ASKING ABOUT ATTITUDE CHANGE

MATTHEW H. GRAHAM*
ALEXANDER COPPOCK

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 report the level of their attitudes rather than the change in them. 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 not know the treatment information. Using a series of experiments embedded in four studies, we show that the counterfactual format greatly reduces bias relative to the change format.

In advance of Alabama's 2017 special election for US 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 percent responded "more likely," 38 percent "less likely," and 33 percent "no difference." Among self-identified

evangelical Christians, "more" outnumbered "less." Many commentators decried the apparent depravity of those whose support for Moore had seemingly increased because of the allegations (e.g., Ballesteros 2017; Wilson 2017).

Surveys often ask respondents to assess the causal effect of exposure to some news event on their attitudes. This article shows that *change questions* of this type exaggerate the extent to which information about events causes attitude change. This bias is at least partially explained by a phenomenon known as *response substitution* (Gal and Rucker 2011; Yair and Huber 2021), wherein respondents use the question to indicate the level of their opinion rather than the change 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; they were 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 percent of Alabama voters wanted to express support for Moore despite the scandal (e.g., Klein 2017). Though this interpretation has merit, 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, which can be measured more directly. This article focuses 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 for some political examples and the Supplementary Material, Section B, for the full set of examples, which also covers nonpolitical topics like sports, drug use, and consumer behavior). 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 affected attitudes, and how events changed social and economic behavior. Many other questions concerned sexual misconduct, including 15 questions about allegations against Supreme Court Justice Brett Kavanaugh. Despite their prevalence, we are not aware of methodological research that evaluates change questions.

This article has two main goals. The first is to explain why the change format generates biased inferences. We test the response substitution

---

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. Questions about attitude change 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*

(*continued*)

Table 1. (*continued*)

Category	Polls	Qs	Example
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.—These questions were compiled using two searches. First, a search of 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, a Google search for this same search string plus the word "poll." Both searches only considered polls conducted between January 1, 2017, and December 31, 2018. The Supplementary Material lists all of the questions.

explanation with a series of experiments in which some subjects are randomly assigned to state the absolute level of their opinion before answering the change question. The results show that answering the level question first substantially reduces the exaggeration in self-reported attitude change.

The second goal is to introduce an alternative, the *counterfactual format*. It proceeds in two steps. First, subjects are reminded of the treatment information and then asked to report the level of their attitude or opinion as in a standard survey question. Second, subjects are asked to imagine how they would have responded if they did not know the treatment information. Comparing the answers given by each subject yields each subject's beliefs about the causal effect of the treatment information.

The relative accuracy of the change and counterfactual formats is evaluated through comparisons to the difference-in-means estimate from a randomized experiment, which provides an unbiased estimate of the causal effect of information. Though neither format is perfect, we find that the counterfactual format consistently produces more accurate estimates of attitude change. This suggests that survey researchers who want to reduce bias in questions about attitude change should abandon change questions in favor of the counterfactual format.

To illustrate, we used the counterfactual format to estimate Americans' beliefs about the effect of the Ukraine revelations on support for the impeachment of Donald Trump. We obtained 4,034 survey responses (field dates: November 21 through December 10, 2019; cooperation rate: 97.7 percent) from Lucid, which quota samples online survey respondents to match

U.S. Census demographic margins. The survey reminded subjects of the treatment information by asking whether they had heard about "Donald Trump withholding military aid to Ukraine while urging the President of Ukraine to investigate allegations of corruption against the Biden family" and whether they thought Trump's actions were appropriate. It then asked subjects for their level of support for impeachment: "How strongly do you oppose or support the impeachment of Donald Trump? [1: Strongly oppose; 7: Strongly support]" The second step asked, "Imagine that you did not know about Donald Trump withholding military aid to Ukraine while urging the President of Ukraine to investigate allegations of corruption against the Biden family. How would you have answered the question: How strongly do you oppose or support the impeachment of Donald Trump?"

The results are presented in figure 1. Panel (a) plots the first question against the second, with a small amount random noise added to distinguish the points. Panel (a) offers prima facie evidence that subjects are not bewildered by the nonstandard question format: the strong correlation (0.82) of the two responses indicates that most subjects believe their counterfactual attitudes would have been very close to their actual attitudes (72 percent report exactly zero change). That said, some subjects do believe the allegations changed their support for impeachment, albeit by a small amount. The average difference between the first and second question is 0.16 (SE = 0.02). Panel (b) shows the average difference across partisan groups. For all groups except strong Republicans, the estimated effect of the Ukraine revelations on impeachment support is small, positive, and statistically significant. Subjects appear to believe that their support or opposition to impeachment is mainly due to other factors but that the Ukraine allegations did increase support ever so slightly.

The intuition behind this approach is that a "counterfactual assist" can help subjects think systematically about causal effects. The change format requires subjects to do three things at once: figure out their opinion, figure out what their opinion *would have been*, and take the sign of the difference. The counterfactual format simply splits these tasks apart, allowing respondents to take them one at a time.

We emphatically do not claim that the counterfactual format is an unbiased measure of individual-level causal effects. Nothing about the format guarantees that respondents' guesses of their counterfactual attitudes will be correct. Indeed, compared with experimental benchmarks, we find direct evidence that subjects are sometimes wrong about what their attitudes would have been. In our view, the counterfactual format should be interpreted as a measure of subject *beliefs* about causal effects.

