The Perils of Self-Assessed Attitude Change

Matthew Graham and Alexander Coppock*

August 20, 2018

Prepared for presentation at the Annual Meeting of the American Political Science Association, Boston, August 30 - September 2, 2018.

Abstract

Surveys often ask respondents to self-report how events or information changed their attitudes. Does [event X] make you more or less likely to vote for a politician?; How did [information X] affect your opinion?; Would you be more supportive of a candidate who took [position X]? We show that the self-assessment question type exhibits poor measurement properties. Using eight mini-experiments, we compare this question type—and two other ways of asking subjects to assess counterfactuals—to randomized experiments. When asked to report how their attitudes change, subjects appear to frequently overestimate the magnitude of treatment effects and sometimes get the sign wrong. Our first set of alternative formats, which "anchors" respondents by eliciting their absolute level of support for a candidate or policy, reduces these tendencies. Our second set of alternative formats, which asks respondents to report both of their potential outcomes, approximates experimental benchmarks surprisingly well.

In advance of Alabama's 2017 special election for U.S. Senator, the polling firm JMC Analytics released a survey which sought to estimate the effect of sexual misconduct allegations on support for Roy Moore's candidacy. The question read, "Given the allegations that have come out about Roy Moore's alleged sexual misconduct against four underage women, are you more or less likely to support him as a result of these allegations?" Among the 575 registered Alabama voters sampled, 29 percent responded "more likely," 38 percent responded

---

*Matthew Grahem is a Doctoral Candidate in Political Science, Yale University. Alexander Coppock is Assistant Professor of Political Science, Yale University.

"less likely," and 33 percent responded "no difference." The poll was widely discussed in the media, with many commentators decrying the apparent depravity of Alabama voters who, according to this poll, *increased* their support for a candidate when allegations of sexual assault against minors surfaced.

Interpreting the poll's results as the causal effect of the Moore scandal requires at least two leaps of faith. First, one must believe that subjects are able to estimate causal effects by introspection alone. That is, they must be able to gauge how much they would support Moore in both the presence and absence of the scandal, then take the difference. Though people draw accurate causal inferences about their own lives all the time (I broke my leg *because* I slipped), some prior research gives us reason to be skeptical of peoples' ability to report the sign of attitude change. Lord et al. (1979)'s classic study found that information about the death penalty produced self-reported attitude polarization: according to their self-assessments, supporters become more supportive and opponents become more opposed. Replications produce the same pattern of self-reports, but otherwise-identical experiments find no attitude polarization (Miller et al. 1993; Guess and Coppock 2018). At least this case, self-reports and experiments yield different conclusions.

Second, in order to interpret these survey responses as causal effects, one must believe subjects answered the question as asked rather some subtly different question. Gal and Rucker (2011) note that survey respondents may answer "unasked questions" as a way of expressing other attitudes about the object of the question. In the case of the Moore poll, one might reasonably imagine that subjects who nonetheless supported Moore—*despite* and not *because of* the allegations—chose the positive response option as an expression of support.

We consider a range of question formats designed to make these two leaps of faith more plausible. The main idea is to help subjects assess the direction of attitude change by anchoring them to one or both of their potential outcomes. The Moore poll used what we call the *direction (no anchor)* format, which asks subjects to state the sign of their treatment effect (more likely, less likely, no difference). The alternatives fit in two groups. *Direction with anchor* formats pair the more/less likely question with questions about the treatment and/or control outcome. *Both PO* formats ask respondents to name both their treatment and control potential outcomes (PO).

We test these formats in eight mini-experiments and report three sets of results. First, we show that asking people to consider their potential outcomes reduces self-reported opinion change. Second, we compare self-reported treatment effects from the *both PO* formats to randomized experiments. The *both PO* formats yield surprisingly accurate estimates.

Third, we show that compared with experimental benchmarks, the *direction* formats yield misleading substantive conclusions in more than half of our mini-experiments.

Survey questions for assessing treatment effects

Our goal is to understand which, if any, question formats enable respondents to accurately estimate the causal effect of information on their attitudes.

As part of the process of developing alternatives to the *direction (no anchor)* format, we consulted an assortment of survey design textbooks (Babbie 2011; Fowler 1995, 2014; Groves et al. 2009; Torangeau and Rips 2000; Sudman and Bradburn 1982). All valuable works, none contained specific guidance for questions that ask subjects to assess causal effects. Consequently, the alternatives we explore are based more on intuition and curiosity than on previous theoretical work.

Our point of departure is the Neyman-Rubin causal model. Under the model, subjects are endowed with a set of potential outcomes. Treatment exposure determines which is revealed. Focusing on the case of a single treatment, subject i has two potential outcomes: untreated and treated, which we respectively denote as $Y_i(0)$ and $Y_i(1)$. Due to the Fundamental Problem of Causal Inference (Holland 1986), we can observe $Y_i(0)$ or $Y_i(1)$ but not both. Accordingly, subject i's individual-level causal effect ($\tau_i \equiv Y_i(1) - Y_i(0)$) is unobservable as well.

The Fundamental Problem of Causal Inference bedevils not only social scientists, but also survey respondents. They *do not know* what effect a piece of information had or would have on their opinions because they cannot observe their own attitudes both in the presence and absence of the treatment. But survey respondents certainly know themselves better than do survey researchers. Individuals have access to something like an interrupted time series of their own opinions: what they remember thinking of Roy Moore before and after the allegations.

Questions that assess direction of change

The *direction (no anchor)* question type asks people to report $\text{sign}(\tau_i)$: did the scandal have a positive effect on support, a negative effect on support, or make no difference? This question type is vulnerable to many sources of bias. One is imperfect introspection. Because subjects cannot know τ_i with certainty, they cannot know its sign.

Our alternative formats are designed to reduce another possible form of measurement error: respondents may use the *direction (no anchor)* question to state their absolute level of support rather than the change in support due to the treatment information. Gal and Rucker (2011) and Yair and Huber (2018) present evidence that giving respondents the opportunity to answer "unasked questions" reduces measurement error in the main question. By a similar logic, we hypothesize that giving respondents the opportunity to state $Y_i(0)$, $Y_i(1)$, or both allows them to express this attitude and, at the same time, makes it clearer that the change question is different from the level question.

