Techniques for Measuring Political Beliefs

Matthew H. Graham *

November 8, 2024

Prepared for the Oxford Handbook of the Cognitive Science of Belief

Abstract

Political beliefs are generally measured using surveys. This chapter organizes prevailing techniques around four concerns about why respondents may not "believe" their answers. Researchers who think of beliefs as latent, general dispositions use multi-question scales to reduce measurement error. Researchers who think that a true belief must be held with some level of strength or credence use "don't know" response options or credence scales to attempt to filter out responses that do not qualify as a true belief. Researchers who worry that respondents are being insincere use techniques designed to encourage accurate reporting. Researchers who worry that respondents are being careless or mischievous, e.g. clicking random buttons or intentionally choosing absurd responses, use screening questions to attempt to filter such respondents out. In general, these techniques do a pretty good job of making sure that the "beliefs" measured in political surveys are related to something the respondent actually thinks, a weak notion of belief. They are a less reliable means of assuring that respondents believe their answers in a stronger sense, e.g. the definitions found in dictionaries or works of philosophy. Assessing whether a survey question measures beliefs in anything more than the weakest sense requires careful evaluation.

^{*}Assistant Professor, Temple University, USA. mattgraham@temple.edu. For helpful comments on drafts, I thank Brian Guay, Neil Van Leeuwen, and the anonymous reviewers. Early versions of this work were presented at the Public Policy Lab, Temple University; the American Politics Graduate Student Workshop, Yale University; and the 2019 meeting of the Midwest Political Science Association.

Quantitative generalizations about political beliefs are a staple of public discourse and academic research. Researchers generally select techniques for measuring political beliefs based on the goal of making such generalizations, resulting in a strong preference for multiple choice questions. Modifications are then made based on different concerns about whether multiple choice questions measure beliefs. This chapter reviews the concept of belief embodied by several of the most popular techniques for measuring political beliefs: multi-question scales, "don't know" (DK) response options, credence scales (e.g. confidence or certainty), techniques for reducing misreporting, and screeners for careless or fraudulent respondents. Collectively, existing techniques generally succeed in capturing some function of respondents' underlying dispositions toward the topic stated in the question—a weak sense of belief that is consistent with Zaller's (1992) and Tourangeau, Rips, and Rasinski's (2000) accounts of the nature of survey responses.

Throughout this chapter, I adopt the notion of beliefs and attitudes that is outlined in this volume's introduction, which may not be familiar to all readers. Consider the statement, "Dee <u>believes</u> that *the 2020 U.S. presidential election was not stolen.*" In the terms laid out in the introduction, the italicized portion is the contents of the belief, and the underlined portion ("believes") is Dee's attitude toward the contents. In contrast, scholars of politics tend to distinguish attitudes and beliefs based on their contents. To the typical scholar of politics, an attitude is a favorable or unfavorable evaluation (Eagly and Chaiken 1993, Holbrook 2011) while a belief is an evaluation related to what is true or false about the state of the world (Bullock and Lenz 2019). Although content-based distinctions are useful in some contexts, the measurement techniques that survey researchers use for attitudes and beliefs tend to follow the same set of structures. Laying aside content-based distinctions between attitudes and beliefs enables a clearer focus on what measurement techniques imply about "what believing is" and "what one is doing with ideas when one believes them" (paraphrased from introduction).

Believing a Survey Response

The most common approach to measuring political beliefs is a structured response (i.e., multiple choice) question in a survey. Some questions have as few as two response options. For example, to measure beliefs about whether the 2020 election was stolen, the 2022 American National Election Survey (ANES) asked, "Referring to the 2020 election, Donald Trump said 'the election was stolen.' Do you believe Trump's statement is true or false?" Similarly, to measure beliefs about what policies are more or less desirable, the Cooperative Election Study (CES) includes a series of binary choice options about which policy comes closest to the respondent's preferences. The universal health care question, for example, asks whether they would support or oppose the proposal "Expand Medicare to a single comprehensive public health care coverage program that would cover all Americans." At the high end, some questions have as many as 101 response options. For example, beliefs about the prevalence of various groups in a population are often measured on a scale from 0 percent to 100 percent (Ahler and Sood 2018; Peyton et al. 2022).

