How robust is evidence of partisan perceptual bias in survey responses? A new approach for studying expressive responding

Omer Yair
Hebrew University of Jerusalem
The Federmann School of Public Policy and Government
Mt, Scopus, Jerusalem, Israel, 9190501
Tel: +972-2-5880015
omer.yair@mail.huji.ac.il

Gregory A. Huber (Corresponding Author)
Yale University, Professor
Department of Political Science
Institution for Social and Policy Studies
77 Prospect Street, PO Box 208209
New Haven, CT 06520-8209
Tel: 203-432-5731
gregory.huber@yale.edu

Running header: An Approach to Studying Expressive Responding

Word count (text and notes): 6,498

OMER YAIR is postdoctoral scholar at the Federmann School of Public Policy and Government at the Hebrew University of Jerusalem, Jerusalem, Israel.
GREGORY A. HUBER is the Forst Family Professor and chair of political science, as well as associate director at the Center of the Study of American Politics, at Yale University, New Haven, CT, USA.

The authors thank Jamie Druckman, Stanley Feldman, Matt Graham, Leonie Huddy, Jennifer Jerit, Ari Malka, Vittorio Mérola, Steven Nicholson, Lior Sheffer, Keren L.G. Snider, seminar participants at Stony Brook University, and participants at the 2018 annual meeting of the Midwest Political Science Association for helpful comments.

This research was funded by the Institution for Social and Policy Studies at Yale University.

The authors declare no conflict of interest.

Address correspondence to Gregory A. Huber, Yale University, 77 Prospect Street, PO Box 208209, New Haven, CT 06520-8209; email: gregory.huber@yale.edu.

Abstract

Partisans often offer divergent responses to survey items ostensibly unrelated to politics. These gaps could reveal that partisanship colors perception or, alternatively, that in answering survey questions, individuals communicate partisan proclivities by providing insincere or, "expressive" responses, to send a partisan message. This study tests two techniques for reducing expressive responding that (1) avoid criticisms about using monetary incentives for accuracy, which have reduced measured partisan differences for objective facts, and (2) can be used in contexts where incentives are infeasible, such as when objective benchmarks for correct responses are unavailable. This study experimentally tests these techniques in replicating a study that found partisanship affected attractiveness evaluations. These interventions, which allow partisans to express their partisan sentiments through other survey items, substantially reduce apparent partisan differences in beauty evaluations and show standard survey items likely confound sincere partisan differences with elements of expressive responding.

Partisan orientations are highly salient in contemporary politics in many countries (e.g., Huddy, Mason, and Aarøe 2015; Huddy, Bankert, and Davies 2018) and are correlated with policy preferences in most contexts (e.g., Bartels 2002; Pew Research Center 2017). These differences are expected because partisan identities are themselves in part manifestations of underlying policy positions (e.g., Carsey and Layman 2006; Chen and Goren 2016). Scholars have identified a more disturbing pattern, however, in which partisanship also appears to alter attitudes apart from policy preferences, such as evaluations of objective facts like the state of the economy (e.g., Bartels 2002) and non-policy subjective evaluations, such as a person's beauty (Nicholson et al. 2016). What explains these unexpected partisan gaps? One view is that partisanship is such a powerful identity that it colors how individuals process identical information (e.g., facts about the economy). Another view is that in answering survey questions, individuals provide responses that also communicate their partisan proclivities, such that partisan differences may indicate "expressive responding" (Bullock et al. 2015) rather than sincere differences.

Recent scholarship reports results from experiments that use financial incentives for accuracy to assess the degree of partisan divergence in factual opinions that arises due to partisan expressive responding in survey response. In those designs, individuals are given a financial incentive to provide correct answers to factual survey items. Compared to survey responses collected without such incentives, estimated partisan differences are smaller, and sometimes vanish entirely, implying that at least some portion of the effect of partisanship on factual evaluations is an artifact of systematic measurement error induced by expressive responding (Bullock et al. 2015; Prior, Sood, and Khanna 2015). (Such expressive responding can arise for a variety of reasons, including survey responses being costless "cheap talk," low attention and engagement, etc.)

But these designs are only feasible both when an objective benchmark can be used to adjudicate whether a survey response is "correct" and when the research environment allows offering financial incentives on the basis of specific survey answers. In areas where these options are infeasible (e.g., rumors and conspiracy theories, as in Berinsky 2018[1]), scholars have tried other techniques to elicit less partisan-tinged answers, but those approaches have generally not reduced partisan differences by large amounts.[2] To some, this is evidence that partisan difference in these evaluations are sincere (e.g., Berinsky 2018), while to others it may simply indicate the relative weakness of these techniques vis-à-vis financial incentives. Additionally, the use of financial incentives has attracted further criticisms from scholars who argue it is unlikely to

---

[1] For example, the statement that "Obama is a Muslim" cannot be falsified because one cannot verify someone's faith.

[2] Berinsky (2018), in a study of partisan rumors, explored the effect of four interventions on measured rumor endorsement: (1) giving individuals the opportunity to answer partisan rumor questions or end a survey more quickly, (2) an admonition to tell the truth, (3) an admonition to tell the truth despite partisan sentiments, and (4) a list experiment. None of these interventions substantially reduced partisan differences in partisan rumor endorsement. Another approach is self-affirmation interventions (Cohen and Sherman 2014), which boosts one's self-esteem and therefore reduce motivational biases. A few studies have shown that self-affirmation can also reduce biases in the political sphere (Cohen, Aronson, and Steele 2000; Cohen et al. 2007). But the effect of this "de-biasing" technique is theoretically ambiguous: it may reduce actual motivational biases or, congruent with the logic of expressive responding, simply reduce respondents' desires to provide insincere survey responses.

reflect the non-material motivations that are normally at play in the political environment (discussed in greater detail below).

In an effort to resolve this uncertainty, this study identifies and tests two new techniques with the potential for reducing expressive responding. Importantly, these approaches can be used for non-factual items and when financial payments are infeasible. These techniques, which we test experimentally, build on ideas presented in market research that survey respondents may respond to survey items "as if" these items asked about another topic that the respondent wished to express their opinion about. In particular, one design allows respondents to "blow off steam" by expressing partisan sentiments before answering other items and a second "warning" design alerts respondents that they will be given an opportunity to provide a potentially partisan sentiment after they answer a different question.

This study applies these techniques to a replication of a recent study that found partisanship affected evaluations of physical attractiveness (Nicholson et al. 2016). Using data gathered in three independent samples, we find that these interventions reduce apparent partisan differences in evaluations of beauty, suggesting conclusions drawn from ordinary survey contexts may overstate the effects of partisanship on evaluations of physical attractiveness. More broadly, they point to the limitations of standard survey items for distinguishing sincere partisan differences from confounding expressive responding.

Partisanship, Bias, and Survey Measurement

The literature on how partisanship colors beliefs is vast and a full review is beyond the scope of this essay. Briefly, the foundational work in this area is Campbell et al. (1960), which famously described partisanship as a "perceptual screen" that affects individuals' perceptions and

information processing (133). Partisan bias therefore arises because individuals see a common reality differently. Subsequent work has expounded on this conclusion to suggest a variety of motivational biases for arriving at a partisan-congenial conclusion (e.g., Lodge and Taber 2013; Leeper and Slothuus 2014). Building on this foundational work, numerous studies document partisan differences in survey assessments both of facts (e.g., Nyhan and Reifler 2010; Jerit and Barabas 2012) and non-factual opinion (e.g., Gaines et al. 2007; Bisgaard 2015, 2019), evidence that is often used to support the claim that partisan bias colors beliefs and perceptions. Indeed, as McGrath (2017, 377) notes, the "preponderance of studies of partisan bias operate under the assumption that partisan divergence on survey questions reflects deeply rooted differences in partisan perception."

Recent scholarship, however, has challenged these conclusions. In a series of experiments scholars have shown that offering respondents money in exchange for correct answers to factual questions substantially reduces the differences in the answers provided by rival partisans (e.g., Bullock et al. 2015; Prior, Sood, and Khanna 2015; Khanna and Sood 2018). Some of these scholars have suggested that rival partisans in fact see reality rather similarly, with a large part of the differences explained by expressive responding, that is, partisans' desire to present the in-party positively and/or the out-party negatively (relatedly, see Schaffner and Luks 2018).