Accordingly, we propose four alternatives to the *direction (no anchor)* format, each of which asking subjects to report one or both of their potential outcomes before reporting the direction of change. We label these *anchored direction* formats:

- *Direction (Y0 first).* Subjects are asked for their control outcome, then are exposed to the treatment and are asked for the direction of change ($\text{sign}(\tau_i)$).

- *Direction (Y0 anchor).* Subjects are exposed to the treatment and then are asked to report what their Y0 *would have been.* Subjects then report the direction of change.

- *Direction (Y1 anchor).* Subjects are exposed to the treatment and then are asked to report what their Y1. Subjects then report the direction of change.

- *Direction (Y0 & Y1 anchors).* Subjects are exposed to the treatment and then are asked to report both their Y0 and Y1. Subjects then report the direction of change.

Our main expectation is that, relative to *direction (no anchor)*, each anchored format will reduce the amount subjects say the treatment changed their opinions. We did not begin our research with strong expectations as to the four variations' relative effect.

Questions that assess both potential outcomes

The sign of τ_i is interesting, but it contains strictly less information than τ_i itself, which encodes both sign and magnitude. Knowing the subjects' best guess of their own τ_i allows rearchers to examine any hypotheses about the treatment's causal effect, not just its sign. It is straightforward to get subjects to reveal one potential outcome or the other. The trouble is getting them to make a good guess about the *missing* potential outcome.

We developed three approaches, all of which are again variations on a theme. In all three, subjects reveal their best guess of both of their potential outcomes. The formats

differ according to which potential outcome is asked first, and whether the questions are asked sequentially (i.e., on two different slides) or simultaneously (i.e., on the same slide).

- *Both POs (Y0 first).* Subjects are asked for their control outcome, then are exposed to the treatment and asked for their treatment outcome.

- *Both POs (Y1 first).* Subjects are exposed to the treatment and asked for their treatment outcome. Subjects are then asked to imagine what their control outcome would have been.

- *Both POs (Simultaneous).* Subjects are exposed to the treatment and asked to report both potential outcomes simultaneously.

Comparing self-reports to experiments

The Fundamental Problem of Causal Inference limits the methods available for assessing the accuracy of these question formats. Because τ_i is unobservable, we cannot (1) assess the accuracy of any individual's self-report of attitude change or (2) compute the true proportion of the population whose attitude changed. This renders our situation somewhat different than experimental validations of self-reports in which the true value of the self-reported quantity is known. For example, in Vavreck (2007) and Jerit et al. (2016), the researchers were able to compare self-reports to true media exposure because they exposed the subjects.

The social scientist's standard response to the unobservability of τ_i is to change the target of inference from the individual level treatment effect τ_i to the average treatment effect (ATE), or $E[\tau_i]$. Because the *both PO* formats elicit each respondent's estimate of their own τ_i, comparing these methods to an experiment is straightforward: simply compare the average self-reported τ_i to the experimental estimate of the ATE. The subject's self-report of τ_i also allows us to compute $\text{sign}(\tau_i)$ for comparison to the *direction* formats.

Inconveniently, the properties of the *direction* formats preclude such straightforward comparison to experimental estimates of the ATE. Because these questions do not ask subjects about τ_i's magnitude, we cannot place directional self-reports on the same scale as an ATE estimate or compute even an average direction of attitude change. These limits make the question format's properties difficult to evaluate. In and of itself, we view this as a reason to prefer some version of the *both PO* formats: to the extent that survey researchers use respondent self-reports to make claims about causal effects, they should elicit self-assessments in a format that can be compared to estimates of the true causal effect.

In light of these limits, we resort to two modes of comparison between the *direction* formats and experiments. First, we compare the effect of question format on the percentage of respondents reporting that their attitude did not change. To the extent that explicit consideration of $Y_i(0)$ and $Y_i(1)$ reduce self-reports of change relative to the *direction (no anchor)* format, we find it plausible that the anchored and both PO formats reduce over-reporting. Second, we compare the substantive takeaways from our experiments to the substantive takeaways that would result from over-interpreting the direction questions. To the extent that these modes of inference imply different substantive interpretations, we will conclude that ignoring the bad properties of direction questions leads to bad inference.

Design

We recruited 989 subjects from Lucid, a firm that offers convenience samples of online survey respondents. In this case, Lucid quota sampled respondents to match census demographics. Experiments conducted on Lucid often recover similar treatment effect estimates to those obtained with national probability samples and the more familiar Mechanical Turk samples (Coppock and McClellan 2017).[1] Our survey first assessed subject demographics (age, gender, race/ethnicity, education, income) and political background (ideology, party identification, and political interest) before beginning the experiment.

Subjects were randomly assigned to one of the 10 question format conditions listed in Table 1. Each subject then participated in a series of mini-experiments. Subjects assessed five candidates for Congress (two Democrats and three Republicans) and three policy areas (immigration, the death penalty, and taxes). The candidates were real current or former state legislators, with vignette information quoted almost verbatim from candidate web sites and public information about bills they cosponsored. Subjects saw all five candidate questions before the three policy questions. We kept the question format the same throughout so as to not confuse our subjects with a barrage of nonstandard question formats.

For each question format, Table 1) displays the sequence of information presented to, and questions asked of, subjects. Subjects saw the treatment or control information, then answered questions designed to elicit their potential outcomes $(Y_i(0), Y_i(1))$ or the direction of their treatment effect $(\text{sign}(\tau_i))$. In the table, $\text{sign}(\tau_i)$ always represents a (quite possibily incorrect) guess, but our notation for the Y's features a slight ambiguity. For the "standard

---

[1]For a paper that uses a Lucid sample to survey a large number of bilingual respondents, see (Flores and Coppock 2018). For evidence that Lucid respondents have similar demographic knowledge patterns as respondents in the American National Election Study (ANES), see (Graham 2018).