Despite these difficulties, we recommend retiring change questions in favor of questions that ask respondents to imagine counterfactuals. This approach greatly reduces bias relative to experimental benchmarks.
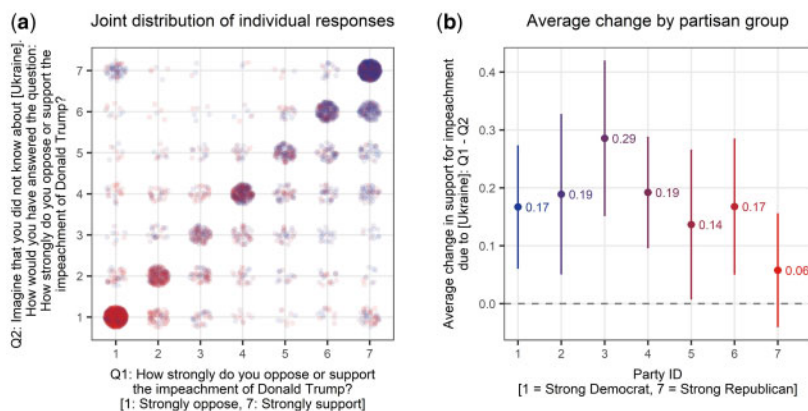
Figure 1. Counterfactual format example. The left panel plots the joint distribution of individual responses. The right panel plots the average change by partisan group.

Questions About Attitude Change

THE CHANGE FORMAT

The standard approach to asking about attitude change is the *change format*, which asks respondents whether some event or information made them more or less supportive of a candidate or policy proposal. For example, amid the healthcare reform debate in summer 2010, angry protestors disrupted town hall meetings across the country. To understand how exposure to the protests affected attitudes, 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 percent of respondents said "more sympathetic," compared with just 21 percent who said "less sympathetic." Republicans split 51 percent to 8 percent, while Democrats split 17 percent to 39 percent. 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). The causal language in the coverage appears to take the survey reports of attitude change at face value.

To formalize 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). The treated potential outcome,

$Y_i(D = 1)$ (or $Y_i(1)$ for short), is the attitude subject i would express if they were aware of the protests. The untreated potential outcome, $Y_i(0)$, is the attitude the subject would express if they were unaware of the protests. Change questions ask respondents to compare the two states of the world, then assess whether their treatment attitude is higher, lower, or the same as their untreated attitude. The difference is the individual-level treatment effect, $\tau_i \equiv Y_i(1) - Y_i(0)$. Effectively, change questions ask respondents to report the sign of their individual-level treatment effect, $sign(\tau_i)$.

Why might it be hard to measure this quantity? First, the fundamental problem of causal inference states that we can observe one potential outcome, but never both (Holland 1986). Though counterfactuals can never be observed, they can sometimes be imputed with reasonable accuracy. Consider a skydiver with a parachute who survives a fall from a plane: $Y_i(\text{parachute} = 1) = $ alive. One cannot be certain as to how the skydiver would have fared without a parachute, but $Y_i(\text{parachute} = 0) = $ dead is a reasonable imputation based on a mix of theory, intuition, and time-series evidence (for related discussion, see Smith and Pell 2003; Yeh et al. 2018).

Imputing counterfactual attitudes may be tougher. Longitudinal work indicates that people tend to misremember their past attitudes as being more consistent with their present attitudes than they really were (Levine 1997; Markus 1986; McFarland and Ross 1987; Wilson, Houston, and Meyers 1998; Schacter 1999) and tend to exaggerate the extent of favorable changes in their personality over time (Conway and Ross 1984; McFarland and Alvaro 2000; Wilson and Ross 2001; Schryer and Ross 2012). In a review, Brinol and Petty (2012) conclude that "people can err in either direction—seeing no change in their attitudes or themselves when there actually has been change, and seeing some change where there actually has been none" (p. 168).

A second reason to mistrust change questions is that respondents may not answer them as intended by survey researchers. In particular, respondents may use them to partially express the *level* of their attitude ($Y_i(1)$) rather than the change in it (τ_i). Then–White House adviser David Axelrod used a version of this argument to critique *USA Today*'s interpretation of the healthcare 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 healthcare reform. Under this interpretation, change questions are vulnerable to *response substitution*, or survey subjects' tendency to sometimes answer a different question

from the one the researcher has asked. Gal and Rucker (2011) use the example of a person who liked a restaurant'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. 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 2021). These patterns are consistent with other findings that respondents exclude information that was measured in preceding questions (Schuman and Presser 1981; Strack and Martin 1987; Sudman, Bradburn, and Schwarz 1996; Tourangeau, Rips, and Rasinski 2000).

When it comes to questions about attitude change, response substitution entails using a question about change (did hearing about the protests change your attitude?) to state the level of one's attitude (I support/oppose healthcare reform). Accordingly, response substitution should usually bias attitude change reports away from zero. If true, the response substitution hypothesis could explain why Democrats tended to say that they became more supportive of healthcare reform while Republicans tended to say they became less supportive. To test for response substitution, we conduct experiments similar in spirit to Gal and Rucker (2011) and Yair and Huber (2021). In our experiments, some subjects are offered 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.

THE COUNTERFACTUAL FORMAT

We propose the *counterfactual format* as an alternative method for asking about attitude change. We present two versions of the format that depend on whether subjects have been "pretreated" with the treatment information under study (Druckman and Leeper 2012; Slothuus 2016; Linos and Twist 2018). When the research goal is to learn about the effects of an event that has already occurred (a common setting for public-facing pollsters), researchers can use a nonrandomized counterfactual format analogous to the impeachment example above. By contrast, the randomized counterfactual format may be a superior option in settings in which subjects have not been pretreated—or, as in our case, when researchers want an unbiased experimental benchmark against which to evaluate other methods of asking about attitude change.

The *nonrandomized counterfactual format* proceeds as follows. Subjects first receive a prompt that reminds them of the treatment information.[2] They are then asked for the level of their opinion, $Y_i(1)$. In the second stage, subjects are asked to guess what their untreated potential outcome would have been if they did not know the treatment information. We refer to the true untreated potential outcome as $Y_i(0)$ and to subjects' best guess of their untreated outcome as $\tilde{Y}_i(0)$. The tilde indicates a guess about a counterfactual state of the world.