But other scholars have criticized the monetary incentives provided in these studies, calling into question the suggestion that these incentives in fact reveal more sincere responses and also contending that these incentives do not help us understand how people in the real world evaluate their political environment (Kahan 2016; Flynn, Reifler, and Nyhan 2017; Berinsky 2018). Relatedly, several scholars have recently suggested that "even a dramatic increase in accuracy incentives [i.e., monetary incentives] would not afford definitive insights into

participants' true beliefs" (Massey, Simmons, and Armor 2011, 280). Furthermore, these monetary incentives cannot help scholars who are interested in respondents' answers to non-factual items.

Cumulatively, the current literature provides conflicting accounts as to the causes of partisan gaps. A common thread, however, is that almost all work showing partisan bias uses survey data, but whether those data are accurate measures of actual beliefs is unknown. As a result, it is unclear whether these gaps are authentic and attest to actual differences in beliefs between rival partisans or not (see also Bullock and Lenz 2019).

Answering "Unasked Questions" and Reducing Expressive Responding

The idea of expressive survey responding introduced above centers on the assumption that a survey response provides a relatively costless means for a respondent to convey a partisan sentiment. Because distorting a survey answer away from one's own true opinion usually imposes no external cost on a respondent, she may be tempted to use survey responses to communicate a partisan message she would like to send. For example, a respondent who reports that the unemployment rate is higher than she knows it to be may be communicating that she thinks the incumbent president is doing a bad job or simply that she does not like the president. In this way, survey respondents may use a question ostensibly about unemployment to answer a different question about evaluations of the president.[3]

[3] In this sense, expressive responding resembles social desirability bias (e.g., Clifford and Jerit 2015; Kuhn and Vivyan 2018), although the psychological motivation to report distorted beliefs differs. In the case of social desirability biases examined previously, respondents provide

The traditional view that this pattern reflects sincere partisan bias assumes that partisanship colors the respondent's actual knowledge of, or perception of, the unemployment rate. But the alternative possibility highlighted here is that individuals may share common beliefs about the question at hand (e.g., unemployment rate) but report partisan-correlated differences because they are allowing their views about a different question—their attitude toward the president—to shape their reported views. Partisans have strong opinions about which party is better for the country and rival partisans substantially differ in their evaluations of particular political leaders. Given a desire to express these core beliefs about the relative superiority of one's partisan "team" vis-à-vis the partisan opposition, the respondent may use the only opportunity available to communicate that view. The respondent may overtly recognize her desire to express her view, or it may be less conscious. In either case, the key idea is that this desire might affect reported beliefs, rather than one's true beliefs, about the economy.

A recent paper in marketing science shows, outside of the political context, that survey respondents' answers to questions are sometimes affected by ideas they want to convey but are not asked about by the researcher (Gal and Rucker 2011). Gal and Rucker point to a hypothetical case in which a person receives bad service at a restaurant but is asked only about the

---

insincere responses to present a positive social image of themselves, while in the case of responses intended to convey a partisan message, the motivation is to support one's partisan allies. A related literature argues that questions are sometimes interpreted in ways that depart from the researcher's underlying measurement goals (see Tourangeau, Rips, and Rasinski 2000, Chapter 2). Our perspective is different in that we suggest individuals sometimes do not misinterpret the question but instead choose to answer a different one.

restaurant's food quality. She may wish to convey her unhappiness about the service, but can only convey her unhappiness by reporting lower food quality, a dimension on which she is not unhappy. These cases of "response substitution" take place when respondents provide "an answer to a question that reflects attitudes or beliefs that they want to convey but that the researcher has not asked about" (186). This study tests whether a similar logic may explain some partisan gaps in (political) survey responses: A respondent wishes to express her partisan feelings and uses available items to express a partisan sentiment because no avenue exists by which to communicate those sentiments.[4] To our knowledge, no other studies have given respondents an opportunity to directly express their partisan sentiments before answering other items to see if doing so affects subsequent estimates of partisan differences.

In the marketing context, Gal and Rucker test two techniques to reduce the effect of "response substitution." The first allowed respondents to answer the "unasked question" prior to answering other items. The second informed respondents that they would have a future opportunity to do so. Both techniques therefore allow respondents to convey a message they might wish to convey but otherwise would not be given the opportunity to do, i.e., to answer the previously "unasked question." Gal and Rucker find that both techniques substantially reduce response substitution. This study builds on these two novel techniques and tests whether they can be similarly effective in the political environment. In particular, before asking partisans questions that they may use to express a partisan sentiment, we gave respondents a chance to directly express those partisan views or alerted them that they would soon have the chance to express

[4] This approach embodies logic somewhat similar to Krupnikov et al.'s (2016) "saving face" technique for reducing social desirability bias.

partisan-related views. If the desire to answer "unasked" questions explains some portion of measured partisan differences, then the differences in responses between rival partisans should be smaller following these interventions vis-à-vis a standard survey design.

Still, it is unlikely that these interventions will completely remove the influence of partisanship on measured attitudes. For one, these are relatively modest treatments; it may be that some desire to express partisan sentiments will remain, and because survey responses are largely costless, even weak motivations to "cheerlead" can continue to affect measured responses. Nor will these interventions undo partisan differences if those differences originate in sincere partisan bias. Accordingly, changes in measured partisan gaps following these treatments provide a lower-bound estimate of the magnitude of expressive responding.

Design and Method

To test the effect of the proposed interventions we replicated a prior study that finds partisan differences in perceptions of ostensibly non-political judgments. We chose Nicholson et al. (2016), which found that partisanship affects evaluations of physical attractiveness, for three reasons. First, it has been interpreted as demonstrating that partisanship alters evaluations of an ostensibly orthogonal characteristic, attractiveness. Second, attractiveness assessments are the kind of outcome for which the use of financial incentives for "accurate" responses is likely infeasible. Judgments of beauty are subjective, and offering people money in exchange for their correct attractiveness assessment is moot. Third, related research (discussed below) provides some evidence that the pattern of partisan differences in attractiveness evaluations may arise due to expressive responding.

The Nicholson et al. (2016) study was conducted online shortly before the 2012 presidential election. Democratic and Republican respondents were shown a picture of a target person and some written information about that person. Respondents were shown a fixed opposite sex photograph. They were then asked to rate the person's attractiveness. The exact question prompt was "The purpose of this question is to learn more about the kinds of characteristics people find attractive. Please look at the picture above and the "About Me" information and rate the attractiveness of the person." Responses were gathered on a 7-point scale from extremely attractive to extremely unattractive. Individuals were assigned to a control condition or one of two partisan conditions. In the control condition, beneath the text "About me:", the person was described as "friendly", "smart", and a "runner." In the Democratic condition, the text "Obama supporter" was added to the list of attributes, while in the Republican condition, the text added was "Romney supporter." (888–890)

Nicholson et al. find that, compared to the control condition, Democratic and Republican respondents (including partisan leaners) evaluate the target person as less physically attractive when the target person supported the out-party's presidential nominee. Women respondents also evaluated the target person as more physically attractive when he supported the in-party's nominee compared to the control condition. Net, Democrats and Republicans evaluated co-partisans as more attractive than out-partisans (890–894). The authors attribute these results as evidence of partisan bias, suggesting partisanship "introduces bias into judgments of physical attractiveness," which shows that that "the perceptual screen induced by party identification travels far beyond political choices" (895).

One reason to suspect that these data do not indicate true partisan difference in attractiveness evaluations comes from a related study of online dating by Huber and Malhotra

(2017, Study 1). In that study, respondents viewed online dating profiles constructed from randomly selected photographs and other respondent characteristics, including partisan affiliations. After viewing each profile, respondents answered several items, including an assessment of the person's values and how attractive the person was. They find that partisans had greater interest in dating co-partisans, but did not systematically rate co-partisans more attractive than out-partisans (275–276). One possible explanation for this null effect is that the items asked before the attractiveness item allowed respondents to express their partisan feelings. But because the study did not randomize the presence of the other items before the attractiveness question, it is not possible to estimate the effect asking different questions before the attractiveness question.

Accordingly, to test the argument that the partisan bias estimates in Nicholson et al.'s study are affected by expressive responding, we use a research design that attempts to closely replicate the approach taken by Nicholson et al to create a baseline "partisan bias" estimate. We then test whether including items that allow or warn about the future opportunity to express potentially partisan sentiments reduces the estimated effect of partisanship on attractiveness evaluations. These interventions are designed to alter the survey context to encourage respondents to think about expressing their physical attractiveness and value judgments separately.

Sampling

We fielded the experiment three times using three different sample. The first iteration was conducted on Mechanical Turk (MTurk) and was fielded on July 13, 2017. 1,004 respondents completed the survey. Analysis is restricted to Democrats or Republicans respondents assigned to either the control condition or the two interventions analyzed here, resulting in a sample of

355 Democrats and 147 Republicans.[5] Because of a survey programming error, we cannot identify partisan "leaners" in this sample.