Table 1: Survey Design

Condition	N	Vignette 1	Question(s) 1	Vignette 2	Question 2
Standard Experiment (control)	103	Control	$Y_i(0)$		
Standard Experiment (treatment)	96	Treatment	$Y_i(1)$		
Both POs (simultaneous)	99	Treatment	$Y_i(0), Y_i(1)$		
Both POs ($Y_i(1)$ first)	99	Treatment	$Y_i(1)$	Control	$Y_i(0)$
Both POs ($Y_i(0)$ first)	99	Control	$Y_i(0)$	Treatment	$Y_i(1)$
Direction (no anchor)	98	Treatment	$\mathrm{sign}(\tau_i)$		
Direction ($Y_i(0)$ anchor)	99	Treatment	$Y_i(0), \mathrm{sign}(\tau_i)$		
Direction ($Y_i(0)$ first)	97	Control	$Y_i(0)$	Treatment	$\mathrm{sign}(\tau_i)$
Direction ($Y_i(0)$ & $Y_i(1)$ anchor)	98	Treatment	$Y_i(0), Y_i(1), \mathrm{sign}(\tau_i)$		
Direction ($Y_i(1)$ anchor)	101	Treatment	$Y_i(1), \mathrm{sign}(\tau_i)$		

experiment" rows, $Y_i(1)$ and $Y_i(0)$ refer to the actual potential outcomes that are revealed. In the remaining rows, they sometimes refer to subjects' best guess of those potential outcomes. We left this ambiguity to spare the reader the additional notational elaboration.

We elicited $Y_i(1)$ and $Y_i(0)$ using seven-point scales from 1 to 7. For the five candidates we asked: "If [candidate] were running for Congress in your district against a moderate [outpartisan], how likely would you be to support [him/her]?" [1: Nearly zero, 2: Very unlikely, 3: Slightly unlikely, 4: No opinion, 5: Slightly likely, 6: Very likely, 7: Nearly certain]. For the three policy topics we asked: "How strongly do you support or oppose [policy]?" [1: Strongly oppose, 2: Oppose, 3: Slightly oppose, 4: Neither support nor oppose, 5: Slightly support, 6: Support, 7: Strongly support]. Whenever respondents had already been exposed to treatment, questions about $Y_i(0)$ instructed respondents to "suppose you did not know" the treatment detail about the candidate or "suppose you had not seen the information" about the policy. To ask about $\mathrm{sign}(\tau_i)$, we asked "does the fact that [name] [action] make you more or less likely to support [pronoun] an election against a moderate [outpartisan]?" for the candidates and "does the information make you more or less supportive of [policy]?" for the policies. We coded this variable $\{-1, 0, 1\}$.

The candidate mini-experiments had treatment and control versions of the vignette, which differed only in their inclusion of the treatment information. For the policy topics, only treatment respondents saw any vignette. Tables 2 and 3 list the full text of each vignette. The online appendix presents full-text vignettes and questions for every combination of mini-experiment and question format.

Table 2: Candidate Evaluation Treatments

|	Control	Treatment
Candidate 1	Kevin C. Kelly is a Republican state senator. He also owns a law firm. He and his wife Cindy have four children and five grandchildren. According to his web site, Kelly • Is an advocate for common sense solutions to reduce government spending and create jobs for middle-class families. • Prioritizes environmental stewardship. • Has made an impact [on] health insurance & hospitals, Medicaid, and consumer protection.	Kevin C. Kelly is a Republican state senator. He also owns a law firm. He and his wife Cindy have four children and five grandchildren. According to his web site, Kelly • Is an advocate for common sense solutions to reduce government spending and create jobs for middle-class families. • Prioritizes environmental stewardship. • Has made an impact [on] health insurance & hospitals, Medicaid, and consumer protection. In the 2016 election, Kelly endorsed Donald Trump for president.
Candidate 2	Ricardo Lara is a Democratic state senator. He is the first openly gay person of color to be elected to the his state's Senate. According to his web site, Lara • Authored the L.E.A.R.N. Initiative, which asked voters to expand multilingual programs for all students. • Has emerged as one of the most effective environmental champions in the state by consistently passing policies that improve health conditions and rein in air pollution.	Ricardo Lara is a Democratic state senator. He is the first openly gay person of color to be elected to the his state's Senate. According to his web site, Lara • Authored the L.E.A.R.N. Initiative, which asked voters to expand multilingual programs for all students. • Has emerged as one of the most effective environmental champions in the state by consistently passing policies that improve health conditions and rein in air pollution. Lara has been accused of using his committee chairmanship to block whistleblower legislation. The bill would have made sure legislative staff who come forward with sexual harassment allegations are not punished.
Candidate 3	Don Shooter has been a Republican member of the state Senate since 2010. He works as a farmer and real estate developer. He has three children and five grandchildren. According to his web site, Shooter • wants to solve illegal immigration by enforcing immigration laws, ending sanctuary policies, and enforcing identification standards • cut spending and produced a balanced budget as chair of the appropriations committee.	Don Shooter has been a Republican member of the state Senate since 2010. He works as a farmer and real estate developer. He has three children and five grandchildren. According to his web site, Shooter • wants to solve illegal immigration by enforcing immigration laws, ending sanctuary policies, and enforcing identification standards • cut spending and produced a balanced budget as chair of the appropriations committee. Shooter wants to expand school choice, including charter schools and vouchers that can be used at private schools.
Candidate 4	Tony Cornish, a Republican, was first elected to the state legislature in 2002. He grew up on a small farm. Before entering politics, he worked as a sheriff and game warden. According to his web site, Cornish • Fought against government waste and opposed the governor's plan to raise sales taxes. • Played a key role in crafting a new policy that allows county attorneys to carry handguns at work. • Increased prison sentences for car thiefs and other criminals.	Tony Cornish, a Republican, was first elected to the state legislature in 2002. He grew up on a small farm. Before entering politics, he worked as a sheriff and game warden. According to his web site, Cornish • Fought against government waste and opposed the governor's plan to raise sales taxes. • Played a key role in crafting a new policy that allows county attorneys to carry handguns at work. • Increased prison sentences for car thiefs and other criminals. Cornish has been accused of making inappropriate sexual comments by fellow legislator Erin Quade, a Democrat. Cornish denied the allegations, saying he was "blindsided." Quade admitted having a "cordial and collegial relationship" with Cornish but said that "doesnt excuse sexual harassment."
Candidate 5	Dean Westlake, a Democrat, is a state representative. He fights for funding for rural schools and strongly supports municipal revenue sharing. As a legislator, Westlake • sponsored legislation to allow prisons to profit off contracts that put inmates to work for private companies. • sponsored a bill to increase regulations on Caribou hunting.	Dean Westlake, a Democrat, is a state representative. He fights for funding for rural schools and strongly supports municipal revenue sharing. As a legislator, Westlake • sponsored legislation to allow prisons to profit off contracts that put inmates to work for private companies. • sponsored a bill to increase regulations on Caribou hunting. After seven state legislative staff accused Westlake of groping and unwanted sexual advances, reporters found old court records that suggest he impregnated a 15-year-old when he was 28.