The counterfactual format estimate of subject beliefs about the causal effect of information on their attitude is then $\tilde{\tau}_i = Y_i(1) - \tilde{Y}_i(0)$. In order to interpret this causal effect estimate ($\tilde{\tau}_i$) as being equal to the true causal effect ($\tau_i \equiv Y_i(1) - Y_i(0)$), one would need to believe that subjects can accurately guess their counterfactual attitudes, that is, that $\tilde{Y}_i(0) = Y_i(0)$. When this condition does not hold, $\tilde{\tau}_i$ remains an estimate of subject *beliefs* about causal effects, which may be of interest to survey researchers regardless of whether those beliefs are accurate.

When subjects have not been pretreated, one could reverse the process: solicit the level of each subject's untreated opinion ($Y_i(0)$), expose them to the treatment information, then ask how they would have responded had they known the treatment information ($\tilde{Y}_i(1)$). This "control-first" version is similar to some common within-subject designs.

However, if subjects have not been pretreated, researchers can do better than the control-first, within-subject design. Instead, they can use the *randomized counterfactual format* to take advantage of researcher control over exposure to information. In the first stage, all subjects participate in a standard two-arm randomized experiment in which m of N subjects are assigned to see the treatment information and the remaining $(N - m)$ control subjects are not. Both groups then report their attitudes, $Y_i(1)$ for the treatment group and $Y_i(0)$ for the control group. Using these data, one can directly estimate the average treatment effect (ATE) using a standard approach like the difference-in-means estimator: $\hat{ATE}_{DIM} = \frac{\sum_{i=1}^{m} Y_i(1)}{m} - \frac{\sum_{i=m+1}^{N} Y_i(0)}{N-m}$. Under typical random assignment procedures, this estimator is unbiased for the ATE (for a textbook proof, see Gerber and Green 2012, chapter 2). In the second stage, subjects imagine what their response would have been in the other treatment condition. This entails exposing control group subjects to the treatment information and asking the outcome question a second time. As in the nonrandomized version, treatment group subjects are asked to imagine they did not know the treatment information before reporting their attitudes a second time.

---

2. This reminder may have a priming or otherwise influencing effect on attitudes, so survey researchers should follow good design practices (e.g., neutral question wording) to ensure that the reminder avoids "pushing" subjects to one response or another.

This procedure is formalized as follows. Control group subjects first report their untreated potential outcome, $Y_i(0)$, then their guess of their treated potential outcome, $\tilde{Y}_i(1)$.[3] Treatment group subjects first report their treated potential outcome, $Y_i(1)$, followed by their guess of their untreated potential outcome, $\tilde{Y}_i(0)$. For subjects in the control group, the individual-level treatment effect estimates $(\tilde{\tau}_i|D = 0)$ is equal to $\tilde{Y}_i(1) - Y_i(0)$. For treatment group subjects, the estimate is $(\tilde{\tau}_i|D = 1) = Y_i(1) - \tilde{Y}_i(0)$. The value of $\tilde{\tau}_i$ is conditioned on the treatment assignment D to emphasize that $\tilde{\tau}_i$ could be affected by treatment. The counterfactual format estimate of the ATE is then $\hat{ATE}_{CF} = \frac{\sum_{i=1}^{N} \tilde{\tau}_i}{N}$, averaging over both the treatment and control groups. If subjects' guesses about their counterfactual attitudes are correct, then for all subjects, $\tilde{Y}_i(1) = Y_i(1)$, $\tilde{Y}_i(0) = Y_i(0)$, and $(\tilde{\tau}_i|D = 0) = (\tilde{\tau}_i|D = 1) = \tau_i$.

The randomized counterfactual format embeds a direct test of whether subjects' guesses are accurate. One can directly compare distributions of $Y_i(0)$, $\tilde{Y}_i(1)$, and $\tilde{\tau}_i|D = 0$ in the control group to the distributions of $\tilde{Y}_i(0)$, $Y_i(1)$, and $\tilde{\tau}_i|D = 1$ in the treatment group using standard hypothesis testing procedures like the two-sample *t*-test. Rejection of the null hypothesis of no difference is direct evidence that at least some subjects' guesses are incorrect. In contrast to the change format, a major virtue of the randomized counterfactual format is the built-in test of the accuracy of subject beliefs.

As we have emphasized, counterfactuals are difficult to imagine and subjects' guesses may for many reasons be incorrect. When they are, the results of the counterfactual format will be biased away from the true causal effects. We hasten to add, however, that any biases that attend to the counterfactual format will apply all the more strongly to the change format, which asks subjects to report the sign of the difference between two counterfactuals. Our results suggest that though the counterfactual format is sometimes inaccurate, the change format is substantially less accurate.

Detailed Example

We illustrate our approach using the example of Tony Cornish, a former Republican state legislator from Minnesota who was accused of sexual harassment. Because Cornish is not well known outside his district, subjects were unlikely to have known the treatment information prior to the study, which justifies the use of the randomized counterfactual format. We present three sets of results: a test for response substitution in the change format, a

---

3. To be fully explicit, this potential outcome could be written as $\tilde{Y}_i(1, D = 0, Y_i(0))$ to insist upon the idea that this counterfactual guess is made for subjects in the control group who have revealed their $Y_i(0)$. The fully explicit $\tilde{Y}_i(1)$ can be defined analogously.

direct comparison of the counterfactual format to the experimental benchmark, and an evaluation of the overall accuracy of each format.

