015; Brandt & Spälti, 2018; Jussim et al., 2016). Future meta-scientific investigations should code research topics at a more granular level to examine what is studied and how research questions are framed, characterizations of competition and incentives, and conclusions about intergroup bias, human rationality, and other topics.

Diverse methodological approaches should be employed to capture the various ways in which political values might intrude into scientific research. For instance, semantic analyses of large databases of scientific articles could be used to assess the valence of the words that appear near to “conservative” and “liberal” (Holtzman, Schott, Jones, Balota, & Yarkoni, 2011). Databases of academic citations (Uzzi, Mukherjee, Stringer, & Jones, 2013) could be leveraged to examine whether politically unpopular ideas and scholars who espouse them are excluded from the intellectual network over time.

Given anecdotal cases of refuted science that continues to be cited (Collins, 2014; Jussim, 2015) and evidence that laypeople set different evidentiary thresholds for desired and undesired conclusions (Ditto & Lopez, 1992), future work should systematically examine citation rates over time for published studies whose findings are politically favorable to liberal or conservative ideology, and which are subsequently refuted or not. Disproven research whose conclusions favor liberal policies may enjoy a higher citation rate than invalidated work supporting conservative policies. The political implications of the finding may also affect willingness to accept the null hypothesis (Greenwald, 1975). In other words, null effects, typically difficult to publish and poorly cited, may be welcomed by the academic community when highly congenial to liberal sensibilities. For instance, an article showing a lack of gender differences on a specific dimension may be more cited (and less easily refuted) than an article reporting a lack of gender discrimination in a particular context.

Data analytic choices represent a potential entry point for the influence of political values on research. As there are typically multiple defensible ways to analyze the same dataset (Bakker, van Dijk, & Wicherts, 2012; Gelman & Loken, 2014; Simmons, Nelson, & Simonsohn, 2011), scientists may (perhaps unconsciously) choose an approach that leads to a politically appealing conclusion. If so, there should be a greater degree of publication bias, as reflected in a p-curve or p-uniform analysis (Simonsohn, Nelson, & Simmons, 2014; van Aert, Wicherts, & van Assen, 2016), evident in quantitative articles reporting conclusions favorable to liberalism. Research on politicized topics may also be more likely to use covariates to achieve significance (Lenz & Sahn, 2017), and scientists' political orientation may moderate their effect size estimates on politicized topics when many analysts use the same dataset to test the same hypothesis (Silberzahn et al., in press).

Our forecasting survey had some significant limitations that should be factored into the design of future studies. The use of social media to recruit participants online for a survey on "politics in science" may have oversampled individuals who perceive academic research as politically biased. Future forecasting surveys should employ representative sampling to more accurately gauge scientific opinion both before and after key empirical investigations are conducted. Completion rates were also lower than hoped, with 65% of individuals who started the forecasting survey finishing it and 35% discontinuing. This could have been due to the challenging nature of the effect size estimation task, lack of clarity in the instructions, and/or the unexpected length of the survey. In addition, the belief updating observed could have been partly due to experimental demand, although the anonymous reporting conditions would arguably mitigate this problem to some extent. That some beliefs about politics in science (i.e., beliefs about the extent of bias) shifted in light of the evidence while others (i.e., beliefs about the underlying causes of bias) did not suggest that the belief updating that did occur was genuine. However, future work is needed to determine the degree to which belief change among scientists in light of new empirical evidence is genuine and durable. Future forecasting studies should also examine scientists' beliefs about the numerous potential aspects of politicized science described above.

3.2. Conclusion

The extent to which investigations of politically charged topics are affected by the generally liberal values of the scientists who carry out the research has been a topic of considerable debate inside and outside academia for some time. The present research informs this discussion by providing the first empirical evidence of systematic effects of political values on research reports in a scientific field. At the same time, a great deal of further meta-scientific work is needed to uncover when politics does (and does not) play a role in the scientific process, and what might potentially be done about it.

Scientific debates often prove intractable, even in the face of accumulating empirical evidence (Kahneman & Klein, 2009; Kuhn, 1962; Mellers, Hertwig, & Kahneman, 2001). We hope that our forecasting survey serves not only as a means of addressing the longstanding public debate over alleged liberal bias in academic research, but also as a model for future scientific exchanges more generally (Tetlock et al., 2014). A similar approach might be profitably applied to other prominent controversies, such as the extent to which measures of implicit associations predict relevant outcomes (Greenwald et al., 2015; Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2015), whether unconsciously activated concepts exert a robust and reliable influence over human behavior (Bargh, 2012; Harris, Coburn, Rohrer, & Pashler, 2013), and other present and future questions that capture the attention of scientists.

