Asking About Attitude Change

Matthew Graham and Alexander Coppock*

September 28, 2019

6,853 words

Abstract

Surveys often ask respondents how information or events changed their attitudes. Does [information X] make you more or less supportive of [policy Y]? Does [scandal X] make you more or less likely to vote for [politician Y]? We show that this type of question (the *change format*) exhibits poor measurement properties, in large part because subjects engage in response substitution. When asked how their attitudes changed, people often appear to report the level of their attitudes instead. As an alternative, we propose the *counterfactual format*, which asks subjects what their attitude would have been in the counterfactual world in which they did (or did not) know a particular piece of information. Using a series of experiments embedded in four studies, we show that our alternative approach reduces response substitution and comes closer to experimental estimates of average causal effects of information.

*Matthew Grahem is a Doctoral Candidate in Political Science, Yale University. Alexander Coppock is Assistant Professor of Political Science, Yale University. We thank Jordan Farenhem for his excellent research assistance, and Don Green and Molly Offer-Westort for providing survey space. For helpful comments on earlier versions of this paper, we are grateful to Greg Huber, seminar participants at Yale, and conference attendees at the annual meetings of the American Political Science Association and the Midwest Political Science Association.

In advance of Alabama's 2017 special election for U.S. Senator, polling firm JMC Analytics released a survey that sought to estimate the effect of sexual misconduct allegations on support for Republican Roy Moore. 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% responded "more likely," 38% "less likely," and 33% "no difference." Among self-identified evangelical Christians, "more" outnumbered "less." Many commentators decried the apparent depravity of those whose support for Moore had allegedly increased as because of the allegations (e.g., Wilson 2017; Ballesteros 2017).

Surveys often ask respondents to assess the causal effect of some news event on their attitudes. In this article, we argue that *change questions* of this type exaggerate the extent to which events and information cause attitude change. We believe this bias is at least partially explained by a phenomenon known as response substitution (Gal and Rucker 2011; Yair and Huber 2019), wherein respondents use the question to indicate the level of their opinion rather than the changes in it. According to the response substitution interpretation, Alabama voters were not saying they support Roy Moore *more* because he was accused of sexual misconduct. Instead, they saying they support him *anyway*.

To some, it may seem obvious that self-reports of attitude change should not be taken at face value. Indeed, some analysts argued that the poll really meant that 29% of Alabama voters wanted to express support for Moore despite the scandal (e.g., Klein 2017). Although we agree with this interpretation, we note that it subtly switches the estimand from the average causal effect of the scandal on support to the average level of support after the scandal. In our view, researchers interested in this second estimand should choose questions that directly measure it. In this article, we focus squarely on the original, causal estimand.

Questions about attitude change are a staple of public opinion surveys. Examples

stretch back to the early days of polling.[1] To better understand their contemporary uses, a research assistant helped us document nearly 200 self-reported attitude change questions that appeared in statewide or national polls during 2017 and 2018 (see Table 1). The largest category of questions concerned the effects of policy positions on candidate support. Other common topics included how candidate endorsements affected support for other candidates; how information affects attitudes; and how events changed social and economic behavior. Many of these concerned sexual misconduct, including 15 questions about controversies surrounding Supreme Court Justice Brett Kavanaugh.

Despite their prevalence, we are not aware of survey methodological research that evaluates change questions. We consulted an assortment of survey design textbooks (Babbie 2011; Fowler 1995, 2014; Groves et al. 2009; Tourangeau et al. 2000; Sudman and Bradburn 1982). While all are valuable works, none contained specific guidance for questions that ask subjects to assess causal effects. Given their common appearance in public discourse—including, in the Roy Moore case, as a basis for profoundly negative judgments about Alabama voters—it is important for survey researchers to understand what change questions measure and whether they measure it accurately.

We have two main goals in this article. First, we seek to explain why asking about attitude change generates biased inferences. As stated above, our main explanation is that subjects engage in response substitution. We demonstrate that this is a plausible explanation with a series of randomized experiments in which some subjects are randomly assigned to state the absolute level of their opinion before answering the change question. We show that this manipulation reduces self-reports of attitude change in line with the response substitution explanation.

Our second goal is to propose an alternative approach to asking about attitude change. We introduce the *counterfactual format*, which proceeds in two steps. The first step is a

[1]The earliest example we have encountered is summarized by Dahl (1961). In 1954, the Survey Research Center asked whether respondents would be more or less likely to vote for a politician who had the support of Senator Joseph McCarthy.

Table 1: Self-Reported Attitude Change Questions in Public-Facing Polls

Category	Polls	Qs	Example
Candidate positions	25	68	If your member of Congress voted for the health care bill currently being considered by Congress, would that make you more or less likely to vote for them in the next election, or would it not make a difference either way? — *Public Policy Polling, July 2017*
Endorsements by or support for other people	25	33	If [Claire McCaskill / Heidi Heitkamp / Joe Donnelly] votes against Brett Kavanaugh's nomination to the Supreme Court, would that make you more likely or less likely to vote for [her / him], or would it not make a difference to your vote for Senate? — *Fox News, September 2018.*
Politics and social or economic behavior	13	17	As you may know, some athletes and sports teams have begun not standing during the national anthem in order to protest police violence against the black community in the United States. Does this make you more likely, less likely or has no influence on you to watch NFL games on television? — *University of North Florida, September 2017.*
Attitudes	12	30	If you knew that the Republican tax plan would cause a significant increase in the national debt over the next 10 years, would that make you more likely to support it, less likely to support it, or would it not have an impact? — *Quinnipiac, November 2017*
Misconduct	10	13	Does the issue of sexual harassment make you more likely to vote for a [Democratic / Republican / woman] candidate, or not? — *Quinnipiac, December 2017*
Candidate attributes	9	12	Stacey Abrams has discussed being more than $200,000 in debt. Does Stacey Abrams' debt make you more likely to consider voting for her? Less likely to consider voting for her? Or does it make no difference? — *SurveyUSA, May 2018*
Political participation	6	7	Has what you've seen in Washington over the last year made you more likely to speak up and let your political views be known, less likely to speak up and let your political views be known, or has there been no change? — *CBS, October 2018*

Note: We compiled these questions using two searches. First, we searched the Roper Center's iPoll database using the search string: more OR less OR make% OR likely OR (change AND your) OR (would AND you AND be) OR (rate AND you%). Second, we searched Google for this same search string plus the word "poll." For both searches, we only considered polls conducted between January 1, 2017 and December 31, 2018. The supplementary materials list all of the questions.

straightforward two-arm experiment. Subjects are randomized into treatment and control groups; the treatment group sees the treatment information and the control group does not. All subjects then report their attitudes. The data from this first step can be analyzed in the same manner as any randomized experiment. In the second step, all subjects are asked a follow-up question that prompts them to imagine how they *would have* responded had they been in the other treatment condition. From the pair of answers given by each subject in the first and second steps, we calculate subject beliefs about the causal effect of the treatment information.

An advantage of this procedure is that its two stages can also be used as a validation exercise. To assess the accuracy of the treatment group's guesses about its untreated outcome, we can compare those guesses to the control group. Likewise, we can compare the control group's guesses about treated outcomes to the treatment group. In our empirical examples, we find that although subjects are often fairly accurate in their counterfactual guesses, they sometimes appear to understate attitude change, as if their counterfactual guesses are anchored to their first stage responses. That said—and this is key—when we compare both the change format and counterfactual formats to the results of the randomized experiments, the counterfactual format is clearly more accurate.

Our short-term recommendation for survey researchers is to replace change questions with questions that ask subjects to imagine counterfactual states of the world. In the longer term, we hope this article will start, not end, a conversation about how survey questions can elicit accurate self-reports about causal effects. Asking about attitude change is fraught with challenges but nevertheless important to get right, as these questions are used to inform political decision-making and our collective understanding of public opinion.

Theoretical Framework

The standard approach to asking about attitude change is what we call the *change format*. These questions ask people whether some event or information made them more or

less supportive of a candidate or policy proposal. For example, amid the health reform debate that produced the Affordable Care Act of 2010, angry protestors disrupted legislators' town hall meetings all over the country during Congress's annual August recess. To understand how the protests affected attitudes toward health reform, Gallup and USA Today asked,

> *From what you know or have read, have these town hall meeting protests against the proposed bills made you more sympathetic to the protestors' views, do the protests not make any difference to you either way, or have the protests made you less sympathetic to the protestors' views?*

Overall, 35% of respondents said "more sympathetic" compared with just 21% percent who said "less sympathetic." Republicans split 51% to 8%, while Democrats split 17% to 39%. News coverage of the protests took these self-reports at face value. The opening sentence of *USA Today*'s coverage of the poll declared that "[t]he raucous protests at congressional town-hall-style meetings have succeeded in fueling opposition to proposed health care bills" (Page 2009). This implies the protests caused support for health reform to be lower than it would have been if the protests had not occurred.