While the MTurk sample is diverse, it is not representative of the larger US population (e.g., Berinsky, Huber, and Lenz 2012). Descriptions of sample demographics for this and other samples, as well as test of randomization, appear in Appendix Table 1. Accordingly, our second data collection effort used respondents recruited by Lucid (https://luc.id/), an online survey platform that has been found to provide samples that are suitable for conducting social science research (Coppock and McClellan 2019). Unlike the MTurk sample, this sample was balanced by partisanship. The survey was fielded on August 11–14, 2017. 1,025 respondents completed the survey. There are 378 Democrats (including leaners) and 388 Republicans (including leaners) in this analysis sample.

Finally, we fielded the experiment a third time using a sample provided by Lucid and designed to match Census benchmarks using Lucid's quota sampling procedure. The survey was fielded between September 28 and October 4, 2017. 2,048 respondents completed the survey; 562 Democrats (including leaners) and 444 Republicans (including leaners) are included in this analysis sample.

---

[5] Two additional treatments that were fielded in the second Lucid sample are discussed in Appendix Section A. As explained in the appendix, these treatments, in which we altered the response format, did not perform as expected because they did not produce baseline differences in attractiveness evaluations.

Experimental Procedure

Where possible, we followed the procedures from Nicholson et al. (2016) exactly. All manipulations are summarized in Table 1. As in that study, all respondents were shown a picture of a target person along with some information about that person. We used the images from Nicholson et al.'s study (provided by the original authors) and matched respondents to opposite gender pictures. As in the original study, underneath an "About me:" box the person was described as "friendly", "smart", and a "runner". Participants were then asked to rate the person's attractiveness using the same 7-point attractiveness scale. This is the baseline control.

[Table 1 here]

In Nicholson et al.'s study, respondents were randomly assigned to this baseline control condition, a Democratic treatment, or a Republican treatment. We refer to these manipulations as baseline conditions. In their study, partisan treatments were operationalized as being a supporter of either of the major party nominees in the 2012 presidential election. Because we fielded our experiments in 2017, it was not appropriate to use Obama or Romney support in 2012 as signals of contemporary partisanship. Accordingly, we provided respondents in the partisan treatment groups with one of four different partisan profiles.

The two Republican profiles indicated that the target person was either a "Republican" or "Supported Trump in the 2016 election." Similarly, the two Democratic profiles indicated that the target person was either a "Democrat" or "Supported Clinton in the 2016 election."[6] To facilitate presentation of the findings, the main results section below combines the results

[6] It was unclear ex ante whether mentioning a party or a (different) candidate (post-election) would be equivalent to the original Nicholson et al.'s treatment.

obtained for each method as "Republican profiles" and "Democratic profiles." Notably, this is not an exact replication of Nicholson et al. because we use different signals of partisanship and because the political context changed between 2012 and 2017.

We added two interventions to this baseline set of conditions. In the "blow-off-steam" arm respondents are given the chance to express their partisan sentiments about the person shown before assessing his or her attractiveness. This follows the logic, outlined by Gal and Rucker (2011), that allowing respondents to answer a potential "unasked question" would reduce subsequent response substitution. In particular, after showing respondents the picture of the target person and the information section, respondents were asked, on a 7-point scale, whether they believe the target person has good values.[7] We chose this item because people regard political choices as indicating important differences in values more than attractiveness. As such, conveying a message about values may reduce partisans' desires to provide insincere attractiveness responses.

This question was followed, on the next page, by the item about the target person's attractiveness. Compared to the baseline conditions, the treatments in the "blow-off-steam" arm differ only in giving individuals an opportunity to express a partisan message (i.e., whether the person shown has good values) before evaluating the person's attractiveness. If this opportunity to express one's partisan views reduces the desire to engage in response substitution it should diminish measured partisan differences in attractiveness evaluations.

---

[7] Specifically: "Please look at the picture above and the "About Me" information. Does this person have good values?" (from "extremely good" to "extremely bad" values).

In the "warning" treatments arm, we told respondents they would have a future opportunity to express an additional, potentially partisan, message. Specifically, before presenting the target individual and descriptive text, respondents were told that they would be shown, on the next page, a person's picture, and would be asked both about the person's attractiveness and their values (See exact wording in Appendix Section B.)

On the following page respondents were then shown the target person's picture and the "About Me" information, and were asked first about the target person's attractiveness and then whether this person had good values. Following Gal and Rucker (2011), this condition tests whether informing respondents about a future opportunity to express a partisan message reduces expressive responding. Because the respondent has not yet had the chance to "blow off steam," however, it is unclear whether this future opportunity will be sufficient to reduce the motivation to engage in partisan cheerleading prior to actually expressing those views.

In all three implementations of the experiment individuals were independently randomized into the three partisan conditions (no partisanship, Democrat, or Republican) and three question response conditions (baseline, blow-off-steam, and warning). In the first implementation (MTurk), the baseline condition was oversampled to adequately power separate comparisons of the two other conditions to this baseline. In the subsequent fieldings treatment assignment was equally balanced. The pooled treatment effect analysis accounts for this different rate of assignment across samples.[8]

[8] Weights are inverse probability weights. In the two Lucid samples, all observations are equally weighted. In the MTurk sample, control group observations have a probability of assignment of 43% and each treatment has a probability of approximately 28.5%.

Analysis Strategy and Results

To test whether the opportunity to express partisan sentiments affects measured partisan bias, we compare estimates of that bias in the control conditions to estimates in the two treatment conditions. To calculate a measure of partisan bias, we created two indicator variables, *Match* and *Mismatch*. The *Match* variable takes the value of 1 if a respondent was presented with a target person who supported either their party or their party's presidential nominee (e.g., a Republican or Trump, respectively, for a Republican respondent) and 0 otherwise. The *Mismatch* variable takes the value of 1 if a respondent was presented with a target person who supported either the opposition party or that party's presidential nominee and 0 otherwise. This approach allows us to examine if measured bias is affected by in- versus out-party evaluations. The excluded category is profiles that do not include a partisan referent. Because analysis is restricted to partisan respondents, these indicators span all possible respondent-target mappings.

To measure baseline partisan bias in a way comparable to prior work, we focus on respondents in the control condition and regress the 7-point attractiveness item, scaled to vary -3 to 3 (higher values denote more attractiveness), on the *Match* and *Mismatch* variables. The difference between the coefficients for the two variable, i.e., *Match* minus *Mismatch*, is an estimate of total partisan bias. It corresponds to how much more attractive individuals report a target whose partisanship matches their own than a target whose partisanship is in opposition to their own, despite all other features of the target profile being held fixed.

Formally, this equation is

(1) $\text{Attractiveness}_i=\beta_0+\beta_1*\text{Match}_i+\beta_2*\text{Mismatch}_i+ \Omega_i + \varepsilon_i$

where the i subscript denotes individuals, Ω is a vector of controls (an indicator for whether a respondent is a Democrat or a Republican and indicators for each survey sample), and ε is an

idiosyncratic error term. This is an across-person analysis exploiting the individual-level randomization. For power reasons, we pool the three samples in our primary analysis.

To test whether estimated partisan bias is different in the "blow-off-steam" and "warning" treatment arms, we add indicators for assigned treatment and those indicators interacted with the Match and Mismatch indicators to equation (1). We first examine the pooled effect of the treatments. The variable *Either Treatment* takes on the value 1 in either the "Blow-Off-Steam" or "Warning" conditions and 0 otherwise. This is:

(2) $\text{Attractiveness}_i=\beta_0+\beta_1*\text{Match}_i+\beta_2*\text{Mismatch}_i+$

$\beta_3*\text{EitherTreatment}_i*\text{Match}_i+\beta_4*\text{EitherTreatment}_i*\text{Mismatch}_i +$

$\beta_5*\text{EitherTreatment}_i + \boldsymbol{\Omega}_i + \varepsilon_i.$

The estimate β_3 minus β_4 is how much smaller the estimate of partisan bias is in the treatment conditions relative to the control (baseline) condition. If the interventions work to reduce measured partisan bias, this estimate should be negative. (The remaining estimated partisan bias in this condition is ($\beta_1 + \beta_3$) minus ($\beta_2 + \beta_4$).) In additional analyses, we also separately estimate treatment effects for the "blow-off-steam" and "warning" conditions.

All equations are estimated using OLS regression.