A naive interpretation of people's multiple choice responses is that people hold in their mind a "belief" about the specific proposition they have endorsed. This is known as the file drawer model: attitudes are stable evaluations of objects, and "[w]hen asked how they feel about something, such as legalized abortion, their Uncle Harry, or anchovies on a pizza, presumably they consult a mental file containing their evaluation. They look for the file marked abortion, Uncle Harry, or anchovies, and report the evaluation it contains" (Wilson and Hodges 1992, 38). A number of classic findings led survey researchers to reject file drawer-like notions of the beliefs measured in surveys. For example, survey responses are often quite unstable over time, with some respondents changing their answers in a seemingly random fashion (Converse 1964). The relative ordering of questions can exercise a large influence on responses, as can changes in question wording that would appear substantively insignificant to anyone with knowledge of the issue at hand (Schuman and Presser 1981). People express beliefs about completely made up conspiracy theories and policy proposals, which they could not possibly have believed in prior to the survey

(Schuman and Presser 1981; Oliver and Wood 2014).

To describe what surveys measure, researchers have developed weaker notions of the sense in which respondents believe their answers. Zaller's (1992) classic account holds that survey respondents are generally "make it up as they go along," drawing inferences about the words on the page based on information that may only be tangentially related to the proposition at hand. Similarly, Tourangeau et al. (2000) describe "a continuum corresponding to how well articulated a respondent's attitude is. At the more articulated end, the respondent has a preformed opinion just waiting to be offered; at the less articulated end, the respondent has no opinion whatsoever. Between these extremes, he or she may have a loosely related set of ideas to use in constructing an opinion" (12). Most responses are not "readouts of stored judgments" (22). To the extent this is true, survey responses are a projection of underlying attitudes and beliefs that may not be as specific as, and are often only loosely related to, the wording of the question and response options. The sense in which responses are "believed" can be as weak as the sense in which one believes a guess on a multiple choice test.

Open-ended questions are one reasonable way around people's tendency to endorse beliefs that had never entered their mind prior to the survey. Open-ended questions also convey information that may prompt responses to say things they would not have otherwise (e.g., Tourangeau et al. 2000, 41-44). However, open-ended questions cannot put words in people's mouths and minds in the same way that researchers and journalists do when they quote multiple choice questions and answers as if respondents "said" or "think" those words.¹ In practice, open-ended questions clash with researchers' desire to make generalizations about the prevalence of beliefs and to use measures of belief in statistical analysis that relates them to other beliefs or personal traits. For example, Thorson (2024) writes that as a first step, "[o]pen-ended questions in surveys can allow respondents to discuss their beliefs without the bias of interviewer-provided responses ... Next, the

¹Sudman et al. (1996) write, "Once respondents have formed a judgment, the interviewer usually does not allow them to report it in their own words. Rather, they are asked to endorse one of the response alternatives provided. In order to do this, they must format their response in line with the options given. Accordingly, the researcher's choice of response alternatives may strongly affect survey results" (73).

set of misperceptions identified in this first step can be used to inform a closed-ended survey to measure their prevalence, along with who holds them and the extent to which they are politicized" (34). Though the open-ended step assures that at least some people hold the specific belief in question, the closed-ended step retains the possibility that many or most responses reflect only the weak sense of belief described by Zaller, Tourangeau, and colleagues.