To formalize this interpretation of change questions, we draw on the potential outcomes framework, a standard model for thinking about causal inference in the social sciences (Neyman 1923; Rubin 1974). In the binary treatment case and assuming no interference across units, subject i has two potential outcomes. The treated potential outcome, $Y_i(Z = 1)$ (or $Y_i(1)$ for short), is the attitude the subject expresses in the state of the world in which the protests occurred. The control potential outcome, $Y_i(0)$, is expressed in the state of the world with no protests. Change questions ask people to compare the two states of the world, then assess whether their treatment attitude is larger, smaller, or the same as their control attitude. More precisely, denote the effect of the protests on an individual's support for health reform as $\tau_i \equiv Y_i(1) - Y_i(0)$. Change questions ask people to report the sign—positive, negative, or zero—of τ_i.

An important obstacle makes it hard to assess the accuracy of these self-reports. Because the protests already occurred, it is no longer possible to observe attitudes in the

counterfactual world in which they did not occur. We can try to make some reasonable assumptions about what the counterfactual outcome would have been, but it cannot be observed—not even by subject i. Holland (1986) called this the fundamental problem of causal inference. As only one state of the world is ever revealed, we can never observe both $Y_i(1)$ and $Y_i(0)$ for any single person. By extension, we can never directly observe τ_i.[2]

We consider two reasons why people might not accurately report how their attitudes changed. First, peoples' honest best guesses may or may not be accurate. In some circumstances, it is reasonable to trust peoples' causal inferences about themselves. When someone falls and breaks their leg ($Y_i(1)$ = broken leg), it is reasonable to assume that the fall caused the broken leg because the counterfactual outcome ($Y_i(0)$ = intact leg) is easy to credibly impute. For health reform attitudes, the respondent could follow the same basic process: recall their level of support for reform from before the protests, assume that no other factor affected their attitudes, and report how their attitudes are different now.

Counterfactual attitudes may be less straightforward to imagine than an intact leg. In particular, we are concerned that peoples' guesses about counterfactual states of the world may be anchored to the actual state of the world. Some evidence of this pattern comes from longitudinal studies. Markus (1986) found that subjects' memories of their past attitudes were more consistent with present attitudes than with actual past attitudes, and Levine (1997) found that respondents' memories of their emotions following political events were more consistent with current emotions than with past emotions. Schacter (1999) calls this pattern "consistency bias." In our setting, consistency bias would attenuate self-reports of attitude change reports towards zero.

A second reason to mistrust self-reports of attitude change is that people might answer a different question than the one that has been asked. We are particularly concerned that people may use change questions to express the *level* of their attitude. Then-White House

[2]Experiments sidestep the fundamental problem of causal inference by randomly assigning some subjects to reveal $Y_i(1)$ and others to reveal $Y_i(0)$. The difference in means between the two groups is an estimate of the average treatment effect (ATE), even though no individual's treatment effect has been (or could be) measured.

adviser David Axelrod used a version of this argument to critique *USA Today*'s interpretation of the health care protest question:

> White House adviser David Axelrod questioned the USA TODAY survey's methodology, saying those who report being more sympathetic to the protesters now were likely to have been on that side from the start. "There is a media fetish about these things," Axelrod said of the protests, "but I don't think this has changed much" when it comes to public opinion. (Page 2009)

Here, Axelrod suggests that health reform opponents may have used the change question to voice their general opposition to health reform. Under this interpretation, change questions are vulnerable to *response substitution*, or survey subjects' tendency to sometimes answer a different question than the researcher has asked. Gal and Rucker (2011) use the example of a person who liked a restuarant's service but hated the food. Asked to rate the service, this person might say "terrible" even though their true rating of the service is "good." The rating of the food, which the person wanted to express, was substituted for the rating of the service, which they have been given the opportunity to express. This same phenomenon can affect political surveys. When asked to rate the physical attractiveness of potential dating partners, people rate co-partisans lower than out-partisans (Nicholson et al. 2016). This gap shrinks substantially when people are also given the chance to rate the potential dating partner's moral values, suggesting that respondents often use the question about physical attractiveness to express their disapproval of something else (Yair and Huber 2019).

In our setting, response substitution entails using a question about change (did the protests change your attitude?) to state the level of one's attitude (I support/oppose health reform). Accordingly, we expect response substitution to bias attitude change reports away from zero. In groups that are relatively homogenous in their attitudes, the bias should create the false impression that the group collectively moved in the direction of its baseline attitude. When two groups start out on opposite sides of an issue, the bias can make it look as though the event or information drove the two sides apart. If true, the response

substitution hypothesis could explain why Democrats tended to say that the protests made them more supportive of their party's health reform effort while Republicans tended to say they became less supportive.

To test for these biases, we would ideally compare various question formats to a measure of the ground truth. Validation studies of self-reported media exposure (Vavreck 2007; Guess 2015; Jerit et al. 2016) and voter turnout (Holbrook and Krosnick 2010) use this approach, comparing self-reports to objective measures of whether the subject actually voted or saw news coverage. Unfortunately, two properties of change questions make this very difficult. First, the quantity in question (τ_i) is fundamentally unobservable, so we can never obtain an individual-level measure of ground truth. We might attempt to get around this problem by considering *average* treatment effects rather than individual level treatment effects. Here, the change format presents a second problem. The data generated by standard self-reported change questions cannot be used to compute an average causal effect because the change format is *unitless*: it asks only about the sign of attitude change, not the magnitude of attitude change. Consequently, it cannot be aggregated into an average amount of change.

As a consequence of these properties, our first approach to assessing the change format is to conduct an experiment similar in spirit to Gal and Rucker (2011) and Yair and Huber (2019). We will offer some subjects the opportunity to express the *level* of support before reporting the *change* in it. If response substitution biases self-reports of attitude change, this treatment should ameliorate the problem by allowing subjects to answer the question they seem to want to answer. Later, we provide additional evidence of the bias in the change format by comparing the substantive conclusions it implies to the substantive conclusions suggested by a randomized experiment.

The Counterfactual Format

We propose an alternative question format that helps subjects to report their beliefs about the causal effects of events and information on their attitudes. Though this format

will of course not overcome the fundamental problem of causal inference, we designed it in the hope of minimizing response substitution and encouraging respondents to explicitly imagine counterfactual states of the world.

The *counterfactual format* is a two-stage process (visualized in Figure 1). First, subjects participate in a two-arm randomized experiment. We assign m of N subjects to treatment and the remaining $(N - m)$ subjects to control. Subjects randomly assigned to the treatment group are exposed to treatment information; subjects assigned to the control group are not. Both groups then report their attitude. Each member of treatment group reveals their treated outcome, $Y_i(1)$, while each member of the control group reveals their untreated outcome, $Y_i(0)$. Using these data, we can directly estimate the average effect of information using a standard approach like the difference-in-means estimator: $\widehat{ATE_{DIM}} = \frac{\sum_{i=1}^{m} Y_i(1)}{m} - \frac{\sum_{i=m+1}^{N} Y_i(0)}{N-m}$.

Second, subjects imagine what their response *would have been* if they had been in the *other* treatment condition. For control group subjects, this means exposing them to treatment information and asking the outcome question a second time. For treatment group subjects, this means asking them to imagine they did not know the treatment information and then asking the outcome question a second time. This second stage is designed to act as a "counterfactual assist" that helps subjects to think concretely about how their attitude would have been different in an alternative state of the world. Rather than asking directly about attitude change, we infer subjects' beliefs about it by comparing the two questions about attitude levels.

We formalize this procedure as follows. Control group subjects first report their untreated potential outcome, $Y_i(0)$, then their guess of their treated potential outcome, $\widetilde{Y}_i(1)$. We use the tilde to indicate that the measurement is a guess about a counterfactual world. Similarly, treatment group subjects first report their treated potential outcome, $Y_i(1)$, followed by their guess about their untreated potential outcome, $\widetilde{Y}_i(0)$. Interpreting the difference between the two responses as equal to the true effect of treatment requires the strong

assumption that the second stage guesses are correct.