Baseline Estimates of Partisan Bias

Table 2 presents baseline (control) condition estimates of partisan bias using the equation (1) specification. Column (1) pools across all respondents and shows that compared to a neutral (non-partisan) profile, respondents, on average, evaluated a partisan matched profile as no more attractive (B=.016, SE=.11, *ns*) but evaluated a mismatched profile as substantially less attractive (B -.563, SE=.122, p<.01). The difference in these coefficients, reported in the bottom rows of

the Table in the row labeled "Diff Match – Unmatch" is .579 ($p<.01$), meaning that matched profiles are evaluated about half a unit (on a 7-point scale) more attractive than unmatched profiles. In the aggregate, these results therefore replicate those reported in Nicholson et al. (2016) that partisanship affects assessments of attractiveness, although the partisan bias estimates in this study are smaller.

[Table 2 here]

Partitioning the data by respondent partisanship, however, makes clear that this effect is driven by Democratic respondents. For Democrats, shown in column (3), the estimated partisan bias is .873 ($p<.01$) units, while for Republicans (column 5) it is only .180 ($p=.27$). There are of course a variety of potential explanations for this difference in results among Republicans compared to the earlier published estimates, including differences in sample composition, a political environment empowering Republicans (Republicans controlled the presidency and both chambers of Congress in 2017) which could reduce the threat posed by outpartisans, and a polarizing incumbent Republican president (Trump). One possibility is that the type of Republicans who participate in studies on MTurk may not representative of all Republicans (is younger, more liberal, etc. Relatedly, see Clifford, Jewell, and Waggoner [2015]). But columns 7-18 in Table 2 show larger gaps for Democrats than Republicans in all three samples.

Analysis shown in in Appendix Table 2 tests whether these result are driven by the use of Trump and Clinton as party exemplars. Estimates of partisan bias are indistinguishable between party cues that did not list candidates or only those party cues that were candidate based. Additionally, analysis reported there shows that the exclusion of party leaners or focusing only on strong partisans (in the Lucid samples) slightly increases the bias estimates, but they remain insignificant for Republicans.

Importantly, this lack of clear baseline evidence for partisan bias among Republicans has several implications. First, it implies that the pattern observed among Republicans in prior work does not hold at all times and for all samples. Second, it means that reliably detecting treatment effect differences among Republicans, for whom the bias estimates are indistinguishable from statistical noise, is unlikely. Accordingly, all subsequent analyses are also presented broken down by party.

Additionally, it is useful to validate whether partisans perceive value differences that are correlated with partisanship. If partisans do not perceive (or report) large value differences depending on the partisanship of the profile they view, then the assumption that being given a chance to offer values assessments provides a desired opportunity to send a partisan message is unlikely to be supported. Pooling across all treatments, the estimated partisan difference in values assessments is 1.18 units ($p<.01$), which is larger than the effect of partisanship on attractiveness in the control condition (See Appendix Table 4). Our respondents therefore report partisan value differences and, unlike attractiveness ratings, these results are similar in magnitude across the parties and statistically significant for each party in all three samples.

Do Treatments Reduce Measured Partisan Bias?

Table 3 presents regression analysis showing the effects of our two treatments on estimates of partisan bias in attractiveness assessments. The first six columns present estimates using equation 2, which pools across interventions. Columns (7) through (18) repeat the column (1) specification by party and sample subgroups.

[Table 3 here]

We focus attention on the baseline estimates of partisan bias shown at the bottom of the table (See "Diff Match – Unmatch") and the reduction in estimated bias associated with the treatments. The results are clear: In the pooled sample shown in column (1), the baseline estimate of partisan bias is .57 ($p<.01$), which is reduced by .19 units ($p=.14$), or 34%, by either treatment.[9] Partitioning by partisanship, we find that the effect is larger and more precisely estimated for Democrats (Column 3): The baseline partisan bias estimate is .87 ($p<.01$), and is reduced by .41 units ($p=.02$) by the treatments, or 48%. For Republicans (Column 5) there is a null result for baseline partisan bias ($B=.18$, $p=.28$) and the treatment effect estimate is small, positive, and far from significant ($B=.08$, $p=.66$). In short, there is clear evidence that for the group for which there is a baseline partisan bias in attractiveness assessments (Democrats) that these treatments reduce measured partisan bias.

In columns (7) through (18) we repeat the analysis, by sample and party. For Republicans (columns 9, 13, and 17) baseline bias estimates are small and statistically insignificant and treatment effects are small, inconsistently signed, and imprecisely estimated. For Democrats, baseline bias is larger in the first two samples (columns 7 and 11) and we find statistically significant reductions in that bias, about 70 percent, associated with the treatments ($p<.05$). In the third sample, the baseline bias estimate is smaller and the treatments have a small insignificant effect that has the wrong sign. Overall, for the group for which we find bias (Democrats), in two of three samples the treatments substantially reduce estimated bias.

---

[9] This baseline partisan bias estimate differs from the estimate reported in Table 2 because Table 3 applies weights to account for different treatment rates across samples.

We also separately estimate treatment effects for the "blow-off-steam" and "warning" conditions (See Appendix Section C for full results.) The warning treatment reduced bias estimates by .48 units (p=.02, 55%), while the blow-off-steam treatment reduced it by .36 units (p=.08, 41%). The difference between the two treatments is insignificant (diff. = .12, n.s.) suggesting both treatments exhibit similar effects on reducing bias estimates. Future studies with much larger sample sizes would be necessary to identify meaningful differences in the effects of these two treatments.

By what mechanism do these treatments work?

We posit that these treatments reduce expressive responding because they allow individuals to express their partisan sentiments using questions other than the attractiveness item. But one might wonder whether the treatments operate not by allowing individuals to express (or anticipate expressing) their partisan sentiments, but instead by changing how individuals interpret the attractiveness item. Specifically, it is not clear that this question refers solely to *physical* attractiveness. The original study protocol, copied in this study's baseline condition, asked respondents to look at a picture of a person and read a short information section and then rate the attractiveness of that person. It is possible that in both the original study and this study's baseline conditions, individuals understood this question as also referring to other "dimensions" of attractiveness, including partisanship or other values, with the two treatments causing individuals to distinguish those considerations from physical attractiveness. In this case, the treatment would operate by changing the meaning of the attractiveness item rather than via an "answering the unasked question" mechanism.

Two primary responses to this concern are in order. First, the exact mechanism is less important than what this account would tell us about interpreting the partisan bias estimates in the baseline conditions and the original study. If the attractiveness item is originally interpreted as not merely about physical attractiveness, then it is incorrect to infer that individuals are engaging in biased assessments of beauty, which is how the prior result has been interpreted. In that case, ambiguous question wording concerns are reduced by emphasizing that physical attractiveness and values are distinct dimensions.

Second, a related paper (Mallinas, Crawford, and Cole 2018) testing the effect of ideology (rather than partisanship) on assessments of attractiveness found that including the word "physical" in the attractiveness question, compared to asking solely about attractiveness, did not alter the finding that individuals (particularly liberals) rated co-ideologues as more attractive than out-ideologues. It thus appears unlikely that the potential ambiguity of the attractiveness item explains the effect observed at baseline.[10]

Discussion

It is commonly believed that partisan differences in responses to survey items concerning both factual and non-factual questions are evidence of the biasing effect of partisanship (see also Ditto

---

[10] Mallinas et al. (2018) also show that while ideological dissimilarity affected respondents' self-reported evaluations of physical attractiveness, it did not affect respondents' evaluations of the attractiveness of the target person in comparison to a morphed version of that same face. They interpret this as showing reported attractiveness, but not actual attractiveness, is affected by ideology and suggest this is compatible with expressive responding (66).

et al. 2019). This study's results provide support for a different explanation for these measured partisan differences: Traditional survey data may overstate the apparent effect of partisanship because some individuals use those items to express partisan sentiments not actually asked in the underlying survey item. Ordinary survey items, despite their direct wording, are therefore used by some respondents as a way to convey a "partisan message"—to declare their support for their party and dislike for the rival party.

Replicating a recent experimental study that showed that rival partisans evaluate ingroup members as more physically attractive than outgroups members (Nicholson et al. 2016), this paper shows that allowing partisan respondents to express sentiment about the value superiority of their party vis-à-vis the opposition can reduce (by as much as 50 percent) the differences between partisans' evaluations of the attractiveness of in- versus out-partisans. This suggests that some portion of the originally estimated effects may have been due to expressive responding.

Importantly, the interventions tested—allowing individuals to express a partisan sentiment or provide a future opportunity to do so—are readily adapted to a variety of survey contexts and do not require financial incentives (which are costly, infeasible for non-factual items, and subject to certain theoretical critiques). They are also readily included in randomized designs, so that scholars can assess the degree of expressive responding that arises due to "unasked questions."