Table 3: Policy Treatments

Policy 1	Today about 14 percent of people in the United States are immigrants, meaning that they were born in another country. This is about the same level as the 1860s to 1920s, but is higher than more recent years. The Census Bureau projects that in about five years, the immigrant population will surpass the previous high of 14.7 percent. By 2050, 18 percent of Americans will be immigrants.

Immigrant population, 1880-2050
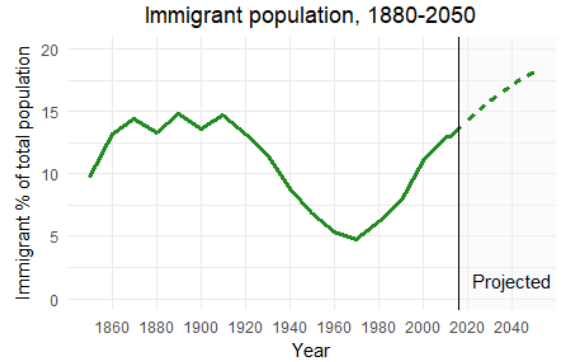
Policy 2	Murder rates in the United States rose in the past couple years, but are only about half as high as in 1990. The decline occurred at about the same pace in states with and without the death penalty. In fact, over the past three decades, states without the death penalty have had a lower murder rate than states that execute people who are convicted of murder.

State murder rates by death penalty law
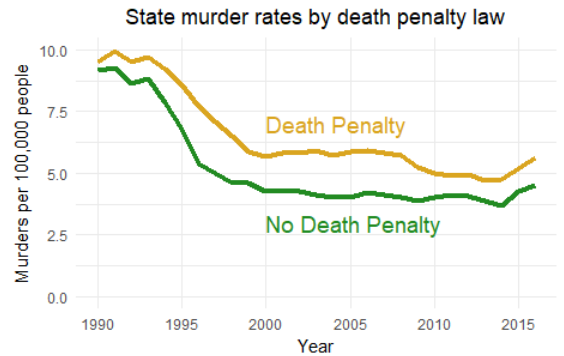
Policy 3	The Tax Cuts and Jobs Act made many changes to America's tax system, including:

- Cuts income taxes for all Americans by lowering rates by a few percent in each tax bracket.
- Cuts the corporate tax rate from 35 percent to 21 percent.
- Limits or eliminates many deductions. The mortage interest deduction would be limited to $750,000, and the deduction for state and local taxes would be limited to $10,000.
- Expands the child tax credit from $1,000 to $2,000. Families who do not pay taxes can receive up to $1,400 instead of the full credit.
- Most of the tax cuts for individuals expire in 2025. Unless Congress extends the cuts, many Americans will receive a tax increase relative to current law |

9

Results

We report three sets of results. First, we describe how the distributions of the "direction" questions change depending on format. Second, we assess the accuracy of the "both potential outcomes" formats relative to an experimental benchmark. Third, we compare the substantive takways from the experiments to the conclusions that would result from over-interpreting the "direction" questions.

Results I: Direction formats

We begin by describing the distributions of self-reported attitude change by topic and question format. Figure 1 plots the percentage of respondents in each condition who reported that the information made them more supportive of the candidate or policy, less supportive, or caused no change. In all eight mini-experiments, the share of subjects who report attitude change (i.e., do not say "no change") tends to be largest in the *direction (no anchor)* format and successively smaller in the other formats. Figure 2 presents a series formal tests of this claim. For each question format in Figure 1, Figure 2 shows the difference in the percentage reporting that their attitudes changed as a result of the treatment information.[2]

The anchoring conditions decrease the percentage of respondents self-reporting attitude change. The average decrease is about 4 percentage points (7 percent) in the two anchored direction formats that include only $Y_i(0)$ and about 12 percentage points (19 percent) in the two direction formats that include $Y_i(1)$. The *both PO* formats reduce self-reported change by an even larger magnitude, about 23 percentage points (38 percent). Out of 56 opportunites across the eight mini-experiments, all but four estimates are negative, with all point positive estimates coming in the *direction (Y0 first)* format. Nearly half (25) are both negative and statistically significant, while no positive point estimate is statistically significant or larger than 2 percent.

We interpret this to mean that subjects asked the standard *direction (no anchor)* questions overstate their attitude change. We also notice the larger differences for the policies than for the candidates. We speculate that subjects may, upon reflection, find policy information less important than information about real but obscure politicians.

---

[2]While some analysts prefer to analyze binary outcome variables with generalized linear models such as logit or probit, this is unnecessary. The difference-in-means is unbiased for the average treatment effect regardless of the outcome space (Gerber and Green 2012, chapter 2). As is common, none of our substantive interpretations depend on this choice.