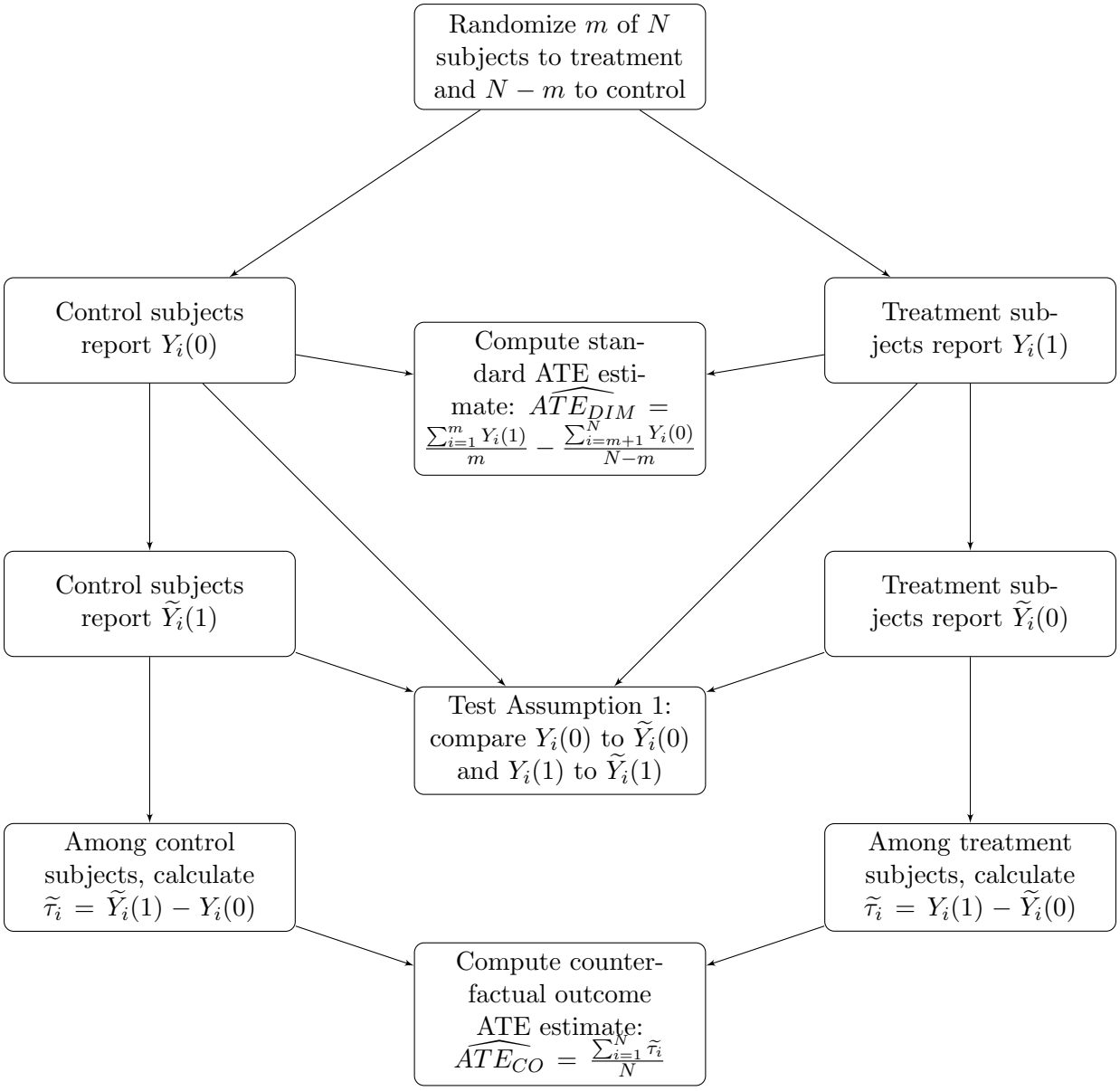
Assumption 1 (Accurate counterfactual guesses) *Subject guesses about counterfactual states of the world are correct:* $Y_i(d) = \widetilde{Y}_i(d), \forall i, d$.

If Assumption 1 holds, we have a direct measure of τ_i, the individual-level treatment effect. For control group subjects, $\widetilde{\tau}_i = \widetilde{Y}_i(1) - Y_i(0)$. For treatment group subjects, $\widetilde{\tau}_i = Y_i(1) - \widetilde{Y}_i(0)$. Taking the average of $\widetilde{\tau}_i$ yields the counterfactual outcome estimate of the ATE, $\widehat{ATE_{CO}} = \frac{\sum_{i=1}^{N} \widetilde{\tau}_i}{N}$.

Assumption 1 could be wrong for many reasons. In particular, responses may suffer from "consistency bias," in the sense that subjects may have a desire to keep their second response closely in line with their first. Or, for an observationally equivalent bias, treatment group subjects might not be able to put the treatment information out of their minds, rendering their guesses about untreated responses artificially closer to their treated responses. Control subjects could overreact to the treatment information and inflate their attitude change or falsely deny the information's influence. We emphatically do not believe Assumption 1 holds in all cases—indeed, it may never hold. We state the assumption as we have in order to clearly delineate what is required to interpret this approach as a method for causal inference, as opposed to a procedure for measuring subjects' biased beliefs about causal effects.

The design of the counterfactual format yields a direct test of Assumption 1. We can directly compare the $Y_i(0)$'s provided by the control group to the $\widetilde{Y}_i(0)$'s provided by the treatment group. Similarly the $Y_i(1)$'s provided by the treatment group cam be compared to the $\widetilde{Y}_i(1)$'s provided by the control group. We can also compare the ATE estimates from the experiment to the ATE estimate implied by the counterfactual guesses. As a matter of measurement, the ability to compare the counterfactual format to three measures of ground truth represents an important advantage over the change format. Below we compare these quantities using a difference-in-means test. As with any hypothesis test, failure to reject the null hypothesis that Assumption 1 does not constitute proof that it is true.

Figure 1: Design of the counterfactual format



The counterfactual format can also be placed in the same units as the change format by taking the sign of each subject's guess about τ_i. Consequently, the counterfactual format can be used to compute quantities that are on the same scale as either the ATE from experiment or the distribution of signs from change format—despite the fact that the unitless nature of change format makes it difficult to compare to an ATE estimate from an experiment.

Example: Tony Cornish (disputed accusation)

Before presenting our full research design, we illustrate our approach using the subset of our results that most closely parallels the Roy Moore example, a set of questions designed to understand the effect of sexual harassment allegations on support for Tony Cornish, a former Republican state legislator from Minnesota. We present the results in same order as our full results below: a test for response substitution in the change format, a direct comparison between the experimental estimates and the counterfactual format, and a judgment as to whether the conclusions implied by the change format or the counterfactual format are more similar to the conclusions implied by the experiment.

The change format

One random subset of our subjects was assigned to answer questions in the change format. These subjects read a short vignette, then answered a question about how one piece of information changed their opinions. The background information was closely paraphrased from Cornish's website. The treatment information describes an allegation against Cornish using real quotes from the Minneapolis *Star Tribune* (Bjorhus and Coolican 2017). The vignette read:

> 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 "doesn't excuse sexual harassment."

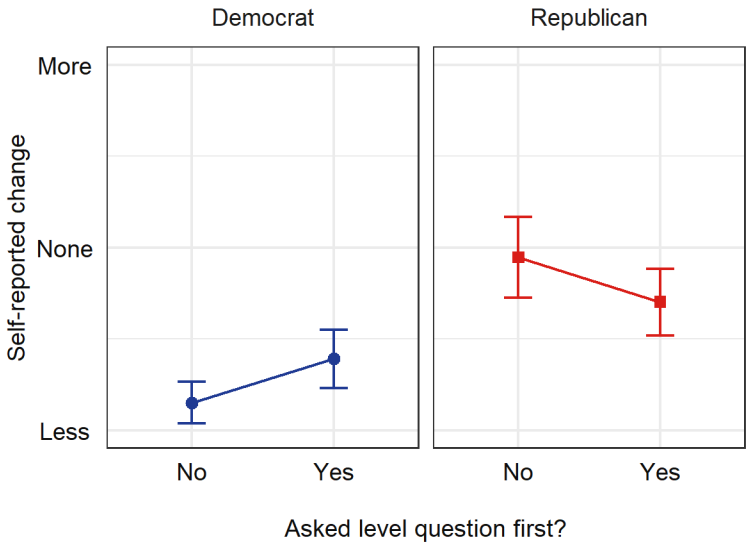
After reading the vignette, subjects were asked: "Does the fact that Cornish was accused of sexual misconduct make you more or less likely to support him in an election against a moderate Democrat? [Less likely, no change, more likely]" Table 2 shows the distribution of responses, broken down by subject partisanship. A huge majority of Democratic respondents (87%) report that the accusation of sexual misconduct made them less likely to vote for Cornish in an election against a moderate Democrat. By contrast, most Republicans (57%) reported that the information had no effect. Similar to the Roy Moore example, we think these self-reports are contaminated by response substitution. We think a sizable share of the Democrats who report becoming less likely to vote for Cornish were simply stating their disapproval. Similarly, we think many Republicans who said "more likely" or "no change" were expressing their continued support despite the allegations.