Open practices

All measures, manipulations, and exclusions in this research are fully disclosed in the paper. To maximize statistical power, ten years of SPSP abstracts available online were included in the analysis, thousands of independent raters were recruited using a crowdsourced internet marketplace, and the forecasting survey was promoted as widely as possible online in order to recruit as many respondents as we could. We did not collect further observations after conducting the analyses. The analysis plans were pre-registered and the data and materials are publicly posted at https://osf.io/zhf98/, https://osf.io/vtyg4/, and https://osf.io/jh47m/.

Author contributions

E. Uhlmann developed the study concept. All authors contributed to the study design. O. Eitan, D. Viganola, & Y. Inbar carried out the data collections. D. Viganola, Y. Inbar, T. Pfeiffer, A. Dreber, M. Johannesson, S. Thau, & O. Eitan performed the data analysis and interpretation. O. Eitan, D. Viganola, Y. Inbar, & E. Uhlmann drafted the manuscript, and all authors provided critical revisions and approved the final version of the manuscript for submission.

Acknowledgements

Eric Uhlmann and Stefan Thau gratefully acknowledge an R&D grant from INSEAD (2520), which funded the data collections for this project. Domenico Viganola gratefully acknowledges financial support from the Jan Wallanders and Tom Hedelius Foundation (H2015-0425:1). Anna Dreber gratefully acknowledges financial support from the Knut and Alice Wallenberg Foundation (through a Wallenberg Academy Fellows grant; A.D. is a Wallenberg Academy Fellow). Anna Dreber and Magnus Johannesson gratefully acknowledge financial support from the Jan Wallander and Tom Hedelius Foundation (Svenska Handelsbankens Forskningsstiftelser grants P2015-0001:1, P2014-0214:1) and the Swedish Foundation for Humanities and Social Sciences (grant NHS14-1719:1). Thomas Pfeiffer gratefully acknowledges financial support from Marsden Fund Grants 16-UOA-190 and 17-MAU-133.

Appendix A. Supplementary data

Online supplements for this article can be found online at https://doi.org/10.1016/j.jesp.2018.06.004.

References

Abramowitz, S. I., Gomes, B., & Abramowitz, C. V. (1975). Publish or politic: Referee bias in manuscript review. *Journal of Applied Social Psychology, 5*(3), 187–200.
van Aert, R. C. M., Wicherts, J. M., & van Assen, M. A. L. M. (2016). Conducting meta-analyses based on p-values: Reservations and recommendations for applying p-uniform and p-curve. *Perspectives on Psychological Science, 11*, 713–729.
Al-Gharbi, M. (in press). Race and the race for the white house: On social research in the age of Trump. *The American Sociologist*.
Arkes, H., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "would Jesse Jackson 'fail' the implicit association test?". *Psychological Inquiry, 15*(4), 257–278.
Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science, 7*, 543–554.
Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2004). No place for nostalgia in science: A response to Arkes and Tetlock. *Psychological Inquiry, 15*, 279–289.
Bargh, J. A. (2012). Priming effects replicate just fine, thanks. *Psychology Today*. Retrieved from https://www.psychologytoday.com/blog/the-natural-unconscious/201205/priming-effects-replicate-just-fine-thanks.
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.
Brandt, M. J., & Proulx, T. (2015). QTIPs: Questionable theoretical and interpretive practices in social psychology. *Behavioral and Brain Sciences, 38*, 19–20.