THE CHANGE FORMAT

One random subset of survey respondents 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—the treatment—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*
>
> 1. *Fought against government waste and opposed the governor's plan to raise sales taxes.*
>
> 2. *Played a key role in crafting a new policy that allows county attorneys to carry handguns at work.*
>
> 3. *Increased prison sentences for car thiefs[4] 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]" The top row of figure 2 shows the distribution of responses, broken down by subject partisanship. A huge majority of Democratic respondents (87 percent) report that the accusation made them less likely to vote for Cornish. By contrast, most Republicans (57 percent) reported that the information had no effect. As in the Roy Moore example, these self-reports may be contaminated by response substitution: Democrats may have simply expressed their disapproval of Cornish while Republicans expressed their continued support despite the allegations.

REDUCING RESPONSE SUBSTITUTION

To test the claim that the self-reports are biased due to response substitution, a separate, also randomly selected group of respondents was offered the

---

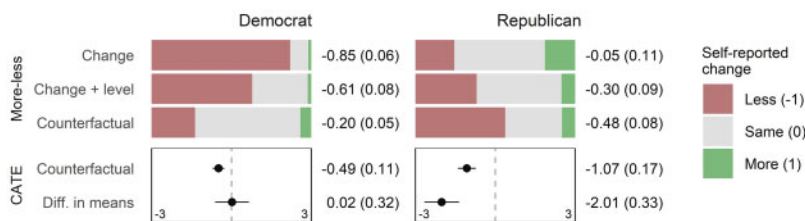4. This use of 'thiefs' was copied directly from Cornish's website.

Figure 2. Self-reported change (Cornish example).

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]." A comparison of the first two rows of figure 2 shows the effect of the level question on responses to the change question. Recoding the change question from -1 to 1, we see that Democrats score -0.85 on this measure, but answering the level question first increases their score to -0.61, for an increase of 0.24 (SE: 0.10). By contrast, the level question decreases this measure among Republicans (-0.24 points, SE: 0.14). Like a valve blowing off steam, asking the level question first appears to reduce subjects' tendency to use the change question to express the level of their attitudes.

EVALUATING THE COUNTERFACTUAL FORMAT

The counterfactual format was tested on yet another separate, also randomly selected subset of respondents. The first stage proceeded like a randomized experiment. The treatment group read the entire vignette, while the control group read a version that left out the accusation. 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 group. The third row of figure 2 shows that this procedure had an even larger effect on self-reports of change: less negative for Democrats, more negative for Republicans.

According to the unbiased difference-in-means estimate that uses first-stage responses only, the information had a very small average effect among Democrats (0.02 scale points, SE: 0.32, 7-point scale) and a large negative average effect among Republicans (-2.01 points, SE: 0.33). Directly contrary to the implications of the change format, the experimental estimate shows that Cornish suffered a heavy loss of support among Republicans, not Democrats, as a consequence of the allegations.

Among Democrats, the counterfactual format estimate that uses responses from both stages is -0.49 scale points (SE: 0.11) and the corresponding

figure for Republicans is -1.07 points (SE: 0.17). As demonstrated in the bottom two rows of figure 2, the counterfactual ATE estimates are biased away from the experimental benchmarks: Democrats overstate negative change by approximately half a point and Republicans understate it by almost a full point.

The main takeaways from this example are that the change format is badly biased and that the counterfactual format represents a large improvement. Subjects still make mistakes when using the counterfactual format—causal inference is hard!—but they express substantially more accurate beliefs about how the information changed their attitudes.

Research Design

For a broader look at the performance of the two question formats, the full research design evaluated 11 total information treatments using the strategies described just above in the detailed example. Study 1 conducted the "reducing response substitution" experiments using eight information treatments. Study 3 applied this strategy to one additional treatment. Using the same eight treatments, Study 1 also evaluated the counterfactual format using the same strategies described above. Study 2 applied these strategies to two additional treatments. Table 2 summarizes all 11 treatments and their corresponding outcome variables.

Our evaluation focuses on treatments for which it is credible to assume that that respondents were not pretreated by the information. This enables the use of the *randomized* counterfactual format, which includes a built-in randomized experiment. If the counterfactual format outperforms the change format in such settings, it is also likely to outperform the change format in settings in which respondents are pretreated. The last set of results applies the nonrandomized counterfactual format to four cases in which we suspect pretreatment.

The empirical analysis is based on four total surveys.[5] The first three were conducted using Lucid, which produces treatment effects that are similar to other commonly used online platforms (Coppock and McClellan 2019). Study 1 was conducted May 8–9, 2018 ($N = 417$, cooperation rate = 97.0 percent), and included eight of the information treatments described in table 2. Study 2a was conducted October 19–31, 2018 ($N = 2,475$, cooperation rate = 97.2 percent), and tests two of the treatments used in our evaluation (table 2), plus four additional treatments that are used to demonstrate the nonrandomized counterfactual format (table 2).

---

5. The impeachment example in the introduction comes from an unrelated fifth survey, also conducted on Lucid.

Table 2. 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? [1: Nearly zero; 7: Nearly certain]	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?
Death penalty	1	States with and without the death penalty have very similar murder rate trends (with graphic).	How strongly do you support or oppose the death penalty? [1: Strongly oppose; 7: Strongly support]	Does the information make you more or less supportive of the death penalty?
Disputed accusation	1	Rep. Tony Cornish (R-MN) was accused of sexual harassment 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? [1: Nearly zero; 7: Nearly certain]	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? [1: Nearly zero; 7: Nearly certain]	Does the fact that Kelly endorsed Donald Trump make you more or less likely to support him in an election against a moderate Democrat?