Table 2: Distribution of responses to the change question (Cornish example)

	Democrats	Republicans
Less likely	87% (46)	24% (9)
No change	11% (6)	57% (21)
More likely	2% (1)	19% (7)

We test the claim that the self-reports are biased due to response substitution by offering a separate, also randomly-selected group the opportunity to express their level of support before answering the change question. The level question read: "If Cornish were running for Congress in your district against a moderate Democrat, how likely would you be to support him? [Nearly zero, Very unlikely, Slightly unlikely, No opinion, Slightly likely, Very likely, Nearly certain]." In Figure 2, we assess the effect of the level question on responses to the change question. Democrats who answered the level question first reported *less* negative attitude change (0.24 scale points, SE: 0.10). By contrast, Republicans reported *more* negative attitude change (-0.24 points, SE: 0.14). Like a valve blowing off steam, asking the level question first appears to allow people to express their baseline attitudes separately from the changes in it.

Figure 2: Response substitution in the change format (Cornish example)



Note: The vertical axis displays the average direction of self-reported attitude change toward Tony Cornish, with "more" scored as a 1, "none" scored 0, and "less" scored -1. The horizontal axis indicates whether subjects were asked about the level of their attitude before they were asked about how their attitude changed. Error bars represent 95 percent confidence intervals.

The counterfactual format

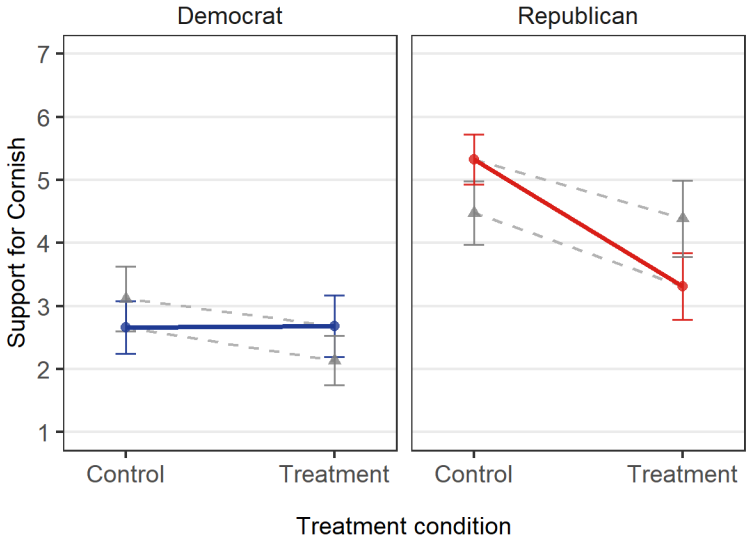
We now turn to the counterfactual format, which we tested on a separate, also randomly selected subset of our respondents. The first stage proceeded like a randomized experiment: while the treatment group read the entire vignette, the control group read a version that left out the information about sexual harrassment. Both groups then answered the level question about their support for Cornish.

In the second stage, subjects were asked to imagine how they would have responded had they been in the other treatment group. Control subjects saw a second slide that began, "Now we would like you to consider the same information about Cornish, plus some new information." After seeing shown the full vignette, they answered the level question again. Treatment subjects saw a second slide that began, "Suppose you had seen the same information, but without any mention of the fact that Cornish was accused of sexual misconduct." After seeing the vignette without the treatment information, they were asked the level question again. Table 3 displays the full sequence of treatments and questions.

Table 3: Counterfactual Format (Cornish example)

Control Group	Treatment Group
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 "doesn't excuse sexual harassment."
If Cornish were running for Congress in your district against a moderate Democrat, how likely would you be to support him? (1: Nearly Zero, 7: Nearly Certain)	If Cornish were running for Congress in your district against a moderate Democrat, how likely would you be to support him? (1: Nearly Zero, 7: Nearly Certain)
(New Slide) Now we would like you to consider the same information about Cornish, plus some new information.	(New Slide) Suppose you had seen the same information, but without any mention of the fact that Cornish was accused of sexual misconduct.
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 "doesn't excuse sexual harassment."	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.
If Cornish were running for Congress in your district against a moderate Democrat, how likely would you be to support him? (1: Nearly Zero, 7: Nearly Certain)	If Cornish were running for Congress in your district against a moderate Democrat, how likely would you be to support him? (1: Nearly Zero, 7: Nearly Certain)

Figure 3: Experiment and Counterfactual Format (Cornish example)



Note: This figure displays all of the data elicited in the counterfactual format. The horizontal axis indicates the respondent's treatment status in the experiment. The vertical axis displays the average level of support for Cornish. Dark colored circles represent the experimental outcomes elicited in the first stage. Light grey triangles represent the counterfactual guesses elicited in the second stage. The solid, dark colored lines represent the experimental estimate of the ATE. The dashed, light grey lines represent the average of subjects' self-reported treatment effects.

Together, the two stages produce an estimate of actual attitude change due to the information (the first stage experiment), as well as subjects' best guesses about how the information changed their attitudes. Figure 3 plots the data from both stages. The first stage estimates are dark colored dots; the counterfactual guesses, light grey triangles. According to the first stage experimental estimate, the information had a very small average effect among Democrats (0.02 scale points, SE: 0.32) and a large negative average effect among Republicans (-2.01 scale points, SE: 0.33). This is the opposite of what the change format implies. Cornish suffered a heavy loss of support among Republicans, not Democrats, as a consequence of the allegations.

Turning next to the ATE estimates derived from the counterfactual format, we find that on average, Democrats believed the information decreased their support for Cornish by 0.49 scale points (SE: 0.11). The corresponding figure for Republicans is 1.07 scale points (SE: 0.17). These estimates are interestingly biased away from the experimental bench-

mark. Democrats overstate their response to treatment whereas Republicans understate it. Democrats think they would respond negatively but they do not. Republicans are correct that they would respond negatively, but are wrong about how much.

Underlying these inaccurate self-reported treatment effects are inaccurate guesses about each of the counterfactual outcomes—a violation of Assumption 1. Treated Republicans under-estimated their control outcome (-0.85 scale points, SE: 0.31), while control Republicans overestimated their treatment outcome ($+1.08$ scale points, SE: 0.40). Among Democrats, the pattern was reversed. Treated Democrats overestimated their control outcome ($+0.45$ scale points, SE: 0.33) while control Democrats overestimated their treatment outcome (-0.55 scale points, SE: 0.31).

Despite this bias, the alternative format yields the more accurate substantive conclusion. The change format suggests that Cornish lost a lot of support among Democrats and none among Republicans—the opposite of the experiment. The counterfactual format is also biased, but it correctly suggests that Cornish lost more support among Republicans than among Democrats.

Research Design

To conduct a broad assessment of attitude change question formats, we tested 14 information treatments, the subject matter of which roughly matches the real world questions documented in Table 1. We drew subjects from three surveys conducted in 2018 by Lucid, which recruits online and quota samples to U.S. census demographic margins (Coppock and McClellan 2019). We also include one result from a survey conducted on Amazon Mechanical Turk (MTurk) in June 2019.

For each treatment, our approach was similar to the Tony Cornish example above. In Studies 1-3, subjects were randomly assigned to answer questions either in the change format or the counterfactual format. Study 1 ($N = 417$, refusal rate = 3.0%) included eight information treatments. The first five used a "candidate profile" format much like the Cornish

example. The remainder concerned issue attitudes and did not include control information. Study 2 ($N = 2,475$, refusal rate = 2.8%) evaluated six information treatments. Eligibility for two of the six, "Opposed Kavanaugh" or "Supported Kavanaugh," was determined based on the respondent's state.[3] Study 3 ($N = 1,110$, refusal rate = 3.2%) fielded four of Study 2's information treatments using binary outcomes rather than Likert scales.[4] Study 4 was the narrowest. The day after special counsel Robert Mueller made his first public comments about his investigation into Russian interference in the 2016 election, we included change questions in an otherwise unrelated MTurk survey ($N = 1,074$, refusal rate = 0.3%).

Table 4 summarizes the information treatments and lists the outcome questions. We split all of our analyses by political party (including leaners) because the biases associated with response substitution appear to be strongly correlated with party. For all of our analyses, we use ordinary least squares (OLS) regression to estimate means and differences-in-means. With two noted exceptions, we report HC2 robust standard errors.

[3]In each state, at least one Senator who was not retiring in 2018 voted with their party on Kavanaugh's confirmation. For states with two such Senators, we selected the Senator whose term expired soonest.