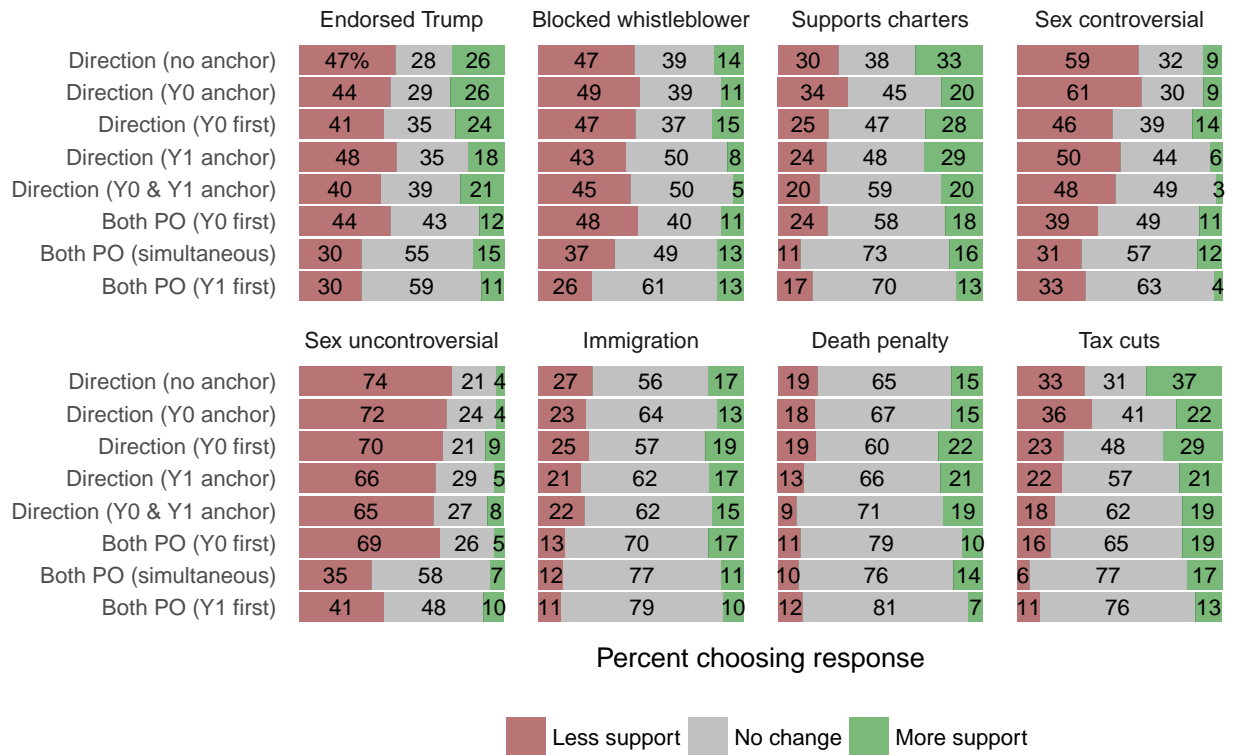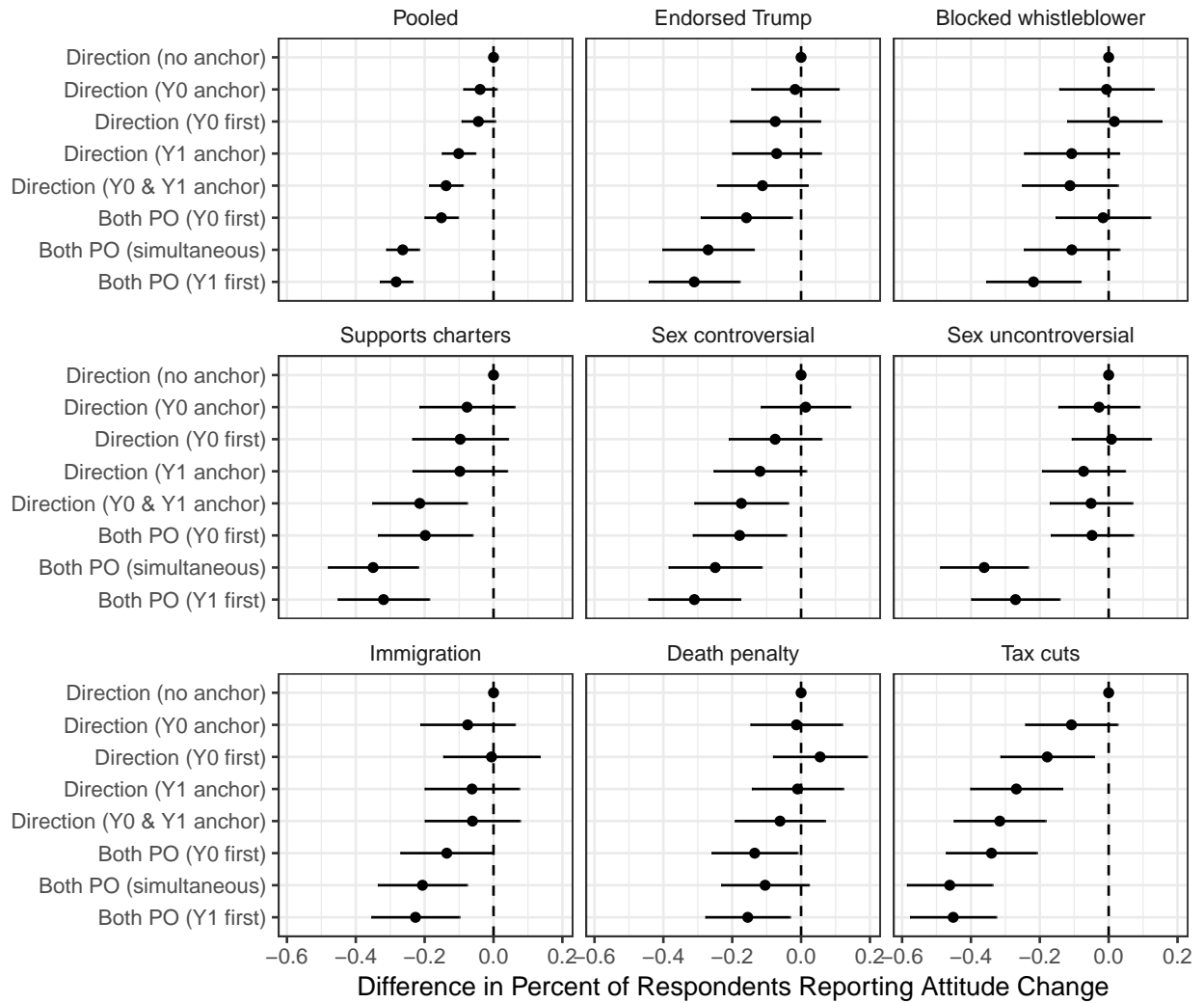Figure 1: Distribution of "Change" Dependent Variable

Figure 2: Effects of Question Format on Percent Reporting Attitude Change

Results II: Both potential outcomes formats

The three *both PO* measurement strategies, as well as *direction (Y0 & Y1 anchors)*, elicited each resondent's best guess of both $Y_i(0)$ and $Y_i(1)$. This allows us to compare the ATE estimate from a standard experiment to the estimate derived from respondents' best guesses.