We conclude with four general observations. First, the treatments do not entirely eliminate measured partisan differences. Thus, one should not conclude on the basis of this work that measured partisan bias is purely expressive. However, the treatments are relatively weak: Individuals were given a chance to express a partisan sentiment. If that desire is strong, there is no reason that one might not want to continue sending a partisan message. As such, it is

premature to assume that the remaining bias reflects actual differences in partisan perception. In studies that have varied the relative value of the incentive for accuracy through financial incentives, larger incentives have reduced partisan bias (e.g., Bullock et al. 2015). Conversely, the current treatments did not allow for a similar variation in strength (Although asking respondents several "blow-off-steam" items might allow for such a variation).

Second, the techniques deployed here should be tested for other questions and in comparison to other approaches that have been tried to reduce measured partisan bias. Assessments of physical attractiveness do not evoke deep partisan feelings, so perhaps these treatments would be less effective in environments where the partisan stakes are higher (e.g., in assessments of responsibility for an economic downturn [Bisgaard 2015, 2019])? Would these treatments work in the context of partisan rumors? And finally, how large are the effects of these treatments vis-à-vis direct financial incentives. All of these questions await future work.

Third, we find substantial differences in baseline partisan bias estimates by partisan subgroup that depart from prior work on this topic. While there are many potential explanations for this pattern, as noted earlier, this study does not adjudicate among those accounts. This does not detract from the fact that the bias that is detected can generally be reduced by these interventions, yet it does call for future work on the origins of bias.

Finally, these results imply that scholars should be more cognizant of, and sensitive to the possibility of, expressive responding in partisans' survey responses (see also Bullock and Lenz 2019). The publication of the Bullock et al. (2015) and Prior et al. (2015) papers has arguably made many scholars aware of the possibility of expressive responding in surveys. Still, thus far only a small number of studies has tried to examine whether partisans' responses to certain political items in surveys are indeed due to expressive responding (for a few exceptions, see

Ahler and Sood 2018; Berinsky 2018; Yair and Sulitzeanu-kenan 2018). Indeed, Van Bavel and Pereira (2018, 218) contend that "the effects of partisanship on perception are important, but controversial, and warrant additional research." This paper contributes to the burgeoning expressive responding literature and may inspire more methods intended to detect expressive responding in public opinion surveys.

Appendix

Appendix Section A – two additional treatments

In the interest of transparency, the study also reports findings from another pair of treatments that did not perform as anticipated. Building on the intuition of Gal and Rucker (2011), the second Lucid sample also tested a unique pair of graphical interventions that yielded unanticipated null results. In particular, in these interventions the standard textual survey items was replaced with a graphical "heatmap" question in which respondents simultaneously answered two questions by clicking in a graphical box. (The profile they viewed was constructed in the same manner used throughout this study.) Their horizontal click indicated their answer to the attractiveness item (with the far left labeled extremely unattractive and the far right labeled extremely attractive) and their vertical click indicated their answer to one of two unrelated items: The person's height or their values.

Ex ante, the height manipulation was viewed as equivalent to a control condition because it was deemed that it would not allow meaningful expression of partisan sentiment, whereas the values manipulation was the core of the "blow-off-steam" treatment presented earlier. This treatment did not perform as anticipated: In neither arm partisan differences in attractiveness assessments were detected. However, for both height and values, substantial and statistically significant differences were detected by whether the profile matched the respondent's own partisanship: On average, partisans report the person has better values or is taller when the profile's partisanship matches their own vis-à-vis an opposing partisan. Ex post, this can be interpreted as meaning that both dimensions (height and values) were used by partisans to express a partisan sentiment, and it would be curious to test whether items that are even more

innocuous than height evaluations (e.g., was the person born on an odd or even day of the month?) would produce similar effects. It could also be the cases that whether the attractiveness item is placed on the vertical or horizontal axis is consequential. Alternatively, it could be that respondents were simply confused by the "heatmap" response format, an ostensibly unfamiliar format that requires respondents to answer two questions in one response. Overall, because these treatments did not produce a baseline partisan bias estimate in attractiveness evaluations in the height condition, further analysis of these conditions was abandoned.

Appendix Section B – wording of the "warning" treatment

Before presenting the target person and descriptive text, we presented respondents in the "warning" treatment arm the following text:

> The purpose of the following questions is to learn more about the kinds of characteristics that people find attractive, as well as about the kinds of characteristics that indicate someone has good values.
>
> On the next page, we will show you a person's picture and some information about them and ask you to answer TWO questions. First, we will ask you to rate their attractiveness. Second, we will ask you whether they have good values.

On the following page respondents were shown the target person's picture and the "About Me" information, and were asked, on the same page, first about the target person's attractiveness and then whether this person had good values.

Appendix Section C – separate analyses of the two treatments

In addition to analyzing the combined effect of our two interventions (Table 3 in the main text), we also separately estimate treatment effects by replacing the pooled *Either Treatment* indicator in equation (2) with indicators for the "blow-off-steam" (*Steam*) and "warning" (*Warn*) conditions and their interactions with the Match and Mismatch variables. This yields:

(1) $\text{Attractiveness}_i=\beta_0+\beta_1*\text{Match}_i+\beta_2*\text{Mismatch}_i+$

$\beta_3*\text{Steam}_i*\text{Match}_i+\beta_4*\text{Steam}_i*\text{Mismatch}_i +$

$\beta_5*\text{Warn}_i*\text{Match}_i+\beta_6*\text{Warn}_i*\text{Mismatch}_i +$

$\beta_7*\text{Steam}_i + \beta_8*\text{Warn}_i + \Omega_i + \varepsilon_i.$

The estimate β_3 minus β_4 is how much smaller the estimate of partisan bias is in the "blow-off-steam" condition relative to control and β_5 minus β_6 is the same calculation for the "warning" condition.

In Appendix Table 3 we present the results of these analyses. Democrats (Column 2) are the focus here because this is the group for which there is baseline evidence of bias. Per these results, the steam treatment reduces measured bias by .36 units (p=.08, 41%) while the warning condition does so by .48 units (p=.02, 55%). While the former estimate is marginally significant and the latter is significant, these estimates are indistinguishable from one another (diff. = .12, n.s.). Both techniques appear to have about the same effect on reducing measured partisan bias and much larger samples would be necessary to determine if the differences across treatment are meaningful, a task for subsequent work.

Supplementary Data

Supplementary data are freely available at *Public Opinion Quarterly* online.

References

Ahler, Douglas J., and Gaurav Sood. 2018. "The Parties in Our Heads: Misperceptions About Party Composition and Their Consequences." *The Journal of Politics* 80:964–81.

Bartels, Larry. 2002. "Beyond the Running Tally: Partisan Bias in Political Perceptions." *Political Behavior* 24:117–50.

Bavel, Jay J. Van, and Andrea Pereira. 2018. "The Partisan Brain: An Identity-Based Model of Political Belief." *Trends in Cognitive Sciences* 22:213–24.

Berinsky, Adam J. 2018. "Telling the Truth about Believing the Lies? Evidence for the Limited Prevalence of Expressive Survey Responding." *The Journal of Politics* 80:211–24.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.Com's Mechanical Turk." *Political Analysis* 20:351–68.

Bisgaard, Martin. 2015. "Bias Will Find a Way: Economic Perceptions, Attributions of Blame, and Partisan-Motivated Reasoning during Crisis." *The Journal of Politics* 77:849–60.

———. 2019. "How Getting the Facts Right Can Fuel Partisan-Motivated Reasoning." *American Journal of Political Science* 63:824-39.

Bullock, John G., Alan S. Gerber, Seth J. Hill, and Gregory A. Huber. 2015. "Partisan Bias in Factual Beliefs about Politics." *Quarterly Journal of Political Science* 10:519–78.

Bullock, John G., and Gabriel Lenz. 2019. "Partisan Bias in Surveys." *Annual Review of Political Science* 22:325–42.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald Stokes. 1960. *The American Voter*. New York: Wiley.

Carsey, Thomas M., and Geoffrey C. Layman. 2006. "Changing Sides or Changing Minds? Party Identification in the American Electorate." *American Journal of Political Science* 50:464–77.

Chen, Philip G., and Paul N. Goren. 2016. "Operational Ideology and Party Identification: A Dynamic Model of Individual-Level Change in Partisan and Ideological Predispositions." *Political Research Quarterly* 69:703–15.

Clifford, Scott, and Jennifer Jerit. 2015. "Do Attempts to Improve Respondent Attention Increase Social Desirability Bias?" *Public Opinion Quarterly* 79:790–802.