Table 2. (*continued*)

Topic	Study	Treatment summary	Level question	Change question
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? [1: Strongly oppose; 7: Strongly support]	Does the information make you more or less supportive of increasing the number of immigrants who can come to the United States?
Mueller comments	3	[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. [0: Definitely false; 100: Definitely true]	Did Robert Mueller's comments yesterday make you more or less likely to believe this statement? [Same statement.]
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? [1: Nearly zero; 7: Nearly certain]	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?
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? [1: Strongly oppose; 7: Strongly support]	Does the information make you more or less supportive of the Tax Cuts and Jobs Act, a law President Trump signed in December 2017?

(*continued*)

Graham and Coppock

Table 2. (*continued*)

Topic	Study	Treatment summary	Level question	Change question
Undisputed accusation	1	Rep. Dean Westlake (D-AK) faces several sexual misconduct allegations and has a long history of misconduct.	If Westlake were running for Congress in your district against a moderate Republican, how likely would you be to support him? [1: Nearly zero; 7: Nearly certain]	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?
Obama torture	2a, 2b	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? [1: Definitely oppose; 6: Definitely support]	How does this change your support for banning the CIA and other government organizations from torturing detainees?
Trump coal	2a, 2b	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? [1: Definitely oppose; 6: Definitely support]	How does this change your support for strict regulations on the disposal of coal ash, the pollutant left over after power plants burn coal?

NOTE.—From left to right, this table lists each treatment, the study in which it appeared, a paraphrase of the treatment question, the full text of the level question, and the full text of the change question. The counterfactual format consists of two level questions, one of which includes a preface asking the respondent to suppose they did not know the treatment information. The full text of all questions is available in the Supplementary Material.

Study 2b ($N = 1,110$, cooperation rate = 96.8 percent), which was conducted November 20–December 7, 2018, replicated four of Study 2a's treatments using binary outcomes rather than Likert scales. As the results are very close to Study 2a, the results of Study 2b are presented in the Supplementary Material, Section E. That survey also included a pilot test of an additional question format–the *simultaneous outcomes format*–that appeared to underperform even the change format (see the Supplementary Material, Sections E and F).

Study 3 was the most limited in scope. On May 28, 2019, the day after special counsel Robert Mueller made his first public comments about the investigation into Russian interference in the 2016 election, we included change questions in an otherwise unrelated survey conducted on Amazon Mechanical Turk ($N = 1,074$, cooperation rate = 99.7 percent). Study 3 did not include the counterfactual format.

All of the analysis is split by political party (including leaners) because the biases associated with response substitution appear to be strongly correlated with party. Consequently, treatment effect estimates are referred to as Conditional Average Treatment Effects (CATEs), conditioning on respondent partisanship. All of the analyses use ordinary least squares (OLS) regression to estimate means and differences-in-means. With two noted exceptions, heteroscadasticity consistent (HC2) robust standard errors are reported.

Results

RESPONSE SUBSTITUTION IN THE CHANGE FORMAT

Studies 1 and 3 included a series of experiments that test our claim that the change format is biased due to response substitution. Subjects were randomly assigned to answer a change question either immediately after seeing the treatment information or after answering a level question. If allowing people to express the level of their attitude reduces self-reports of change, we will infer that response substitution is a meaningful explanation for the bias in the change format.

As shown in figure 3a, asking the level question first greatly reduces self-reports of attitude change. Overall, the effect is a 10 percentage point decrease in reporting any change (SE: 2 points). The effect is larger among Democrats (−14 points, SE: 3 points) and Republicans (−8 points, SE: 4 points) than among pure independents (−2 points, SE: 6 points).[6] Response

---

6. Since respondents answered more than one question, the standard errors in this paragraph are clustered by respondent.
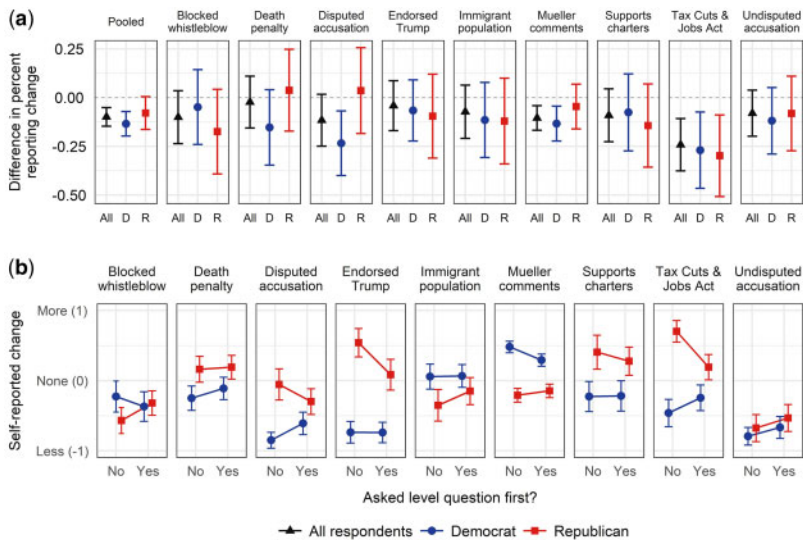
Figure 3. (a) Estimated effects of asking the level question first on reporting any attitude change. (b) Estimated effects of asking the level question first on the sign of self-reported attitude change.

substitution may be less severe among independents if they hold more moderate or ambivalent positions on issues that are polarized by party (though see Ahler and Broockman 2018).

Further support for the response substitution hypothesis comes from figure 3b. 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 a common procedure for analyzing the change format. We refer to this as the *more-minus-less* estimator.

The key pattern is that when level questions reduced self-reports of change, they tended to make Democrats' and Republicans' reactions to the information look more similar.[7] For example, in the *Endorsed Trump* facet, asking the level question first reduced Republican claims that their support of Kevin C. Kelly, a moderate Republican, increased because he endorsed Donald Trump. Similarly, 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.