[4]The original motivation for switching to binary questions in Study 3 was to evaluate another format of our invention, "simultaneous outcomes." Because it performed poorly and only works with binary questions to begin with, we leave that analysis to the supplementary materials.

Table 4: Summary of Information Treatments and Outcome Questions

Topic	Study	Treatment summary	Level question	Change question
Blocked whistle-blower	1	Sen. Ricardo Lara (D-CA) blocked a bill protecting staffers making sexual misconduct allegations.	If Lara were running for Congress in your district against a moderate Republican, how likely would you be to support him?	Does the fact that Lara blocked a whistleblower protection bill make you more or less likely to support him in an election against a moderate Republican?
Disputed accusation	1	Rep. Tony Cornish (R-MN) was accused of sexual harrassment by Rep. Erin Quade (D-MN).	If Cornish were running for Congress in your district against a moderate Democrat, how likely would you be to support him?	Does the fact that Cornish was accused of sexual misconduct make you more or less likely to support him in an election against a moderate Democrat?
Endorsed Trump	1	Rep. Kevin Kelly (R-CT), a centrist, endorsed Donald Trump for president.	If Kelly were running for Congress in your district against a moderate Democrat, how likely would you be to support him?	Does the fact that Kelly endorsed Donald Trump make you more or less likely to support him in an election against a moderate Democrat?
Supports charters	1	Rep. Don Shooter (R-AZ) supports expanding charter schools.	If Shooter were running for Congress in your district against a moderate Democrat, how likely would you be to support him?	Does the fact that Shooter supported expanding charter schools make you more or less likely to support him in an election against a moderate Democrat?
Undisputed accusation	1	Rep. Dean Westlake (D-AK) faces severe sexual misconduct allegations.	If Westlake were running for Congress in your district against a moderate Republican, how likely would you be to support him?	Does the fact that Westlake was accused of sexual misconduct make you more or less likely to support him in an election against a moderate Republican?
Death penalty	1	States with and without the death penalty have very similar muder rate trends (with graphic).	How strongly do you support or oppose the death penalty?	Does the information make you more or less supportive of the death penalty?
Immigrant population	1	Immigrants will soon constitute their largest-ever share of the U.S. population (with graphic).	Do you support or oppose increasing the number of immigrants who can come to the United States?	Does the information make you more or less supportive of increasing the number of immigrants who can come to the United States?
Tax Cuts and Jobs Act	1	List of provisions in the Tax Cuts and Jobs Act.	How strongly do you support or oppose the Tax Cuts and Jobs Act, a law President Trump signed in December 2017?	Does the information make you more or less supportive of the Tax Cuts and Jobs Act, a law President Trump signed in December 2017?
Biden / Hill	2, 3	Joe Biden was skeptical of Anita Hill's sexual harrassment allegations against Clarence Thomas.	Do you support Joe Biden's possible run for president in 2020?	How does this change your support for Joe Biden's possible run for president in 2020?
DREAM Act	2, 3	The DREAM Act would make the economy grow by $30,000 per beneficiary.	Do you support or oppose the DREAM Act, which would allow unauthorized immigrants who were brought to the United States as children to apply for citizenship?	How does this change your support for the DREAM Act, which would allow unauthorized immigrants who were brought to the United States as children to apply for citizenship?
Kavanaugh	2	Respondent's Senator [opposed / supported] Brett Kavanaugh's nomination to the Supreme Court.	Will you support [last name] or [her / his] [Republican / Democratic] opponent?	How does this change your support for [last name] against [her / his] [Republican / Democratic] opponent?
Obama torture	2, 3	President Obama issued an executive order banning the CIA and other agencies from torturing detainees.	Do you support or oppose banning the CIA and other government organizations from torturing detainees?	How does this change your support for banning the CIA and other government organizations from torturing detainees?
Trump coal	2, 3	President Trump issued an executive order that reduced restrictions on coal ash disposal.	Do you support or oppose strict regulations on the disposal of coal ash, the pollutant left over after power plants burn coal?	How does this change your support for strict regulations on the disposal of coal ash, the pollutant left over after power plants burn coal?
Mueller comments	4	[No treatment. The response substitution test was part of an unrelated survey conducted just after public comments by Mueller.]	Is this statement true or false? Robert Mueller's report said there is "undeniable proof" that President Trump personally conspired with Russian agents to influence the 2016 election.	Did Robert Mueller's comments yesterday make you more or less likely to believe this statement? [Same statement.]

Note: All politicians named in Study 1 are real state legislators. At the time of the survey, Cornish and Westlake had recently resigned.

Results

Response Substitution in the Change Format

Our first survey included a series of tests for response substitution in the change format. We randomly assigned subjects to answer a change question (1) after answering a question about their level of support for the candidate or policy or (2) immediately after seeing the treatment information. This experiment answers the question, *does allowing people to express the level of their attitude reduce self-reports of change?*

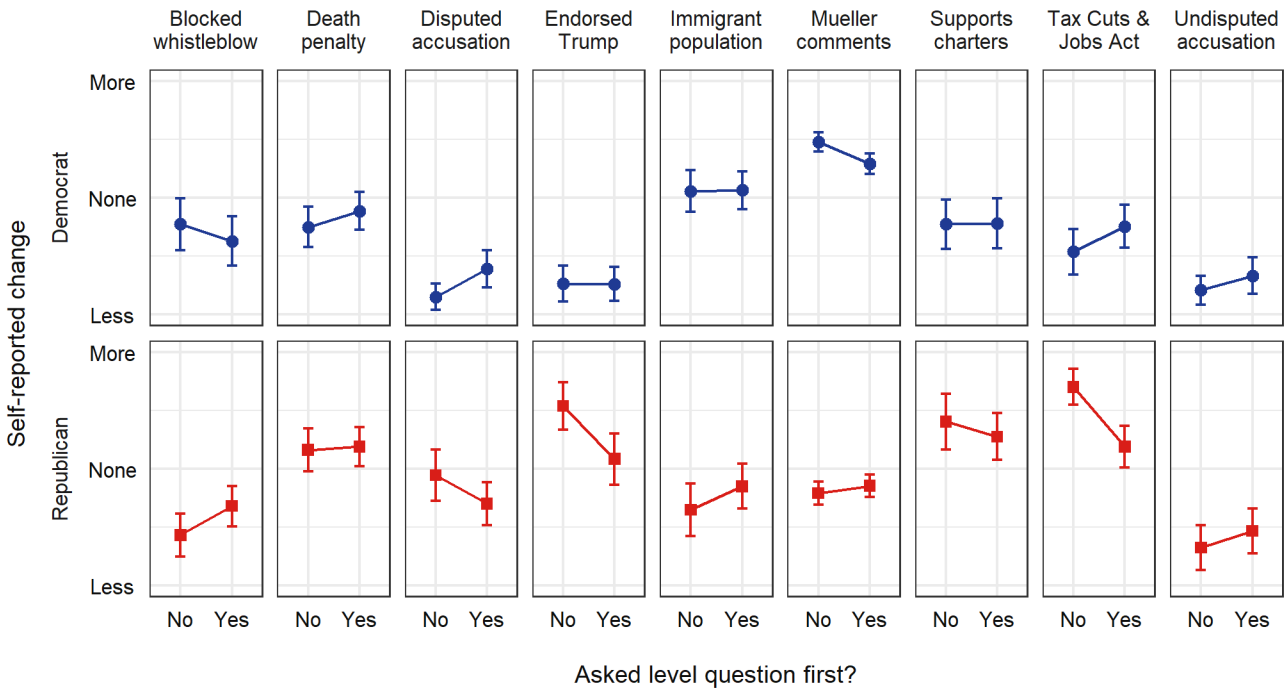
Overall, the "level question first" treatment reduced the probability of self-reporting attitude change by 10 percentage points (SE: 2 points).[5] In other words, giving people the opportunity to state the level of their attitude reduces claims of attitudes change. Splitting by party, we find evidence of an effect among Democrats (−14 points, SE: 3 points) and Republicans (−8 points, SE: 4 points), but not pure independents (−2 points, SE: 6 points). This pattern is consistent with the idea that they have less of the baseline attachment to parties and issue positions that, when interacted with response substitution, lead to misleading inferences.

Further support for the response substitution hypothesis comes from examining which types of self-reports of change were eliminated by the treatment. To aggregate change questions, we scored "more" as 1, "less" as -1, and "no change" as 0. This is mathematically equivalent to subtracting the percentage of respondents who said "less" from the percentage of subjects who said "more," which is the standard practice in media coverage that derives substantive conclusions from this question format. We refer to this as the *more minus less* estimator. Figure 4 plots it for each of the nine treatments for which we conducted this question format experiment.