Clifford, Scott, Ryan M. Jewell, and Philip D. Waggoner. 2015. "Are Samples Drawn from Mechanical Turk Valid for Research on Political Ideology?" *Research & Politics* 2:1–9.

Cohen, Geoffrey L., Joshua Aronson, and Claude M. Steele. 2000. "When Beliefs Yield to Evidence: Reducing Biased Evaluation by Affirming the Self." *Personality and Social Psychology Bulletin* 26:1151–64.

Cohen, Geoffrey L., and David K. Sherman. 2014. "The Psychology of Change: Self-Affirmation and Social Psychological Intervention." *Annual Review of Psychology* 65:333–71.

Cohen, Geoffrey L., David K. Sherman, Anthony Bastardi, Lillian Hsu, Michelle McGoey, and Lee Ross. 2007. "Bridging the Partisan Divide: Self-Affirmation Reduces Ideological Closed-Mindedness and Inflexibility in Negotiation." *Journal of Personality and Social Psychology* 93:415–30.

Coppock, Alexander, and Oliver A. McClellan. 2019. "Validating the Demographic, Political,

Psychological, and Experimental Results Obtained from a New Source of Online Survey Respondents." *Research & Politics*. https://doi.org/10.1177/2053168018822174.

Ditto, Peter H., Brittany S. Liu, Cory J. Clark, Sean P. Wojcik, Eric E. Chen, Rebecca H. Grady, Jared B. Celniker, and Joanne F. Zinger. 2019. "At Least Bias Is Bipartisan: A Meta-Analytic Comparison of Partisan Bias in Liberals and Conservatives." *Perspectives on Psychological Science* 14:273-91.

Flynn, D. J., Jason Reifler, and Brendan Nyhan. 2017. "The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs about Politics." *Political Psychology* 38(Suppl. 1):127–50.

Gaines, Brian J., James H. Kuklinski, Paul J. Quirk, Buddy Peyton, and Jay Verkuilen. 2007. "Same Facts, Different Interpretations: Partisan Motivation and Opinion on Iraq." *The Journal of Politics* 69:957–74.

Gal, David, and Derek D Rucker. 2011. "Answering the Unasked Question: Response Substitution in Consumer Surveys." *Journal of Marketing Research* 48:185–95.

Huber, Gregory A., and Neil Malhotra. 2017. "Political Homophily in Social Relationships: Evidence from Online Dating Behavior." *The Journal of Politics* 79:269–83.

Huddy, Leonie, Alexa Bankert, and Caitlin L. Davies. 2018. "Expressive Versus Instrumental Partisanship in Multi-Party European Systems." *Political Psychology* 39(Suppl. 1):173–99.

Huddy, Leonie, Lilliana Mason, and Lene Aarøe. 2015. "Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity." *American Political Science Review* 109:1–17.

Jerit, Jennifer, and Jason Barabas. 2012. "Partisan Perceptual Bias and the Information Environment." *The Journal of Politics* 74:672–84.

Kahan, Dan M. 2016. “The Politically Motivated Reasoning Paradigm.” *Emerging Trends in Social & Behavioral Sciences*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2703011.

Khanna, Kabir, and Gaurav Sood. 2018. “Motivated Responding in Studies of Factual Learning.” *Political Behavior* 40:79–101.

Krupnikov, Yanna, Spencer Piston, and Nichole M. Bauer. 2016. “Saving Face: Identifying Voter Responses to Black Candidates and Female Candidates.” *Political Psychology* 37:253–73.

Kuhn, Patrick M., and Nick Vivyan. 2018. “Reducing Turnout Misreporting in Online Surveys.” *Public Opinion Quarterly* 82:300–21.

Leeper, Thomas J., and Rune Slothuus. 2014. “Political Parties, Motivated Reasoning, and Public Opinion Formation.” *Political Psychology* 35(Suppl. 1):129–56.

Lodge, Milton, and Charles S. Taber. 2013. *The Rationalizing Voter*. New York: Cambridge University Press.

Mallinas, Stephanie R., Jarret T. Crawford, and Shana Cole. 2018. “Political Opposites Do Not Attract: The Effects of Ideological Dissimilarity on Impression Formation.” *Journal of Social and Political Psychology* 6:49–75.

Massey, Cade, Joseph P. Simmons, and David A. Armor. 2011. “Hope over Experience: Desirability and the Persistence of Optimism.” *Psychological Science* 22:274–81.

McGrath, Mary C. 2017. “Economic Behavior and the Partisan Perceptual Screen.” *Quarterly Journal of Political Science* 11:363–83.

Nicholson, Stephen P., Chelsea M. Coe, Jason Emory, and Anna V. Song. 2016. “The Politics of Beauty: The Effects of Partisan Bias on Physical Attractiveness.” *Political Behavior*

38:883–98.

Nyhan, Brendan, and Jason Reifler. 2010. “When Corrections Fail: The Persistence of Political Misperceptions.” *Political Behavior* 32:303–30.

Pew Research Center. 2017. “The Partisan Divide on Political Values Grows Even Wider.” 2017. http://www.people-press.org/2017/10/05/the-partisan-divide-on-political-values-grows-even-wider/.

Prior, Markus, Gaurav Sood, and Kabir Khanna. 2015. “You Cannot Be Serious: The Impact of Accuracy Incentives on Partisan Bias in Reports of Economic Perceptions.” *Quarterly Journal of Political Science* 10:489–518.

Schaffner, Brian F., and Samantha Luks. 2018. “Misinformation or Expressive Responding? What an Inauguration Crowd Can Tell Us About the Source of Political Misinformation in Surveys.” *Public Opinion Quarterly* 82:135–47.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge University Press.

Yair, Omer, and Raanan Sulitzeanu-kenan. 2018. “When Do We Care about Political Neutrality? The Hypocritical Nature of Reaction to Political Bias.” *PLoS ONE* 13:e0196674.

Table 1: Summary of experimental conditions

ARM	**Control**	**Democrat**	**Republican**	**Supported Clinton in 2016**	**Supported Trump in 2016**
Baseline (replication)	Respondents were shown a profile picture of a target person of the opposite sex and the following information in the 'About me' section: "Friendly", "Smart", and "Runner". They were then asked to answer a 7-point attractiveness item.	Adds to control condition the word "Democrat" at the bottom of the 'About me" section.	Adds to control condition the word "Republican" at the bottom of the 'About me" section.	Adds to control condition the words "Supported Clinton in the 2016 election" at the bottom of the 'About me" section.	Adds to control condition the words "Supported Trump in the 2016 election" at the bottom of the 'About me" section.
"Blow off steam" intervention	Identical to baseline cells above, except that Respondents first answered a question about the values of the person shown to them.				
"Warning" intervention	Identical to baseline cells above, except that Respondents were told that on the next page they will evaluate a person's attractiveness as well as that person's values.				

<<NB: Table 2 is broken into 2 parts to get it to fit>>

Table 2: Effect of Partisan Match on Attractiveness, Control Condition Observations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Attractiveness Pooled (-3 to 3)									
Sample:	Pooled		Democrats		Republicans		Dems. Sample 1: Mturk		Reps. Sample 1: Mturk	
VARIABLES	coef	se	coef	se	coef	se	coef	se	coef	se
Person is Democrat (1) or Republican (0)	-0.182+	(0.094)								
Respondent Partisanship Matches Profile (1=yes, 0=No)	0.016	(0.110)	0.159	(0.139)	-0.166	(0.175)	0.275	(0.224)	-0.327	(0.400)
Respondent Partisanship Mismatches Profile (1=yes, 0=No)	-0.563**	(0.122)	-0.714**	(0.163)	-0.346+	(0.183)	-0.808**	(0.275)	-0.441	(0.381)
Sample = Lucid #2	-0.220+	(0.114)	-0.128	(0.161)	-0.301+	(0.160)				
Sample = MTurk	0.127	(0.120)	0.150	(0.152)	0.056	(0.196)				
Constant	1.591**	(0.125)	1.372**	(0.146)	1.631**	(0.178)	1.500**	(0.200)	1.786**	(0.296)
Observations	814		461		353		149		67	
R-squared	0.060		0.101		0.024		0.173		0.018	
Diff Match - Unmatch	0.579		0.873		0.180		1.082		0.114	
p-value	0.000		0.000		0.272		0.000		0.754	

Unweighted analysis. Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1 (two-tailed test)