---

7. This pattern is consistent with research that suggests that in most cases, different types of people respond to persuasive evidence by updating their beliefs in the same direction (Guess and Coppock 2018; Wood and Porter 2018).

In sum, figure 3 shows that change format responses are sensitive to whether a level question comes first, from which we infer that response substitution is a serious source of bias in the change format. The next section shows that the counterfactual format had an even larger effect on self-reports of attitude change, which suggests that the level-first treatment is not sufficient to fully purge the bias attending to the change format.

EVALUATING THE RANDOMIZED COUNTERFACTUAL FORMAT

This section evaluates the randomized counterfactual format's performance in estimating the effects of ten information treatments to which subjects had likely not been pre-exposed. Figure 4 presents the complete results, again split by party.

In terms of matching the substantive conclusions of the experimental benchmark, the counterfactual format easily outperforms the change format. The more-minus-less estimates derived from the change format have the opposite sign as the experimental difference-in-means estimates in 12 out of 20 opportunities, compared with 3 of 20 for the counterfactual format. In all three of those cases, neither the counterfactual estimate nor the difference in means estimate can be distinguished from zero. More often than not, the change format gets the *sign* wrong, whereas the counterfactual format does not.

A clear example of the relative performance of the two formats comes from the *Tax Cuts and Jobs Act* treatment, a bulleted list of provisions contained in the 2017 tax reform bill. In the change format, Democrats overwhelmingly report that the information made them less supportive of the tax cuts; by an even larger margin, Republicans report the opposite. This pattern is wholly contradicted by the experiment, which indicates small, nonsignificant effects in both parties. Asking a level question prior to the change question eliminated some of the bias, but not all. By contrast, the counterfactual format corresponds to the experimental benchmark very well in this case.

All told, this collection of tests yields 20 opportunities to compare the difference-in-means estimate to the counterfactual guess of the ATE (10 experiments × 2 parties). The difference between the two estimates was statistically significant in six cases (30 percent), indicating that subjects' guesses are sometimes but not uniformly correct.[8] Separating the treatment and control outcomes gives us 40 additional opportunities to evaluate subjects' performance (10 experiments × 2 parties × 2 potential outcomes). Of

8. Since the two estimates are not independent of one another, 95 percent confidence intervals are bootstrapped using the percentile method. The Supplementary Material presents complete results of these tests.
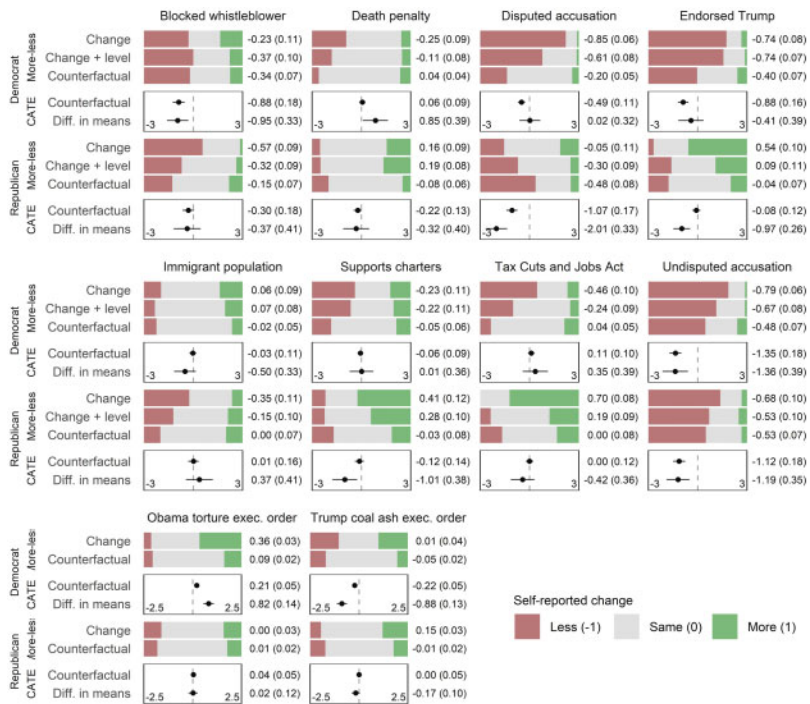
Figure 4. Comparison of change format, randomized counterfactual format, and experiment. This figure displays the full set of estimates for 10 treatments in studies 1 and 2. The horizontal bars represent the distribution of self-reported change. The white boxes contain CATE estimates derived from difference-in-means with first-stage responses and from the counterfactual format. Point estimates with robust standard errors are presented on the right-hand side of each facet.

these, difference-in-means tests reject the null hypothesis of no difference in 12 cases (30.0 percent).

THE NONRANDOMIZED COUNTERFACTUAL FORMAT

Many potential applications of the counterfactual format will likely arise in the wake of large political events that would have been hard to anticipate. In such cases, subjects are pretreated, meaning they cannot reveal their true untreated potential outcomes. Table 3 shows three examples in which it is likely that many subjects could have been exposed to the treatment information before the survey: former Vice President Joe Biden's skepticism of Anita Hill's allegations against Supreme Court Justice Clarence Thomas; the

Table 3. Summary of information treatments and outcome questions, pre-treated examples

Topic	Study	Treatment summary	Level question	Change question
Biden / Hill	2a, 2b	Joe Biden was skeptical of Anita Hill's sexual harassment allegations against Clarence Thomas.	Do you support Joe Biden's possible run for president in 2020? [1: Definitely oppose; 6: Definitely support]	How does this change your support for Joe Biden's possible run for president in 2020?
DREAM Act	2a, 2b	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? [1: Definitely oppose; 6: Definitely support]	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	2a, 2b	Respondent's Senator [opposed / supported] Brett Kavanaugh's nomination to the Supreme Court.	Will you support [last name] or [her / his] [Republican / Democratic] opponent? [1: Definitely oppose [last name]; 6: Definitely support [last name]]	How does this change your support for [last name] against [her / his] [Republican / Democratic] opponent?

projected economic benefits of the Development, Relief, and Education for Alien Minors (DREAM) Act; and whether the US Senator from the respondent's home state voted to confirm Brett Kavanaugh to the Supreme Court. We offer these as illustrative examples of what researchers can do to

estimate subject beliefs about causal effects when subjects may have been pretreated.