Figure 3 plots five estimates of the ATEs for each of the eight topics. Pooling across the eight mini-experiments, three of the methods appear conservative: the mean absolute treatment effecs for *both PO (simultaneous)*, *both PO (Y1 first)*, and *direction (Y0 & Y1 anchors)* are respectively 0.24, 0.32, and 0.25, compared with 0.55 for the experiment. The *both PO (Y0)* condition does not appear similarly conservative: its mean absolute treatment effect of 0.54 is quite close to the experiments.

In terms of their ability to get the sign right, all of the methods perform similarly: in only one instance does even a borderline-significant experimental treatment effect take a different sign than a borderline-significant self-report (Republicans, "supports charters", *both PO (Y0 first)* condition). In a few cases, all four self-report methods are systematically lower

Figure 3: Actual versus Self-Reported Average Treatment Effect

Average Treatment Effect Estimate

Party — Democrat — Republican
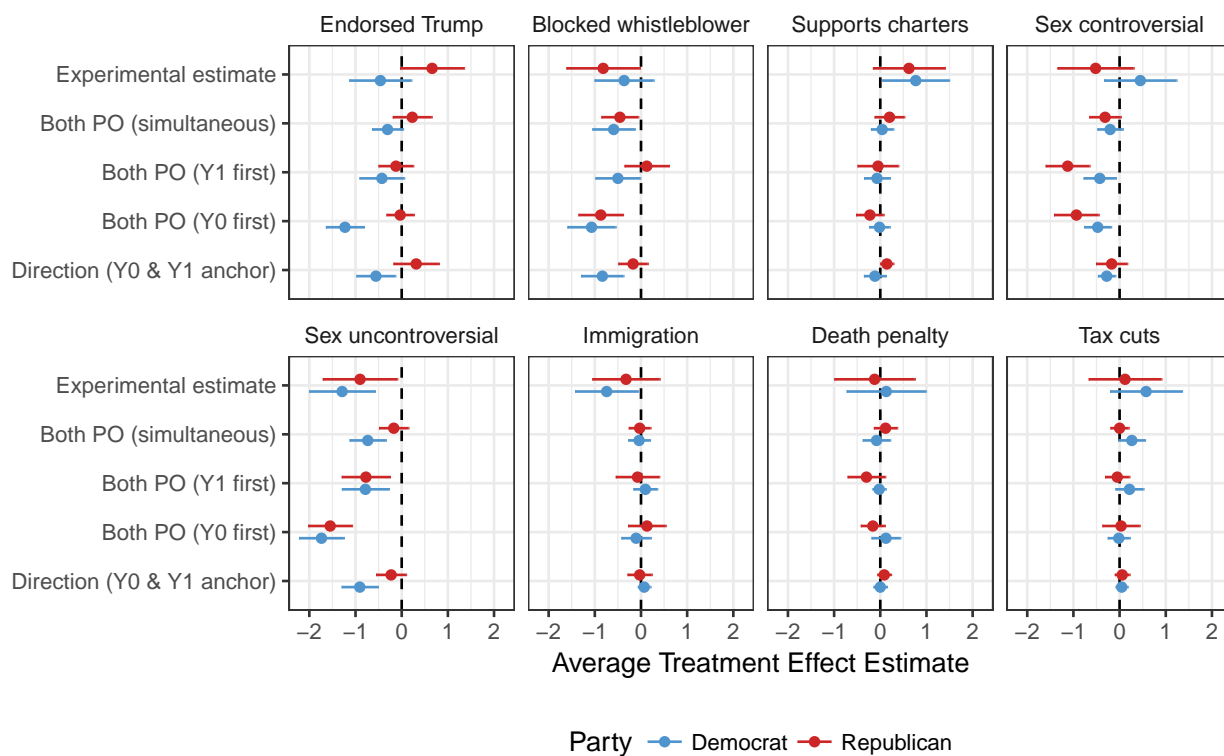
in absolute value than the experiment (supports charters, immigration, perhaps Democrats on tax cuts). In only one case does the substantive story appear incorrect: in the "sex controversial" condition, Democrats self-report small but statistically significant average treatment effects of -0.47 to -0.20, but the statistically insignificant experimental estimate has a point estimate of about 0.45 (se= 0.40).

We take this as evidence that the *Both PO* formats provide a reasonable, if sometimes conservative, approximation of how information affects attitudes.

## Results III: Question format can change the substantive story

Above, we demonstrated that (1) our alternative formats reduce self-reports of attitude change relative to *direction (no anchor)* and (2) the *both PO* formats provide a reasonably accurate, possibly conservative estimate of the average treatment effect. Does this mean that public opinion researchers should prefer the *both PO* format? One could argue that rather than over-reporting attitude change, subjects in the *direction (no anchor)* condition are accurately reporting small amounts of change. The obstacles to inference we described above make this proposition difficult to test: we cannot directly test the accuracy of an individual's self-report, place the average self-report on the same scale as the ATE, or even meaningfully aggregate the direction responses into an average amount or direction of change.

Despite the fundamental challenges associated with benchmarking direction questions against an experiment, we wish to provide evidence that may convince readers who are more sympathetic to the direction formats. Rather than ask the unanswerable ("are directional self-reports accurate"), this section asks a more tractable question: *how often does the balance of directional self-reports imply the same substantive story as an experiment*? To aid our analysis, Figure 4 displays, for each study and question format, the distribution of the change variable for Democrats and Republicans. The online appendix presents a corresponding numerical table with standard errors.