The key pattern is that level questions reduced self-reports of change the most in settings where where Republicans and Democrats are especially likely to have divergent baseline

[5]The estimates in this paragraph pool across questions, meaning that each respondent contributed more than one observation. Consequently, we report clustered standard errors.

Figure 4: Effect of Level Question on Self-reports of Change, Studies 1 and 4.



Note: For each of nine experiments, this figure displays the mean of the change question, with "more" scored as a 1, "no change" scored 0, and "less" scored -1 (see text). The horizontal axis indicates whether subjects were asked about the level of their attitude before they were asked about how their attitude changed. The vertical axis indicates subjects' self-reported change in support for a candidate, support for a policy proposal, or belief in a factual statement. The error bars represent 95 percent confidence intervals.

attitudes. For example, the *Disputed accusation* facet reproduces the Cornish example from above. Democrats become less likely to claim that the accusation harms their already low support for the Republican candidate and Republicans admit that the accusation hurts their already high support for him. In the *Endorsed Trump* facet, Republicans given the opportunity to give the level of their opinion first become less likely to claim that their support Kevin C. Kelly, a moderate Republican from Connecticut, increased because he endorsed Donald Trump. In the *Muller comments* facet, Democrats became less likely to claim that special counsel Robert Mueller's comments made them believe Donald Trump had personally colluded with Russian agents. Finally, the *Tax Cuts and Jobs Act* facet shows that Republicans became less likely to claim that a short summary of its provisions made them more supportive of the law. Democrats become less likely to claim that the same information

made them less supportive.

Evaluating the Counterfactual Format

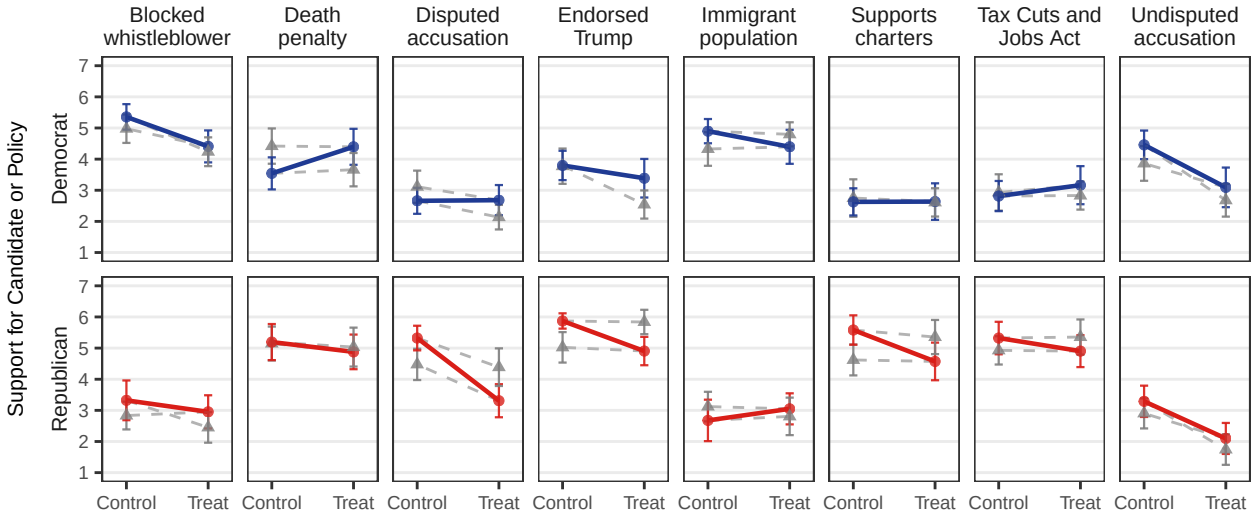
Across our studies, we tested the counterfactual format using 14 information treatments, four of which appeared in both Study 2 and Study 3. Figure 5 plots the mean of all of the outcome measures elicited in this format. The experimental outcomes elicited in the first stage appear as dark colored circles, connected by a solid line representing the ATE. The counterfactual guesses elicited in the second stage appear as light gray triangles, connected by light gray dotted lines representing the self-reported ATEs within each treatment group. If subjects made perfectly accurate guesses about counterfactuals, the solid and dotted lines would overlap almost perfectly, only differing due to sampling error.

As it happens, the counterfactual guesses often appear fairly accurate, as indicated by the fact that many of the circles, triangles, and connecting lines tend to be very close together. Yet many cases show substantial differences between the experimental outcomes and the counterfactual guesses.

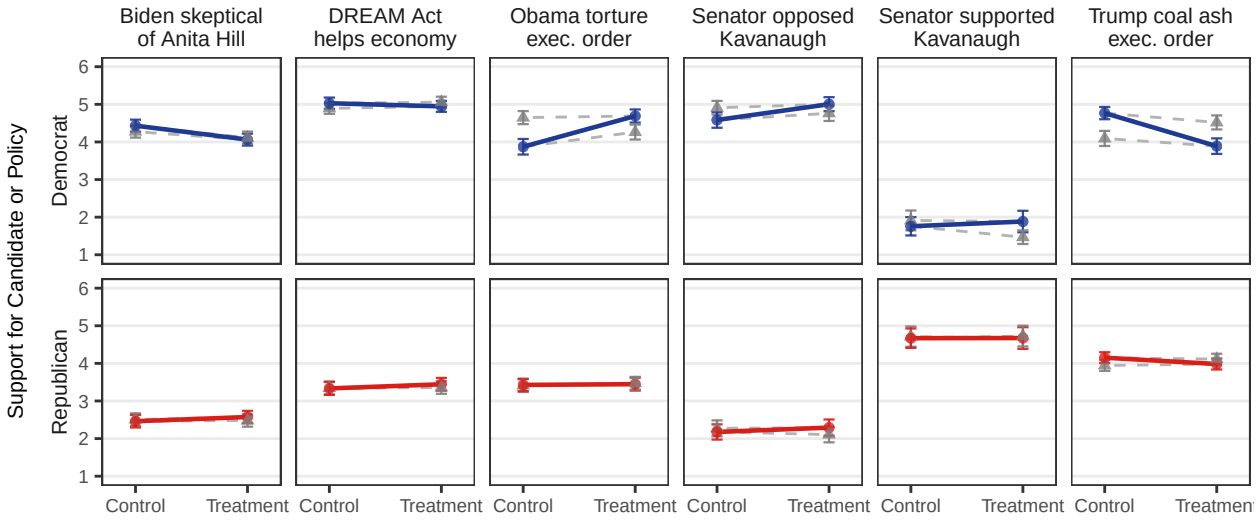
We begin by testing Assumption 1, the assumption that people can accurately state the counterfactual outcome. Across our studies, we have 72 opportunities to compare an actual outcome to a counterfactual guess (18 experiments $\times$ two parties $\times$ two outcomes). Of these, difference-in-means tests reject the null hypothesis of no difference in 17 cases (23.6%) at the $p \leq 0.05$ level. Another look comes from a comparison of the experimental and self-reported ATEs. Since these two estimates are not independent of one another, we bootstrapped our estimates and computed 95% confidence intervals using the percentile method. Out of 36 comparisons (18 experiments $\times$ two parties), the 95% confidence interval excluded zero in 9 cases (25%). Both approaches find that direct evidence against Assumption 1 in about a quarter of the cases. The supplementary materials present complete results of these tests.