Given researchers' commitment to measuring the same belief across the entire sample, techniques for measuring belief generally take structured response questions as their foundation, then add features to address some specific reason why responses may not reflect beliefs. In what follows, I review the empirical strategies that political researchers use to address what I see as their three biggest concerns about the relationship between beliefs and survey responses: that survey measures are accurate on average but noisy in specific cases, that people answer survey questions even when they do not have beliefs, and that people systematically misreport their beliefs in surveys. Together, each concern and the corresponding empirical solutions embody some conception of belief. Table 1 summarizes the following sections.

Multi-Question Scales

One strategy for dealing with findings that many respondents "deduc[e] their position on the issue at hand from ideological predispositions or broad values" (Tourangeau et al. 2000, 172) is to try to measure the underlying disposition rather than more specific beliefs. Researchers generally carry out this task by asking a series of questions that are thought to be functions of the same underlying disposition. For example, to measure conspiratorial thinking, Uscinski et al. (2016) measure the following items on a 5-point scale from "strongly agree" to "strongly disagree."

- Much of our lives are being controlled by plots hatched in secret places.
- Even though we live in a democracy, a few people will always run things anyway.
- The people who really 'run' the country, are not known to the voters.
- Big events like wars, the current recession, and the outcomes of elections are controlled by small groups of people who are working in secret against the rest of us.

Concern	Strategies for addressing	Embodied conception of belief	
Survey measures of belief are accurate on average but may be noisy.	Ask multiple questions that are functions of the belief and combine them into a composite measure.	Belief is a latent, general disposition. Survey responses are noisy reflections of it.	
People answer survey questions even when they don't have a true belief.	 Provide a "don't know" or "not sure" response option. Ask respondents about credence (e.g., how confident or certain they are). 	To be called a belief, a proposition must be believed with sufficient credence (the threshold view). DK options and certainty scales raise the threshold.	
People use survey questions in a systematically different way than the researcher intends, e.g. to express partisan sentiments.	 Offer payments for correct answers. Provide an opportunity to express desired sentiments. Measure indirectly using a list experiment. Additional strategies (see text). 	Belief is the response option that is closest to the respondent's underlying beliefs. The measure of belief may still be noisy, and beliefs can be of any degree (i.e., no minimum threshold).	
People respond to surveys in a careless or mischievous manner.	Use screening questions that are likely to discriminate between careless or mischievous respondents and others.	Compatible with any conception of belief (because the strategy for addressing it is pre-screening, not a modification to the measure of belief).	

Table 1: Summary	of Measurement	Approaches	Discussed	Below

Once measured, the separate questions (or "items") are then aggregated into a single scale that is described as measuring the latent construct, in this case conspiratorial thinking. Researchers use a range of aggregation methods, from simply taking the average to more complex statistical methods like factor analysis or item response theory models (Ansolabehere et al. 2008; Boateng et al. 2018); the more closely the individual items captures the construct, the less the choice of the procedure matters. This approach is motivated by quantitative models that posit that each response is a combination of a "true" attitude plus noise (Achen 1975; Erikson 1979; Ansolabehere et al. 2008).

Multi-item scales are well motivated in that they embrace the minimal sense of belief in survey responses described by Zaller and others. Scales do not attempt to measure belief in specific propositions, instead asserting only that the individual questions reflect some latent disposition. Research on scale development recommends a number of steps to verify that this is the case, in-

cluding pre- and post-testing with content experts and the study population ("cognitive interviewing"; Beatty and Willis 2007), as well as statistical checks on the items' relationship to one another and to the latent construct (Boateng et al. 2018).

The focus on latent dispositions is a significant limitation for researchers who want to measure belief in a specific proposition (e.g., all of the examples in this volume's introduction). Indeed, some researchers explicitly define beliefs as being more specific than the type of generic predisposition that scales attempt to measure (Uscinski et al. 2016, 58). Relatedly, multi-item scales do not provide any assurance that the measured belief is held with a high degree of credence.² Multi-item scales also do not address systematic sources of measurement error that affect all of the component items similarly (Ansolabehere et al. 2008, 218), e.g. a systematic tendency to over-estimate small quantities (Landy et al. 2018) or give responses that are more partisan than one's underlying beliefs (Bullock et al. 2015; Prior et al. 2015).