We find the balance of more/less directional answers to be a misleading proxy for the true direction of attitude change. In contrast to the *both PO* formats, where apparent mis-reporting usually erred on the side of caution, our mini-experiments present situations in which over-interpretation of directional self-reports would cause an observer to (1) overstate the magnitude of attitude change, (2) mis-state the direction, and (3) perceive attitude polarization where none existed. We describe these instances in turn.

In the first set of cases, the experimental treatment effect is near-zero but respondents overwhelmingly report change in one direction. In "sex controversial" 88 percent of

Figure 4: Distribution of "Change" Dependent Variable by Party
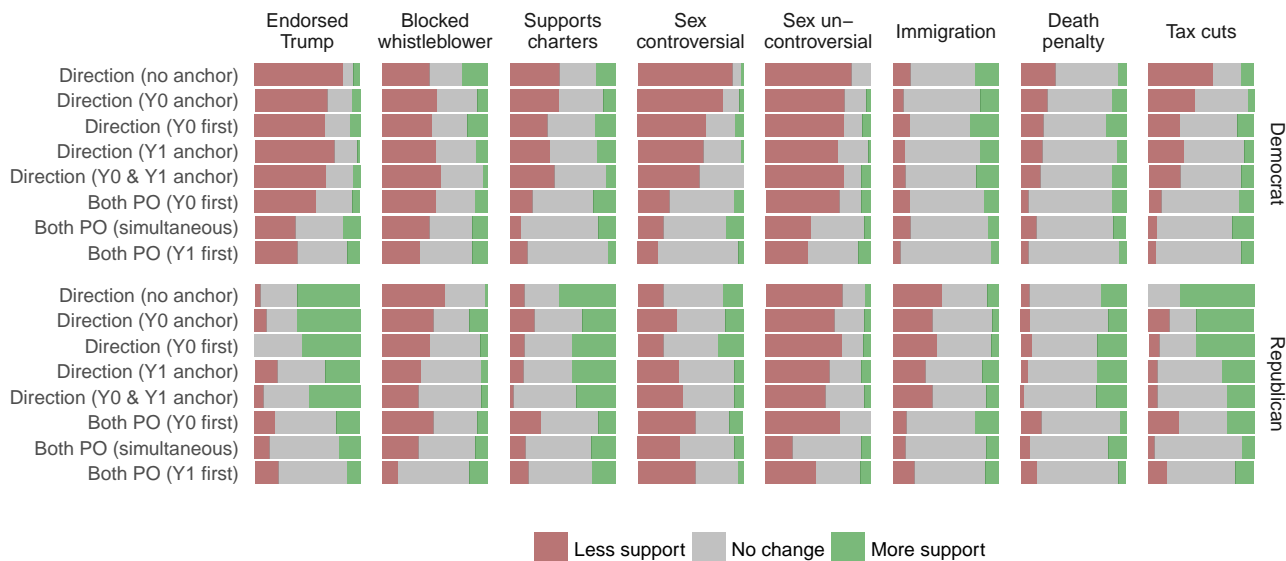
Democrats report becoming less supportive of Republican Tony Cornish as a consequence of the allegations against him, compared with zero percent more supportive. Yet the experimental point estimate is slightly positive and indistinguishable from zero (Figure 3). Similarly, 70 percent of Republicans say details about the Tax Cuts and Jobs Act made them more supportive, against zero less supportive; the experimental treatment effect is almost exactly zero.

In a second pair of cases, over-interpretation of the balance of directional self-reports would cause an observer to predict the wrong sign for the treatment effect. In "supports charters" a substantial plurality of Democrats report that Don Shooter's support for charter schools reduced their support for him (without anchors, 47 percent "less" versus 18 percent "more"). Yet in the experiment, the treatment effect is positive ($\beta = 0.77, \text{se} = 0.36$). On "immigration" Democrats are slightly more likely to report that the information made them more supportive of immigration,[3] but the experimental treatment effect is negative ($\beta = -0.74, \text{se} = 0.34$).

In a third set of cases, over-interpretation of the direction question would lead an observer to infer attitude polarization where none exists. Just as Republicans overwhelmingly

---

[3]This difference is only statistically significant if one pools across the direction formats. Combining these formats, 21.3 percent of Democrats report that the information increased their support for immigration, compared with 12.9 percent who say it reduced their support (difference = 0.083, $\sigma = 0.037$).

suggested that the Tax Cuts and Jobs Act's provisions made them more supportive, most Democrats claimed they became less supportive (without anchors, 62 percent "less" versus 12 percent "more"). The experiment revealed no evidence of attitude polarization, reducing the difference between the parties by a statistically insignificant amount.[4] Likewise, both parties have a treatment effect of zero in "death penalty," yet over-interpretation of the more/less question would suggest attitude polarization.[5] Similarly, directional self-reports in "supports charters" suggest polarization but the experiment suggests nearly identical movement in Shooter's favor.

All told, five of the eight mini-experiments suggest substantive differences in interpretation between experiments and directional self-reports. These differences are usually due to overstatement of the direction of attitude change, which is most severe in the *direction (no anchor)* format.

Discussion

Our first goal was to assess the properties of the *direction (no anchor)* format. Of the eight self-assessment methods we tested, our results suggest that it is the most misleading. Respondents reported more change in this format than any other format, leading to substantive conclusions that are even more misleading than the other direction methods. Survey researchers should not ask questions in this format.

Our second goal was to come up with alternatives that perform better. Our first set of alternative formats, *anchored direction*, reduced apparently-misleading self-reports of attitude change, but in most cases this reduction was not enough for the formats to imply a conclusion consistent with a randomized experiment. Discouraging as this may be, we cannot entirely foreclose the possibility that some direction format could lead to better answers. In future iterations we may focus on $Y_i(1)$ anchors and vary the instructions as another possible way to make subjects introspect more accurately.

To our surprise, the *both PO* formats were a qualified success. These formats allow

---

[4]Republicans rated the Act 2.16 scale points higher than Democrats in control and 1.70 scale points higher in treatment. To test the statistical significance of this difference in differences, we restricted our sample to Democrats and Republicans and ran the OLS regression $Y_i = \beta_0 + \beta_1 z_i + \beta_2 R_i + \beta_3 z_i R_i + \epsilon_i$, where z is a treatment indicator and R indicates Republican partisanship. The estimate for $\hat{\beta}_3$ is -0.46 with a HC2 robust standard error of 0.56.