A familiar pattern emerges in figure 5. Relative to the change format, the counterfactual format greatly reduces self-reports of attitude change. In fact, to make the estimates visible at all, we shrank the scale for the counterfactual CATE to just 40 percent of the width used in figure 4.

Consider the impact that a Senator's vote on Brett Kavanaugh's confirmation might have had on US Senate elections in 2018, which was the topic of at least 15 change format questions in public-facing polls. According to the change format, voting for or against Kavanaugh has implausibly large electoral consequences for Senators, with large majorities of Democratic and Republican respondents saying it impacted their candidate preference. By contrast, the counterfactual format offers the more realistic conclusion that the effect was small, perhaps ever-so-slightly boosting the steadfastness of Democratic Senators' support within their own party.

In the case of Biden's skepticism of Hill's allegations, the difference across formats was also striking. The change format suggests that Biden's handling of the allegations cost him slightly more support among Republicans than among Democrats, while the counterfactual format suggests that any loss of support is concentrated among Democrats. The DREAM Act treatment produces a similar pattern: the change format suggests that the information immensely improves Democrats' already high support for the Act, whereas the counterfactual format suggests a small boost mainly for Republicans.
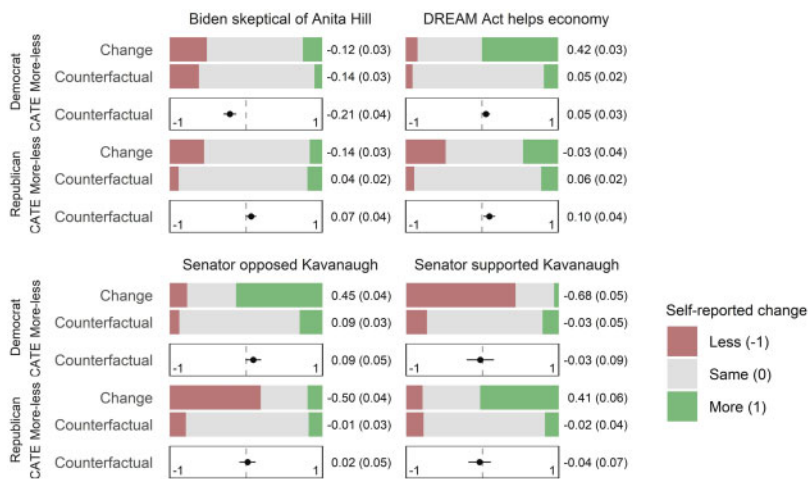

Figure 5. Comparison of change format and nonrandomized counterfactual format.

Because these survey subjects may have been pre-exposed to the treatments in each question, a comparison to an experimental benchmark is not possible. Nevertheless, the change format results are almost certainly not credible in these cases, whereas the counterfactual format produces plausible, useful information about subject beliefs about the effects of each treatment.

Discussion

The first goal of this article was to evaluate the standard approach to asking about attitude change. We confirmed that change questions are biased and that response substitution appears to be at least partially to blame. The main evidence in support of the latter claim is that change questions are sensitive to whether level questions are asked immediately beforehand.

The second goal was to propose an alternative. The counterfactual format improves upon the change format by inducing subjects to imagine the level of their attitudes in counterfactual worlds; the difference is a measure of subject beliefs about causal effects. Across 10 treatments, we found that the counterfactual format yields far more accurate estimates of the effect of information on attitudes. Even as we tout the counterfactual format's advantages relative to the change format, we have endeavored to be clear about its limitations. Because survey subjects can be wrong about what their responses would have been had things been different, we do not regard the difference between observed outcomes and the counterfactual guesses as true individual-level causal effects.

Together, the results suggest that relative to asking directly about attitude change, methods that ask respondents to explicitly imagine their counterfactual attitudes are more trustworthy. Depending on the context, researchers who want to ask about attitude change can choose between the nonrandomized and the randomized counterfactual formats. The nonrandomized version is appropriate when the research goal is to estimate the attitudinal effects of an event that has already occurred or information that has already been revealed. This involves first asking subjects to state the level of their attitude, then to imagine what their attitude would have been if they did not know about the event.

The randomized version is more appropriate when subjects have not encountered the treatment information before entering the study. This is likely to be the case for many of the questions in table 1; for example, many questions about attitude change ask how specific candidate positions would change support for a hypothetical candidate. In these settings, a standard treatment-versus-control comparison remains the best way to obtain an unbiased estimate of the ATE. In the short run, the randomized version is likely to be most useful as a template for evaluating the measurement properties of questions about attitude change.

In the longer term, the counterfactual format may hold promise as a tool for the estimation of heterogeneous treatment effects, which is usually plagued by large standard errors. At the cost of a single additional follow-up question, obtaining estimates of individual-level causal effects greatly reduces the variance in heterogeneous treatment effect estimates, at the cost of some bias. Rather than estimating highly imprecise treatment-by-covariate interaction terms, researchers could instead directly inspect how the subject-level treatment effect estimates covary with other subject-level characteristics, as in the estimates by partisanship in figure 1.