In each case in which the experimental and counterfactual estimates differ, subjects' failure to recognize attitude change was the culprit. Democrats failed to recognize that their

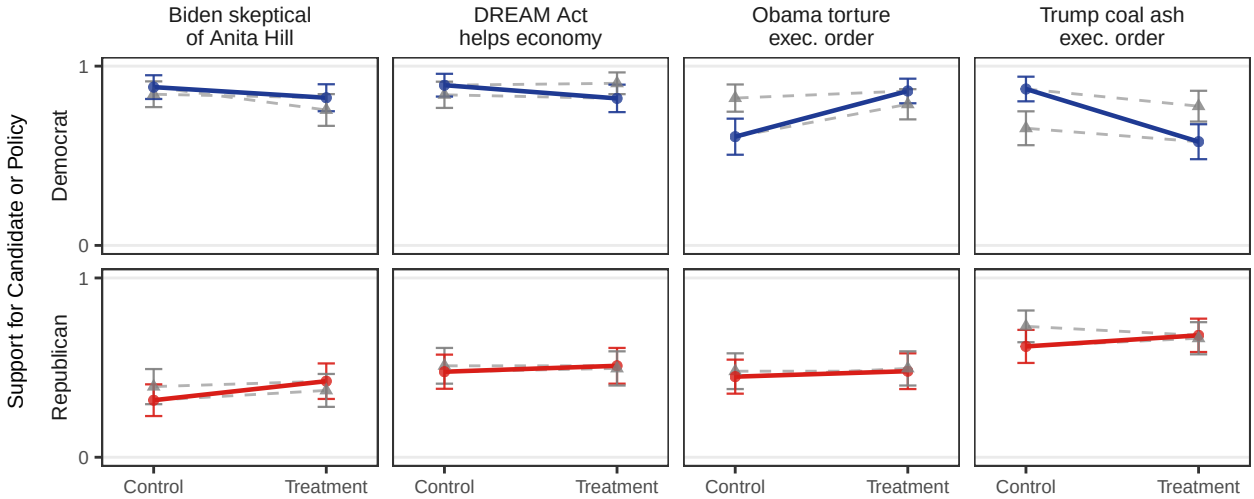
Figure 5: Experiment versus Counterfactual Format



(a) Study 1



(b) Study 2



(c) Study 3

Note: For each of information treatment we tested using the counterfactual format, this figure plots the same information as Figure 3.

support for an executive order banning torture increased when the *Obama torture* treatment revealed that President Obama issued such an order (studies 2 and 3); that the *opposed Kavanaugh* treatment made them more supportive of their Senator (study 2); and that the *death penalty* treatment made them more supportive of the death penalty. Republicans failed to recognize that their support for an executive order easing restrictions on coal ash disposal increased when the *Trump coal* treatment revealed that President Trump issued such an order (studies 2 and 3); that the *supports charters* treatment reduced their support for a Republican candidate; and that the *endorsed Trump* treatment reduced their support for a moderate Republican candidate. As in the example above, Republicans also underestimated the extent to which the *disputed accusation* treatment reduced their support for Cornish.

The bias in the counterfactual format appears to be drawing treatment effects towards zero. One reasonable explanation is that subjects want to maintain consistency between the first and second responses they give. They anchor their second response to their first, and adjust it too little when considering the counterfactual. In short, subjects using the counterfactual format may sometimes deny attitude change that did occur, but are unlikely to raise "false alarms" that imply a lot of attitude change where none occurred.

Which Format is More Accurate?

So far, we have examined the change and counterfactual formats separately. In this final section of results, we combine data from the previous two sections. Using the first stage experiment as an estimate of true attitude change, we argue that the counterfactual format is more accurate than the change format.

To guide our analysis, Figure 6 presents a side-by-side comparison of the results from the two question formats. Under the name of each treatment, the colored horizontal bars plot the distribution of self-reported change using both the change format and the counterfactual format. The triangles on top of the bars represent the more minus less estimator, always on a [-1, 1] scale. Below each set of colored bars is a white box containing two average

treatment effect estimates: the subjects' best guess from the counterfactual format, and the estimate of true attitude change from the experiment. In each box, the X-axis covers half the possible range of the ATE.[6] The separation of the colored bars and the white boxes captures a core analytical challenge, elaborated above, that stems from the properties of the change format. Because the change format is unitless, it cannot be placed on the same numerical scale an average causal effect from an experiment. Consequently, the analysis below is based on a comparison of the substantive conclusions that arise from each approach to measuring attitude change.

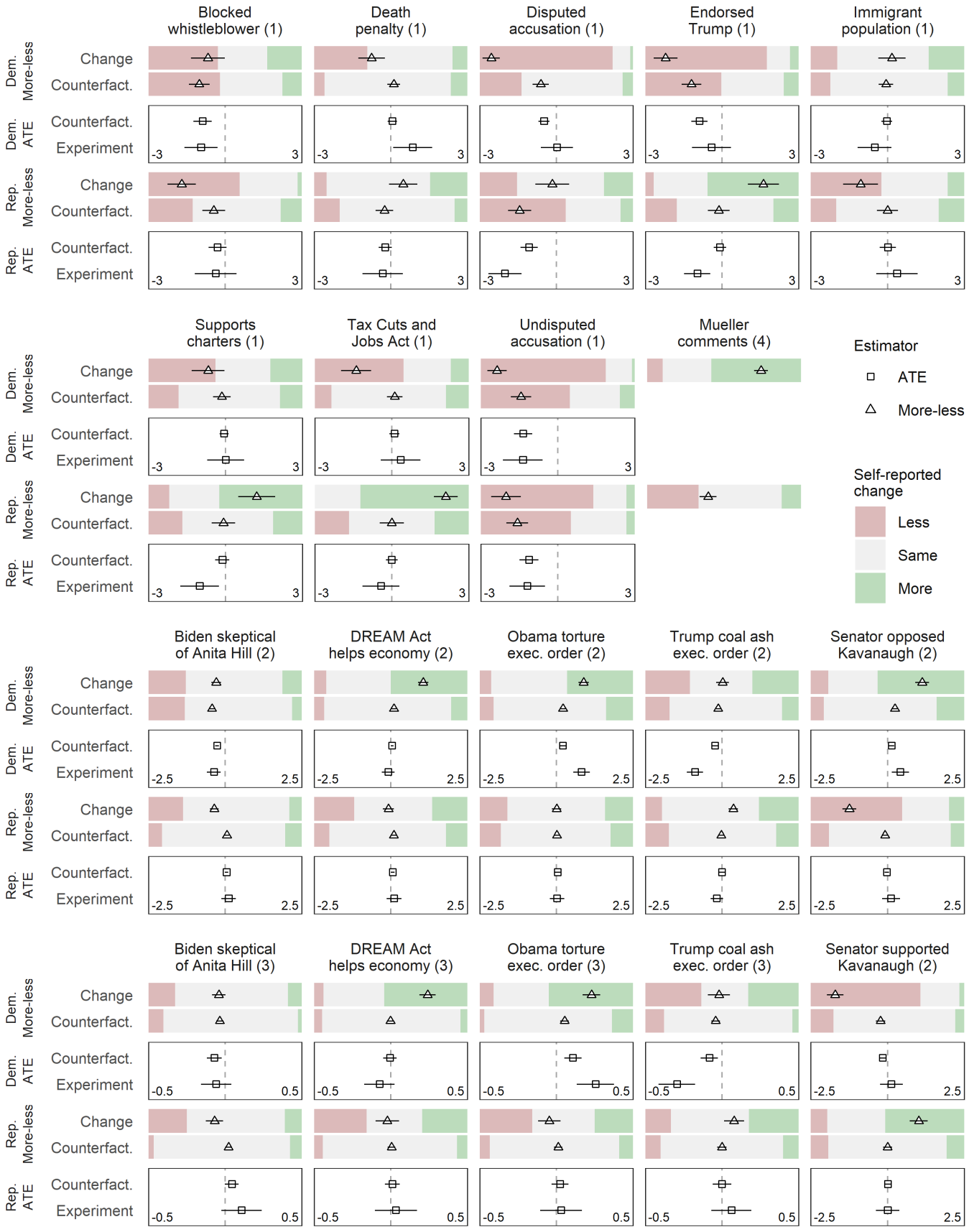
In most cases, the counterfactual format provides more accurate inferences than the change format. In one set of cases, the change format suggests overwhelming attitude change that is not present in the experiment. The starkest example comes from the *Tax Cuts and Jobs Act* treatment, a bulleted list of provisions contained in the tax reform law President Trump signed in 2017. Democrats overwhelmingly report that information made them less supportive of the tax cuts; by an even larger margin, Republicans report the opposite. In the experiment, the point estimates suggest a small effect that we could not distinguish from zero in either party. Five other treatments produced a similar pattern of results: *endorsed Trump*, *Senator opposed Kavanaugh*, *Senator supported Kavanaugh*, *blocked whistleblower* (among Republicans), and *DREAM Act* (among Democrats).

In a second set of cases, the change format elicits self-reports of attitude change that are less overwhelming, but imply a different conclusion about patterns of attitude change than the counterfactual format and the experiment. When asked how Joe Biden's past skepticism of Anita Hill's sexual harassment allegations against Clarence Thomas changed their support for Biden's presidential candidacy,[7] Republicans in both studies 2 and 3 were more likely to say "less supportive" than "more supportive." In the experiment, the point

[6]For example, Study 1 used 7-point scales. The largest possible ATE was 6 and the smallest possible was -6. In Figure 6, the scale ranges from -3 to 3, which covers the middle half of the possible range.

[7]Our study took place a few months before a wave of media attention to Biden's handling of Hill's allegations. Subsequently, pollsters asked self-reported attitude change questions similar to the ones we fielded. See Emerson Polling, "Nevada 2020: Biden and Sanders lead Democratic Primary Field," March 31, 2019.