Brandt, M. J., & Spälti, A. K. (2018). Norms and explanations in social and political psychology. In J. T. Crawford, & L. Jussim (Eds.). *Frontiers of social psychology series: Politics of social psychology* (pp. 26–43). Psychology Press.
Brown, A. W., & Allison, D. B. (2014). Using crowdsourcing to evaluate published scientific literature: Methods and example. *PLoS One, 9*(7), e100647.
Buhrmester, M. D., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3–5.
Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., ... Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science, 351*(6280), 1433–1436.
Cardiff, C., & Klein (2005). Faculty partisan affiliations in all disciplines: A voter-registration study. *Critical Review, 17*, 237–255.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J.: Lawrence Erlbaum.
Collins, K. (2014). Why researchers keep citing retracted papers. *Quartz*. Retrieved from https://qz.com/583497/researchers-keep-citing-these-retracted-papers/.
DellaVigna, S., & Pope, D. G. (2016). *Predicting experimental results: Who knows what?* (Unpublished manuscript).
DellaVigna, S., & Pope, D.G. (in press). What motivates effort? Evidence and expert forecasts. *Review of Economic Studies*.
Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology, 60*, 817–830.
Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: The use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology, 63*, 568–584.
Dreber, A., Pfeiffer, T., Almenberg, J., Isaksson, S., Wilson, B., Chen, Y., ... Johannesson, M. (2015). Using prediction markets to estimate the reproducibility of scientific research. *Proceedings of the National Academy of Sciences, 112*, 15343–15347.
Duarte, J. L., Crawford, J. T., Stern, C., Haidt, J., Jussim, L., & Tetlock, P. (2015). Political diversity will improve social and personality psychological science. *Behavioral and Brain Sciences, 38*, 1–13.
Eagly, A. (2014). Is social psychology biased against conservatives? *Psychology Today*. Retrieved from https://m.psychologytoday.com/blog/sound-science-sound-policy/201411/is-social-psychology-biased-against-conservatives.
Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.
Gelman, A., & Loken, E. (2014). The statistical crisis in science. *American Scientist, 102*, 460–465.
Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*, 493–498.
Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82*, 1–20.
Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology, 108*, 553–561.
Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*, 17–41.
Gross, N., & Simmons, S. (2006). *Americans' attitudes toward academic freedom and liberal "bias" in higher education.* Working Paper, May 22, 2006. Harvard University and George Mason University.
Haidt, J. (2011, January). The bright future of post-partisan social psychology. Talk given at the annual meeting of the Society for Personality and Social Psychology, San Antonio, TX. Retrieved from http://people.virginia.edu/~jdh6n/postpartisan.html.
Harris, C. R., Coburn, N., Rohrer, D., & Pashler, H. (2013). Two failures to replicate high-performance-goal priming effects. *PLoS One, 8*, e72467. http://dx.doi.org/10.1371/journal.pone.0072467.
Hegarty, P., & Pratto, F. (2001). The effects of social category norms and stereotypes on explanations for intergroup differences. *Journal of Personality and Social Psychology, 80*, 723–735.
Holtzman, N. S., Schott, J. P., Jones, M. N., Balota, D. A., & Yarkoni, T. (2011). Exploring media bias with semantic analysis tools: Validation of the contrast analysis of semantic similarity (CASS). *Behavior Research Methods, 43*, 193–200.
Inbar, Y., & Lammers, J. (2012). Political diversity in social and personality psychology. *Perspectives on Psychological Science, 7*, 496–503.
John, L., Loewenstein, G. F., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth-telling. *Psychological Science, 23*, 524–532.
Jussim, L. (2015). Slow & nonexistent scientific self-correction in psychology. *Psychology Today*. Retrieved from https://www.psychologytoday.com/us/blog/rabble-rouser/201505/slow-nonexistent-scientific-self-correction-in-psychology.
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.
Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64*(6), 515–526. http://dx.doi.org/10.1037/a0016755.
Klein, D. B., & Stern, C. (2005). Professors and their politics: The policy views of social scientists. *Critical Review, 17*, 257–303.
Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences, 27*, 313–327.
Krugman, P. (2014, April 18). On the liberal bias of facts. *NYTimes.com*. Retrieved from https://krugman.blogs.nytimes.com/2014/04/18/on-the-liberal-bias-of-facts/?_r=0.
Kuhn, T. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.
Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480–498.
Lai, M. H. C., & Kwok, O. M. (2014). Standardized mean differences in two-level cross-classified random effects models. *Journal of Educational and Behavioral Statistics, 39*, 282–302.
Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos, & A. Musgrave (Eds.). *Criticism and the growth of knowledge*. Cambridge: University Press.
Lenz, G., & Sahn, A. (2017). *Achieving statistical significance with covariates.* (Unpublished manuscript). Retrieved from https://osf.io/preprints/bitss/s42ba/.
Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*, 2098–2109.
Martell, R. F., Lane, D. M., & Emrich, C. (1996). Male–female differences: A computer simulation. *American Psychologist, 51*, 157–158.
McClintock, C. G., Spaulding, C. B., & Turner, H. A. (1965). Political orientation of academically affiliated psychologists. *American Psychologist, 20*, 211–221.
Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science, 12*, 269–275.
Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749.
Miller, D. T., Taylor, B., & Buck, M. L. (1991). Gender gaps: Who needs to be explained? *Journal of Personality and Social Psychology, 61*, 5–12.
Monteith, M. J., Lybarger, J. E., & Woodcock, A. (2009). Schooling the cognitive monster: The role of motivation in the regulation and control of prejudice. *Social and Personality Compass, 3*, 211–226.
Oswald, F., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Revisiting the predictive validity of the Implicit Association Test. *Journal of Personality and Social Psychology, 105*(2), 171–192.
Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science, 23*, 184–188.
Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*, 411–419.
R Core Team (2016). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.
Redding, R. (2001). Sociopolitical diversity in psychology: The case for pluralism. *American Psychologist, 56*, 205–215.
Richard, F. D., Bond, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7*(4), 331–363.
Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist, 45*, 775–777.
Rothman, S., & Lichter, S. R. (2008). The vanishing conservative: Is there a glass ceiling? In R. Maranto, R. E. Redding, & F. M. Hess (Eds.). *The politically correct university: Problems, scope, and reforms* (pp. 60–76). AEI Press.
Sawilowsky, S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods, 8*(2), 467–474.
Sears, D. O. (1994). Ideological bias in political psychology: The view from scientific hell. *Political Psychology, 15*, 547–556.
Sherman, D. K., & Cohen, G. L. (2002). Accepting threatening information: Self-affirmation and the reduction of defensive biases. *Current Directions in Psychological Science, 11*, 119–123.
Silberzahn, R., Uhlmann, E. L., Martin, D., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., Fong, N., Gamez-Djokic, M., Glenz, A., Gordon-McKeon, S., Heaton, T. J., Hederos, K., E., Heene, M., Hofelich Mohr, A. J., Högden, F., Hui, K., Johannesson, M., Kalodimos, J., Kaszubowski, E., Kennedy, D., Lei, R., Lindsay, T. A., Liverani, S., Madan, C. R., Molden, D., Molleman, E., Morey, R. D., Mulder, L. B., Nijstad, B. A., Pope, N. G., Pope, B., Prenoveau, J. M., Rink, F., Robusto, E., Roderique, H., Sandberg, A., Schlüter, E., Schönbrodt, F. D., Sherman, M. F., Sommer, S., Sotak, K., Spain, S., Spörlein, C., Stafford, T., Stefanutti, L., Tauber, S., Ullrich, J., Vianello, M., Wagenmakers, E., Witkowiak, M., Yoon, S., & Nosek, B.A. (in press). Many analysts, one dataset: Making transparent how variations in analytical choices affect results. *Advances in Methods and Practices in Psychological Science*.
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allow presenting anything as significant. *Psychological Science, 22*, 1359–1366.
Simonsohn, U. (2013). Just post it: The lesson from two cases of fabricated data detected by statistics alone. *Psychological Science, 24*(10), 1875–1888.
Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file drawer. *Journal of Experimental Psychology: General, 143*, 534–547.
Sniderman, P. M., & Tetlock, P. E. (1986). Symbolic racism: Problems of motive attribution in political analysis. *Journal of Social Issues, 42*, 129–150.
Tarman, C., & Sears, D. O. (2005). The conceptualization and measurement of symbolic racism. *The Journal of Politics, 67*(3), 731–761.