Survey researchers often want to learn the causal effects of events and information to which many subjects have already been exposed. Unable to conduct an experiment that would estimate the effect of such a "treatment," researchers often resort to question formats that directly ask subjects how their attitudes changed in response to information. Subjects have some self-knowledge that could be of use, but the change format attempts to extract this information in a clumsy way that yields misleading answers. A better way of asking about attitude is to give subjects a "counterfactual assist" that encourages them to imagine what their attitude might otherwise have been. We hope this article alerts researchers to the problems with the change format and sparks further development of alternatives rooted in counterfactual thinking.

Data Availability Statement

REPLICATION DATA AND DOCUMENTATION are available at https://doi.org/10.7910/DVN/GFF78K.

Supplementary Material

SUPPLEMENTARY MATERIAL may be found in the online version of this article: https://doi.org/10.1093/poq/nfab009.

References

Ahler, Douglas J., and David E. Broockman. 2018. "The Delegate Paradox: Why Polarized Politicians Can Represent Citizens Best." *Journal of Politics* 80(4):1117–33.

Ballesteros, Carlos. 2017. "Evangelicals Are More Likely to Vote for Roy Moore Now Because of Sexual Assault Charges." *Newsweek*, 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." *Minneapolis Star Tribune*, http://www.startribune.com/two-women-including-fellow-state-lawmaker-accuse-rep-tony-cornish-of-sexual-harassment/456503633/.

Brinol, Pablo, and Richard E. Petty. 2012. "Knowing Our Attitudes and How to Change Them." In *The Handbook of Self-Knowledge*, edited by Vazire Simine and Timothy D. Wilson, 157–80. New York: Guilford Press.

Conway, Michael and Michael Ross. 1984. "Getting What You Want by Revising What You Had." *Journal of Personality and Social Psychology* 47:738–48.

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*. New Haven: 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* 56:875–96.

Gal, David, and Derek D. Rucker. 2011. "Answering the Unasked Question: Response Substitution in Consumer Surveys." *Journal of Marketing Research (JMR)* 48:185–95. https://doi.org/10.1509/jmkr.48.1.185.

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

Guess, Andrew, and Alexander Coppock. 2020. "Does Counter-Attitudinal Information Cause Backlash? Results from Three Large Survey Experiments." *British Journal of Political Science* 50: 1517.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396):945–60.

Klein, Ezra. 2017. "The Roy Moore Allegations and the Media's Crisis of Trust." *Vox*, 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* 126:165–77.

Linos, Katerina, and Kimberly Twist. 2018. "Diverse Pre-Treatment Effects in Survey Experiments." *Journal of Experimental Political Science* 5:148–58.

Markus, Gregory B. 1986. "Stability and Change in Political Attitudes: Observed, Recalled, and 'Explained.'" *Political Behavior* 8:21–44.

McFarland, C., and C. Alvaro. 2000. "The Impact of Motivation on Temporal Comparisons: Coping with Traumatic Events by Perceiving Personal Growth." *Journal of Personality and Social Psychology* 79:327–43.

McFarland, C., and M. Ross. 1987. "The Relation Between Current Impressions and Memories of Self and Dating Partners." *Personality and Social Psychology Bulletin* 13:228–38.

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):883–98.

Page, Susan. 2009. "Poll: Health Care Views Take Sympathetic Tilt." *USA Today*, https://abcnews.go.com/amp/Politics/story?id=8316019\&page=1.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized 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.

Schryer, Emily, and Michael Ross. 2012. "People's Thoughts About Their Personal Pasts and Futures." In *Social Metacognition*, eds. Kenneth G. DeMarree, Pablo Briñol, 141–57. New York: Psychology Press.

Schuman, Howard, and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Content*. New York: Academic Press.

Slothuus, Rune. 2016. "Assessing the Influence of Political Parties on Public Opinion: The Challenge from Pretreatment Effects." *Political Communication* 33(2):302–27.

Smith, Gordon C. S., and Jill P. Pell. 2003. "Parachute Use to Prevent Death and Major Trauma Related to Gravitational Challenge: Systematic Review of Randomised Controlled Trials." *British Medical Journal* 327(7429):1459–61.

Strack, Fritz, and Leonard L. Martin. 1987. "Thinking, Judging, and Communicating: A Process Account of Context Effects in Attitudes Surveys." In *Social Information Processing and Survey Methodology*, edited by J. Hippler Hans, Schwarz Norbert, and Sudman Seymour, 123–48. New York: Springer-Verlag.

Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz. 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey Bass.

Tourangeau, Robert, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. New York: Cambridge University Press.

Wilson, A. E., and M. Ross. 2001. "From Chump to Champ: People's Appraisals of Their Earlier and Present Selves." *Journal of Personality and Social Psychology* 80:572–84.

Wilson, Rick. 2017. "Roy Moore's GOP Stands for 'Grand Old Pedobears.'" *The Daily Beast*. The Daily Beast Company.

Wilson, T. D., C. E. Houston, and J. M. Meyers. 1998. "Choose Your Poison: Effects of Lay Beliefs About Mental Processes on Attitude Change." *Social Cognition* 16:114–32.

Wood, Thomas, and Ethan Porter. 2018. "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence." *Political Behavior*.

Yair, Omer, and Gregory A. Huber 2021. "How Robust Is Evidence of Perceptual Partisan Bias in Survey Responses? A New Approach for Studying Expressive Responding." *Public Opinion Quarterly* 84: 469–92.

Yeh, Robert W., Linda R. Valsdottir, Michael W. Yeh, Changyu Shen, Daniel B. Kramer, Jordan B. Strom, Eric A. Secemsky, et al. 2018. "Parachute Use to Prevent Death and Major Trauma When Jumping from Aircraft: Randomized Controlled Trial." *British Medical Journal* 363.