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	Attractiveness Pooled (-3 to 3)							
Sample:	Dems. Sample 2: Lucid 1		Reps. Sample 2: Lucid 1		Dems. Sample 3: Lucid 2		Reps. Sample 3: Lucid 2	
VARIABLES	coef	se	coef	se	coef	se	coef	se
Person is Democrat (1) or Republican (0)								
Respondent Partisanship Matches Profile (1=yes, 0=No)	-0.157	(0.212)	-0.168	(0.343)	0.247	(0.253)	-0.110	(0.223)
Respondent Partisanship Mismatches Profile (1=yes, 0=No)	-1.184**	(0.286)	-0.247	(0.326)	-0.343	(0.268)	-0.400	(0.268)
Sample = Lucid #2								
Sample = MTurk								
Constant	1.692**	(0.153)	1.594**	(0.272)	1.050**	(0.205)	1.325**	(0.173)
Observations	126		128		186		158	
R-squared	0.135		0.005		0.035		0.017	
Diff Match - Unmatch	1.026		0.079		0.590		0.291	
p-value	0.000		0.778		0.011		0.245	

Unweighted analysis. Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1 (two-tailed test)

<<NB: Table 3 is broken into 2 parts to get it to fit>>

Table 3: Effect of Treatments on Estimated Partisan Bias in Attractiveness Evaluations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Attractiveness Pooled (-3 to 3)									
Sample:	Pooled		Democrats		Republicans		Dems. Sample 1: Mturk		Reps. Sample 1: Mturk	
VARIABLES	coef	se	coef	se	coef	se	coef	se	coef	se
Person is Democrat (1) or Republican (0)	-0.196**	(0.057)								
Respondent Partisanship Matches Profile (1=yes, 0=No)	0.014	(0.112)	0.157	(0.141)	-0.161	(0.176)	0.275	(0.224)	-0.327	(0.399)
Respondent Partisanship Mismatches Profile (1=yes, 0=No)	-0.553**	(0.124)	-0.709**	(0.166)	-0.341+	(0.184)	-0.808**	(0.275)	-0.441	(0.380)
Match X Either	-0.056	(0.135)	-0.196	(0.174)	0.105	(0.209)	-0.219	(0.298)	0.023	(0.479)
Mismatch X Either	0.137	(0.151)	0.215	(0.205)	0.021	(0.219)	0.524	(0.360)	-0.067	(0.457)
Either treatment	0.070	(0.109)	0.071	(0.144)	0.071	(0.166)	-0.140	(0.261)	0.155	(0.356)
Sample = Lucid #2	-0.198**	(0.068)	-0.162+	(0.096)	-0.228*	(0.095)				
Sample = MTurk	0.073	(0.074)	0.049	(0.095)	0.104	(0.117)				
Constant	1.598**	(0.104)	1.413**	(0.127)	1.586**	(0.156)	1.500**	(0.200)	1.786**	(0.295)
Observations	2,274		1,295		979		355		147	
R-squared	0.049		0.069		0.025		0.083		0.032	
Diff Match - Unmatch	0.567		0.866		0.179		1.082		0.114	
p-value	0.000		0.000		0.273		0.000		0.753	
Reduction Either	-0.193		-0.411		0.084		-0.743		0.090	
p-value Either	0.139		0.018		0.665		0.001		0.835	

Unweighted analysis. Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1 (two-tailed test)

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	Attractiveness Pooled (-3 to 3)							
Sample:	Dems. Sample 2: Lucid 1		Reps. Sample 2: Lucid 1		Dems. Sample 3: Lucid 2		Reps. Sample 3: Lucid 2	
VARIABLES	coef	se	coef	se	coef	se	coef	se
Person is Democrat (1) or Republican (0)								
Respondent Partisanship Matches Profile (1=yes, 0=No)	-0.157	(0.211)	-0.168	(0.342)	0.247	(0.252)	-0.110	(0.223)
Respondent Partisanship Mismatches Profile (1=yes, 0=No)	-1.184**	(0.285)	-0.247	(0.325)	-0.343	(0.267)	-0.400	(0.267)
Match X Either	-0.308	(0.280)	0.197	(0.384)	-0.127	(0.292)	0.058	(0.281)
Mismatch X Either	0.494	(0.352)	-0.202	(0.376)	-0.214	(0.321)	0.284	(0.330)
Either treatment	0.093	(0.209)	0.045	(0.306)	0.234	(0.237)	0.063	(0.222)
Sample = Lucid #2								
Sample = MTurk								
Constant	1.692**	(0.153)	1.594**	(0.270)	1.050**	(0.204)	1.325**	(0.172)
Observations	378		388		562		444	
R-squared	0.076		0.026		0.053		0.012	
Diff Match - Unmatch	1.026		0.079		0.590		0.291	
p-value	0.000		0.776		0.001		0.242	
Reduction Either	-0.802		0.399		0.087		-0.226	
p-value Either	0.019		0.210		0.752		0.448	

Unweighted analysis. Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1 (two-tailed test)

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	Attractiveness Pooled (-3 to 3)							
Sample:	Dems. Sample 2: Lucid 1		Reps. Sample 2: Lucid 1		Dems. Sample 3: Lucid 2		Reps. Sample 3: Lucid 2	
VARIABLES	coef	se	coef	se	coef	se	coef	se
Person is Democrat (1) or Republican (0)								
Respondent Partisanship Matches Profile (1=yes, 0=No)	-0.157	(0.212)	-0.168	(0.343)	0.247	(0.253)	-0.110	(0.223)
Respondent Partisanship Mismatches Profile (1=yes, 0=No)	-1.184**	(0.286)	-0.247	(0.326)	-0.343	(0.268)	-0.400	(0.268)
Sample = Lucid #2								
Sample = MTurk								
Constant	1.692**	(0.153)	1.594**	(0.272)	1.050**	(0.205)	1.325**	(0.173)
Observations	126		128		186		158	
R-squared	0.135		0.005		0.035		0.017	
Diff Match - Unmatch	1.026		0.0786		0.590		0.291	
p-value	.000		0.778		0.0107		0.245	

Unweighted analysis. Robust standard errors in parentheses.

** p<0.01, * p<0.05, + p<0.1

<<NB: Table 3 is broken into 2 parts to get it to fit>>

Table 3: Effect of Treatments on Estimated Partisan Bias in Attractiveness Evaluations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Attractiveness Pooled (-3 to 3)									
Sample:	Pooled		Democrats		Republicans		Dems. Sample 1: Mturk		Reps. Sample 1: Mturk	
VARIABLES	coef	se	coef	se	coef	se	coef	se	coef	se
Person is Democrat (1) or Republican (0)	-0.196**	(0.057)								
Respondent Partisanship Matches Profile (1=yes, 0=No)	0.014	(0.112)	0.157	(0.141)	-0.161	(0.176)	0.275	(0.224)	-0.327	(0.399)
Respondent Partisanship Mismatches Profile (1=yes, 0=No)	-0.553**	(0.124)	-0.709**	(0.166)	-0.341+	(0.184)	-0.808**	(0.275)	-0.441	(0.380)
Match X Either	-0.056	(0.135)	-0.196	(0.174)	0.105	(0.209)	-0.219	(0.298)	0.023	(0.479)
Mismatch X Either	0.137	(0.151)	0.215	(0.205)	0.021	(0.219)	0.524	(0.360)	-0.067	(0.457)
Either treatment	0.070	(0.109)	0.071	(0.144)	0.071	(0.166)	-0.140	(0.261)	0.155	(0.356)
Sample = Lucid #2	-0.198**	(0.068)	-0.162+	(0.096)	-0.228*	(0.095)				
Sample = MTurk	0.073	(0.074)	0.049	(0.095)	0.104	(0.117)				
Constant	1.598**	(0.104)	1.413**	(0.127)	1.586**	(0.156)	1.500**	(0.200)	1.786**	(0.295)
Observations	2,274		1,295		979		355		147	
R-squared	0.049		0.069		0.025		0.083		0.032	
Diff Match - Unmatch	0.567		0.866		0.179		1.082		0.114	
p-value	1.54e-07		1.22e-09		0.273		7.13e-07		0.753	
Reduction Either	-0.193		-0.411		0.0835		-0.743		0.0895	
p-value Either	0.139		0.0180		0.665		0.00992		0.835	

Unweighted analysis.
Robust standard errors in parentheses.
** p<0.01, * p<0.05, + p<0.1

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	Attractiveness Pooled (-3 to 3)							
Sample:	Dems. Sample 2: Lucid 1		Reps. Sample 2: Lucid 1		Dems. Sample 3: Lucid 2		Reps. Sample 3: Lucid 2	
VARIABLES	coef	se	coef	se	coef	se	coef	se
Person is Democrat (1) or Republican (0)								
Respondent Partisanship Matches Profile (1=yes, 0=No)	-0.157	(0.211)	-0.168	(0.342)	0.247	(0.252)	-0.110	(0.223)
Respondent Partisanship Mismatches Profile (1=yes, 0=No)	-1.184**	(0.285)	-0.247	(0.325)	-0.343	(0.267)	-0.400	(0.267)
Match X Either	-0.308	(0.280)	0.197	(0.384)	-0.127	(0.292)	0.058	(0.281)
Mismatch X Either	0.494	(0.352)	-0.202	(0.376)	-0.214	(0.321)	0.284	(0.330)
Either treatment	0.093	(0.209)	0.045	(0.306)	0.234	(0.237)	0.063	(0.222)
Sample = Lucid #2								
Sample = MTurk								
Constant	1.692**	(0.153)	1.594**	(0.270)	1.050**	(0.204)	1.325**	(0.172)
Observations	378		388		562		444	
R-squared	0.076		0.026		0.053		0.012	
Diff Match - Unmatch	1.026		0.0786		0.590		0.291	
p-value	0.000296		0.776		0.00995		0.242	
Reduction Either	-0.802		0.399		0.0868		-0.226	
p-value Either	0.0186		0.210		0.752		0.448	

Unweighted analysis. Robust standard errors in parentheses.