Figure 6: Comparison of change format, counterfactual format, and experiment.



Note: For a description of this figure, see the text.

estimates were positive and borderline statistically significant, suggesting that many Republicans' "less supportive" self-reports were really statements of their baseline opposition to Biden. The other four cases in which the change format suggested the opposite sign as the experiment were *death penalty* (among Democrats), *immigrant population* (among Republicans), *supports charters* (among Republicans), and the Tony Cornish example, *disputed accusation*.

In some of the cases just mentioned, our evaluation of the counterfactual format identified a misleading inference. In these cases, it is not that the counterfactual format is perfectly accurate—it is just more accurate than the change format. The *endorsed Trump* experiment, for example, suggests that the average Republican was more supportive of state Sen. Kevin Kelly (R-CT) when his allegiance to President Trump was unknown than when it was revealed that he had endorsed President Trump. The counterfactual format did not detect this preference. On average, Republican respondents using this format reported no attitude change. But the change format was even less accurate, with a majority of Republicans reporting that Kelly's endorsement of Trump made them more supportive of Kelly.

For one information treatment, the change format yielded more accurate inferences than the counterfactual format. In the case of the *Obama torture* treatment, Democrats answering the change question were much more likely to say that Obama's involvement made them more supportive of the executive order. By contrast, Democrats using the counterfactual format reported little change on average. The experiment yielded positive ATEs that were, in both studies 2 and 3, larger than the ATEs implied by respondents' counterfactual guesses.

Finally, there were two treatments for which we can make no clear judgment as to which format yields more accurate inferences. On the *Trump coal* treatment, neither self-report format anticipated the negative ATE among Democrats, and Republicans' greater tendency to say "more" than "less" to the change question is only modestly inconsistent with the experiment. In the *undisputed accusation* format, respondents answering the counterfactual

format are right on the money with their guess of the ATE. However, the change format also suggests that strong negative reactions to the allegations against former state Rep. Dean Westlake (D-AK).

In summary, our fourteen information treatments produced six cases in which the change format is less accurate than the counterfactual format because it exaggerates attitude change; five cases in which the change format gets the sign wrong, yields a misleading comparison of Democrats and Republicans, or both; one case in which the change format is more accurate than the counterfactual format; and two cases in which the comparison of the two formats is indeterminate.

Discussion

The first and most important goal of this paper was to evaluate the standard approach to asking about attitude change. We confirmed the hypothesis that change questions are susceptible to response substitution. When asked how their attitudes changed, many people will use the question to state the level of their attitude instead. Our first test of the response substitution theory was a question order experiment in which some subjects answered a change question immediately after seeing some information, while others were asked to state the level of their attitude first. We found that change questions are sensitive to whether level questions are asked immediately beforehand. For each information treatment that produced a substantial difference, the differences in self-report behavior were consistent with the response substitution story.

Our second goal was to propose an alternative. The counterfactual format improves upon the change format by asking subjects for the level of their opinion twice. In the control group, we first obtain subjects' untreated outcome, then we ask how they would have responded had they seen the treatment information. In the treatment group, we ask for the treated outcome first, then we ask how subjects would have responded had they not seen the treatment information. We demonstrated the counterfactual format's potential in

our first three studies, finding that it produces more accurate self-reports than the change format. Despite this success, our subjects were in some cases plainly wrong about what their counterfactual attitudes would have been. In particular, we think they tend to underestimate the causal effects of treatment, possibly because of a desire to maintain consistency between their first and second response.

Based on our theory and evidence, we recommend that the change format be abandoned altogether. For researchers who are interested in asking about individual-level causal effects, we recommend the counterfactual format as a starting point. Across a wide range of topic areas, the counterfactual format's results were more similar to a randomized experiment than were the change format's. Admittedly, this portion of the evaluation was less precise than we would have liked. As we have stressed, the intersection of two issues—the fundamental problem of causal inference and the change format's focus on the sign of τ_i—make it difficult to compare the quantities produced by the change format to any measure of true attitude change (e.g., an experimental estimate of the ATE). Inconvenient as it was for our evaluation, this challenge ultimately highlights another reason to prefer the counterfactual format, which is designed to be easier to validate.

For future applications of the counterfactual format, an important consideration is the possibility that the control group *already knows* the treatment information. When subjects are plausibly "pre-treated," the first-stage experimental comparison recovers the effect of being reminded of the treatment information rather than the effect of the information itself (Druckman and Leeper 2012; Slothuus 2016; Linos and Twist 2018). We endeavored to avoid this problem by focusing on obscure state legislators, little-known facts, and old controversies.[8] However, since survey researchers often want to examine recent events that already attracted widespread media coverage, this problem may be quite serious in other settings. In such circumstances, we recommend a "treatment-arm only" version of the counterfactual

[8]The exception to this rule was the Kavanaugh treatments, which we included because so many similar questions were being asked in real-world polls. The treatment on Biden's skepticism of Anita Hill's allegations would seem to fall into this category, but we fielded it before the recent round of media attention to this event.

format, in which subjects are reminded of the event and asked for the *level* of their support. In a follow-up question, they can be asked what their support would have been had the event not taken place. At a minimum, such a procedure would give subjects the "counterfactual assist" that can help overcome the response substitution problem.

For better or for worse, observers of politics frequently use questions about attitude change to learn about public reactions to information and events. We proposed a framework for thinking about what these questions attempt to measure, identified theoretical and empirical problems with the standard approach, and suggested an alternative strategy, asking about counterfactual outcomes, that gives more accurate answers. We hope this article alerts academics and practicioners to the problems with the standard about attitude change and sparks further development of alternatives.

References

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

Ballesteros, Carlos. 2017. "Evangelicals are more likely to vote for Roy Moore now because of sexual assault charges.".
URL: *https://www.newsweek.com/roy-more-sexual-assault-alabama-evangelicals-709015*

Bjorhus, Jennifer and J. Patrick Coolican. 2017. "Minnesota lawmaker, lobbyist accuse Rep. Tony Cornish of sexual harassment.".
URL: *http://www.startribune.com/two-women-including-fellow-state-lawmaker-accuse-rep-tony-cornish-of-sexual-harassment/456503633/*

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* 6(1):1–14.

Dahl, Robert. 1961. *Who Governs? Democracy and Power in an American City*. Yale University Press.

Druckman, James N. and Thomas J. Leeper. 2012. "Learning More from Political Communication Experiments: Pretreatment and Its Effects." *American Journal of Political Science* 56(4):875–896.

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.

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

Guess, Andrew M. 2015. "Measure for measure: An experimental test of online political media exposure." *Political Analysis* 23(1):59–75.

Holbrook, Allyson L. and Jon A. Krosnick. 2010. "Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique." *Public Opinion Quarterly* 74(1):37–67.

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.

Klein, Ezra. 2017. "The Roy Moore allegations and the media's crisis of trust.".
URL: *https://www.vox.com/policy-and-politics/2017/11/13/16642458/roy-moore-doug-jones-poll-alabama-senate*

Levine, Linda J. 1997. "Reconstructing memory for emotions." *Journal of Experimental Psychology: General* 126(2):165–177.

Linos, Katerina and Kimberly Twist. 2018. "Diverse Pre-Treatment Effects in Survey Experiments." *Journal of Experimental Political Science* 5(2):148–158.

Markus, Gregory B. 1986. "Stability and change in political attitudes: Observed, recalled, and 'explained'." *Political Behavior* 8(1):21–44.

Neyman, Jerzey. 1923. "On the application of probability theory to agricultural experiments." *Annals of Agricultural Sciences* 10:1–51.

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(4):883–898.

Page, Susan. 2009. "Poll: Health care views take sympathetic tilt.".
URL: *https://abcnews.go.com/amp/Politics/story?id=8316019&page=1*

Rubin, Donald B. 1974. "Estimating causal effects of treatments in randomized and non-randomized studies." *Journal of Educational Psychology* 66(5):688–701.

Schacter, Daniel L. 1999. "The Seven Sins of Memory: Insights From Psychology and Cognitive Neuroscience." *The American Psychologist* 54(3):182–203.

Slothuus, Rune. 2016. "Assessing the Influence of Political Parties on Public Opinion: The Challenge from Pretreatment Effects." *Political Communication* 33(2):302–327.

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

Tourangeau, Robert, Lance J. Rips and Kenneth Rasinski. 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* (4):325–343.

Wilson, Rick. 2017. "Roy Moore's GOP Stands for 'Grand Old Pedobears'.".

Yair, Omer and Gregory A. Huber. 2019. "How robust is evidence of perceptual partisan bias in survey responses? A new approach for studying expressive responding." *Unpublished Manuscript* .