[5]Democrats were 24 percent more likely to say "less" than "more" and Republicans were 16 percent more likely to say "more" than "less." The difference in differences, 41 percent, is statistically significant (SE = 12 percent).

us to infer a self-reported treatment effect from subjects' treated and untreated potential outcomes. Both numerically and substantively, we found that self-reported treatment effects were fairly consistent with experimental treatment effects, with apparent errors tending to under-state the magnitude of attitude change. In future iterations we plan to give more attention to the *both PO* format. We may also ask respondents to directly state τ_i.

We note that the choice among these approaches for survey researchers amounts to a bias-variance tradeoff. Standard information experiments are unbiased but are high variance. They are high variance because we literally only observe half the data: treated potential outcomes in the treatment group and untreated potential outcomes in the control group. The *both PO* formats are lower variance because we have estimates of the full dataset: treated and untreated potential outcome guesses for all subjects. But of course, subjects could be *wrong* about their own potential outcomes, in which case the resulting estimates are biased. Our study shows that the bias appears to be small, suggesting that this format could be viable in settings where experimentation is too expensive or otherwise infeasible.

Our results permit some informed speculation about the nature of responses to the Roy Moore poll described in the introduction. Laying aside our personal feelings about allegations, the poll was taken at a time of controversy over their substance. To roughly approximate these circumstances we included the "sex controversial" treatment, which (using text from news reports) framed the allegations against former state representative Tony Cornish (R-MN) in a manner designed to make them seem debateable. In the *direction (no anchor)* format, as well as direction formats with only a $Y_i(0)$ anchor, Republicans were relatively divided in their self-reports, with many reporting that they had become more supportive of Cornish. $Y_i(1)$ anchors changed the story: Republicans and Democrats alike were at least 30 percent more likely to say "less" than "more." In substantive terms, the *both PO* formats appeared even more accurate: more Republicans than Democrats reported losing support for Cornish, consistent with the experiment.[6] By analogy, we suspect that some respondents might have been stating their continued support for Moore despite the allegations rather than the direction of their attitude change.

Together, these results illustrate the risks of attempting to non-experimentally esti-mate causal effects. Absent some form of experimental validation of the question format, researchers should refrain from asking survey respondents to self-report their attitude change. Though we would not full-throatedly endorse any method on the basis of one study, we were

---

[6]For Democrats and Republicans in the "sex controversial" experiment, we used OLS to estimate $Y_i = \beta_0 + \beta_1 z_i + \beta_2 R_i + \beta_3 z_i R_i + \epsilon_i$, where z is a treatment indicator and R indicates Republican partisanship. The estimate for $\hat{\beta}_3$ is -0.161 with a HC2 robust standard error of 0.95.

surprised to find that asking respondents to report both of their potential outcomes can somewhat accurately approximate an experiment. This format is conservative: when it appears to be wrong, it under-states the extent of attitude change, an important characteristic in a world that lauds much more focus on apparent change than null results. We hope this work lays the foundation for continued investigation of whether and how respondents can accurately self-report how information and events changed their attitudes.

References

Babbie, Earl R. 2011. *The Practice of Social Research.* SAGE Publications.

Coppock, Alexander and Oliver A. McClellan. 2017. "Validating the Demographic, Political, Psychological, and Experimental Results Obtained from a New Source of Online Survey Respondents." *Unpublished manuscript* .

Flores, Alejandro and Alexander Coppock. 2018. "Do Bilinguals Respond More Favorably to Candidate Advertisements in English or in Spanish?" *Political Communication* .

Fowler, Floyd J. 1995. *Improving Survey Questions.*

Fowler, Floyd J. 2014. *Survey Research Methods.* 5 ed. SAGE Publications.

Gal, David and Derek D Rucker. 2011. "Answering the Unasked Question: Response Substitution in Consumer Surveys." *Journal of Marketing Research (JMR)* 48(1):185–195.

Gerber, Alan S. and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation.* New York: W.W. Norton.

Graham, Matthew H. 2018. "Self-Awareness of Political Knowledge." *Political Behavior* forthcoming.

Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Elanor Singer and Robert Torangeau. 2009. *Survey Methodology.* Wiley.

Guess, Andrew and Alexander Coppock. 2018. "Does Counter-Attitudinal Information Cause Backlash? Results from Three Large Survey Experiments." *British Journal of Political Science* forthcomin.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396):945–960.

Jerit, Jennifer, Jason Barabas, William Pollock, Susan Banducci, Daniel Stevens and Martijn Schoonvelde. 2016. "Manipulated vs. Measured: Using an Experimental Benchmark to Investigate the Performance of Self-Reported Media Exposure." *Communication Methods and Measures* 10(2-3):99–114.

Lord, Charles G., Lee Ross and Mark R. Lepper. 1979. "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology* 37(11):2098–2109.

Miller, Arthur G., John W. McHoskey, Cynthia M. Bane and Timothy G. Dowd. 1993. "The Attitude Polarization Phenomenon: Role of Response Measure, Attitude Extremity, and Behavioral Consequences of Reported Attitude Change." *Journal of Personality and Social Psychology* 64(4):561–574.

Sudman, Seymour and Norman M. Bradburn. 1982. *Asking Questions: A Practical Guide to Questionnaire Design.* San Francisco: Jossey-Bass.

Torangeau, Robert and Lance J. Rips. 2000. *The Psychology of Survey Response.* Cambridge University Press.

Vavreck, Lynn. 2007. "The Exaggerated Effects of Advertising on Turnout: The Dangers of Self-Reports." *Quarterly Journal of Political Science* 2(4):325–343.
URL: *http://www.nowpublishers.com/article/Details/QJPS-6005*

Yair, Omer and Gregory A Huber. 2018. "How robust is evidence of perceptual partisan bias in survey responses? A new approach for studying expressive responding." *Unpublished Manuscript* .