Appendix Table 1: Summary Statistics and Summary of Randomization Tests

Variable	Sample 1: Mturk	Sample 2: Lucid 1	Sample 3: Lucid 2
Person is Democrat (1) or Republican (0)	0.707	0.494	0.559
	[.4555]	[.5003]	[.4968]
Race white (1=yes)	0.799	0.845	0.749
	[.4013]	[.3625]	[.4341]
Race black (1=yes)	0.054	0.094	0.127
	[.2258]	[.292]	[.3334]
Race latino (1=yes)	0.068	0.068	0.098
	[.2515]	[.2517]	[.298]
Female (1=yes, 0=male)	0.470	0.551	0.556
	[.4996]	[.4977]	[.4971]
Education (0=<HS, 3=grad school)	1.702	1.384	1.324
	[.8115]	[.9852]	[.9528]
Age in years	35.394	48.971	45.244
	[11.4836]	[16.0989]	[16.7669]
Ideology (0=V. Consv; 4=V. Lib)	2.552	2.103	2.024
	[1.2479]	[1.3473]	[1.1519]
Observations	502	766	1006
Chi2 Randomization Test	18.682	15.289	11.047
Chi2 p-value	0.286	0.504	0.807

Unweighted summary statistics. Standard deviations in brackets. Chi2 randomization test is from a weighted multinomial logit predicting treatment assignment with all covariates listed in table. Insignificant Chi2 p-value indicates these demographics do not explain assigned treatment.

Appendix Table 2: Robustness of Table 2 Results

	(1)	(2)	(3)	(4)	(5)	(6)
	Excluding Clinton/Trump Profiles	Excluding Dem/Rep Profiles	Democrats No Leaners	Republicans No Leaners	Strong Democrats	Strong Republicans
Person is Democrat (1) or Republican (0)	-0.130	-0.236				
	[0.111]	[0.123]*				
Respondent Partisanship Matches Profile (1=yes, 0=No)	0.167	-0.136	0.231	-0.093	0.223	0.038
	[0.119]	[0.136]	[0.148]	[0.197]	[0.250]	[0.283]
Respondent Partisanship Mismatches Profile (1=yes, 0=No)	-0.351	-0.737	-0.681	-0.356	-0.744	-0.169
	[0.148]**	[0.147]***	[0.173]***	[0.204]*	[0.304]**	[0.289]
Sample = Lucid #2	-0.279	-0.265	-0.065	-0.327	-0.119	-0.378
	[0.134]**	[0.147]*	[0.175]	[0.185]*	[0.237]	[0.228]*
Sample = MTurk	-0.045	0.226	0.143	0.049		
	[0.143]	[0.158]	[0.159]	[0.211]		
Constant	1.630	1.618	1.334	1.616	1.466	1.607
	[0.142]***	[0.143]***	[0.161]***	[0.206]***	[0.235]***	[0.264]***
Observations	481	511	399	290	162	155
R-squared	0.042	0.081	0.109	0.027	0.092	0.022
Diff Match - Unmatch	0.518	0.601	0.912	0.262	0.967	0.207
p-value	0.000	0.000	0.000	0.150	0.000	0.442

Unweighted analysis. Robust standard errors in brackets.

* significant at 10%; ** significant at 5%; *** significant at 1%

Appendix Table 3: Effect of Individual Treatments on Estimated Partisan Bias in Attractiveness Evaluations

	(1) Attractiveness, Pooled (-3 to 3)	(2) Democrats	(3) Republicans
Person is Democrat (1) or Republican (0)	-0.200		
	[0.053]***		
Respondent Partisanship Matches Profile (1=yes, 0=No)	0.015	0.160	-0.162
	[0.112]	[0.141]	[0.177]
Respondent Partisanship Mismatches Profile (1=yes, 0=No)	-0.552	-0.709	-0.340
	[0.124]***	[0.166]***	[0.184]*
Match X Blow off Steam	0.020	-0.114	0.185
	[0.155]	[0.207]	[0.233]
Mismatch X Blow off Steam	0.130	0.243	-0.015
	[0.176]	[0.244]	[0.249]
Match X Warning	-0.134	-0.273	0.037
	[0.155]	[0.198]	[0.242]
Mismatch X Warning	0.141	0.201	0.052
	[0.171]	[0.229]	[0.255]
Treatment Blow Off Steam	-0.001	-0.042	0.047
	[0.125]	[0.173]	[0.182]
Treatment Warning	0.142	0.177	0.095
	[0.126]	[0.162]	[0.198]
Sample = Lucid #2	-0.189	-0.174	-0.200
	[0.061]***	[0.088]**	[0.085]**
Sample = MTurk	0.058	0.014	0.127
	[0.070]	[0.093]	[0.104]
Constant	1.599	1.427	1.569
	[0.102]***	[0.127]***	[0.152]***
Observations	2274	1295	979
R-squared	0.046	0.061	0.026
Diff Match - Unmatch	0.568	0.869	0.178
p-value	0.000	0.000	0.276
Reduction Steam	-0.110	-0.357	0.200
p-value Steam	0.471	0.082	0.375
Reduction Warn	-0.275	-0.474	-0.016
p-value Warn	0.060	0.016	0.941

Weighted analysis. Robust standard errors in brackets.

* significant at 10%; ** significant at 5%; *** significant at 1%

Appendix Table 4: Effect of Partisan Match on Values Assessments

	(1) Good Values, Pooled (-3 to 3)	(2) Democrats	(3) Republicans	(4) Dems. Sample 1: Mturk	(5) Reps. Sample 1: Mturk	(6) Dems. Sample 2: Lucid 1	(7) Reps. Sample 2: Lucid 1	(8) Dems. Sample 3: Lucid 2	(9) Reps. Sample 3: Lucid 2
Person is Democrat (1) or Republican (0)	-0.199								
	[0.058]***								
Respondent Partisanship Matches Profile (1=yes, 0=No)	0.247	0.258	0.216	0.275	0.241	0.001	0.263	0.427	0.172
	[0.064]***	[0.083]***	[0.099]**	[0.132]**	[0.217]	[0.165]	[0.177]	[0.130]***	[0.140]
Respondent Partisanship Mismatches Profile (1=yes, 0=No)	-0.930	-1.081	-0.743	-1.730	-1.149	-1.126	-0.613	-0.651	-0.715
	[0.076]***	[0.103]***	[0.112]***	[0.183]***	[0.230]***	[0.199]***	[0.193]***	[0.153]***	[0.166]***
Sample = Lucid #2	-0.167	-0.027	-0.313						
	[0.067]**	[0.096]	[0.095]***						
Sample = MTurk	-0.013	0.043	-0.052						
	[0.078]	[0.102]	[0.123]						
Constant	1.628	1.409	1.641	1.684	1.742	1.529	1.570	1.140	1.336
	[0.074]***	[0.095]***	[0.100]***	[0.111]***	[0.146]***	[0.136]***	[0.145]***	[0.106]***	[0.110]***
Observations	2273	1294	979	355	147	378	388	561	444
R-squared	0.142	0.170	0.104	0.361	0.225	0.125	0.074	0.114	0.085
Diff Match - Unmatch	1.177	1.339	0.958	2.005	1.389	1.126	0.876	1.079	0.886
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Unweighted pooled analysis. Robust standard errors in brackets.

* significant at 10%; ** significant at 5%; *** significant at 1%