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. (2005). *Expert political judgment: How good is it? How can we know?* Princeton: Princeton University Press.
Tetlock, P. E., & Mellers, B. (2014). Judging political judgment. *Proceedings of National Academy of Sciences, 111*(32), 11574–11575.
Tetlock, P. E., Mellers, B., Rohrbaugh, N., & Chen, E. (2014). Forecasting tournaments: Tools for increasing transparency and the quality of debate. *Current Directions in Psychological Science, 23*(4), 290–295.
Tetlock, P. E., & Mitchell, G. (2009). Implicit bias and accountability systems: What must organizations do to prevent discrimination? In B. M. Staw, & A. Brief (Vol. Eds.), *Research in Organizational Behavior*. *Vol. 29*. (pp. 3–38). New York: Elsevier.
Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. F. (2013). Atypical combinations and scientific impact. *Science, 342*, 268–472.
Vallone, R., Ross, L., & Lepper, M. R. (1985). The hostile media phenomenon: Biased perception of media bias in coverage of the Beirut massacre. *Journal of Personality and Social Psychology, 49*, 577–585.
Van't Veer, A., & Giner-Sorolla, R. (2016). Pre-registration in social psychology: A discussion and suggested template. *Journal of Experimental Social Psychology, 67*, 2–12.
Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7*, 627–633.
Wicherts, J. M., & Bakker, M. (2012). Publish (your data) or (let the data) perish! Why not publish your data too? *Intelligence, 40*, 73–76.
Wolfers, J., & Zitzewitz, E. (2004). Prediction markets. *Journal of Economic Perspectives, 18*(2), 107–126.