One common concern about scales is that different survey respondents may interpret them differently. This phenomenon, known as "differential item functioning," is thought to invalidate between-group comparisons in many commonly used scales (Pietryka and MacIntosh 2022). How-ever, single-item measures can suffer from the same fundamental problem (King et al. 2004). For example, Jefferson (2024) finds that single-item measures of political ideology are interpreted differently by people of different racial groups. Allen et al. (2022) offer a useful perspective on validity differences between single-item measures and multi-item scales.

In sum, multi-question scales are a good choice for researchers who are comfortable with the weak notion of belief that appears in theories of the survey response. Their key vulnerabilities are shared by single-item measures, and their theoretical premise is consistent with research on what surveys actually do. Researchers who are interested in beliefs in specific propositions or are

²This is sometimes obscured by language like "true attitude" or "true belief," which to reasonable readers could imply some form of acceptance. In the underlying theoretical models, however, the true belief is just a parameter that is assumed to exist, and the computer will fit a value for it regardless of the sense in which people believe their answers. In a footnote, Erikson (1979) is especially plainspoken on this point: "the term 'non-attitude' is technically a misnomer in the sense that every respondent has a theoretical mean (true) position" (100).

concerned about systematic sources of measurement error must turn to other techniques.

Don't Know Response Options

A classic worry in the study of political beliefs is that survey responses do not reflect beliefs about the proposition contained in the question—and may not reflect anything at all. Converse (1964) famously argued that many survey responses were best thought of as "non-attitudes," too weakly believed to be considered meaningful.³

The most common approach to the concern that some responses don't really reflect beliefs is to provide a "don't know" (DK) option, or some closely worded variant like "not sure." Researchers who use DK options generally hope that respondents who do not think they know the answer will choose DK, while only those who do think they know the answer will choose between the substantive response options. This motivation is related to the threshold view of belief (Foley 1992): one's level of credence in their choice must cross some latent threshold in order to be considered a belief.⁴

Research suggests that DK options act as a low and permeable barrier. In terms of what DK options *do* filter out, studies of political knowledge find that respondents who choose DK do not answer correctly than one would predict by chance (Sturgis et al. 2008; Bullock 2011). This suggests that DK options primarily filter out total ignorance. However, DK options do not filter out *all* of the least meaningful responses. Studying peoples' confidence in beliefs about the economy and politicized controversies, Graham (2021*b*) estimates that DK options filter out about half of the least confident responses and about one-quarter of responses that would have been stated with

³The non-attitudes thesis, particularly the black-and-white model, is sometimes read as implying that a binary conception of belief truly fits the data. However, Converse only said that the model was a good fit to one question, the power and housing item; for the remainder, "a distribution of the population continuously across the total range of response probabilities is entirely compatible with the data" (72). Later, Converse (2000) wrote that "I tried to make clear that the [power and housing] issue was a limiting case, because of its location at the extreme boundary" but that both his supporters and detractors "had ... a basic incomprehension of the role of limiting cases in inquiry" (338).

⁴Although this measurement technique mixes together knowledge and beliefs in a way that philosophers are likely find uncomfortable, survey researchers tend not to dwell on this issue. The operational definition of knowledge in survey research is "true belief," without any consideration of issues like how the belief or originated or whether it is justified.



Figure 1: "Don't Know" Response Options as a Permeable Barrier

Note: Figure depicts a stylized version of the relationship between underlying beliefs and DK response option. In each panel, the top box represents the distribution of beliefs in the population, while the bottom boxes represent the distribution of survey responses that would obtain if half of respondents without a belief (hollow) chose DK. Comparing the two panels illustrates that when fewer people have meaningful beliefs, more choose DK *and* a smaller share of those who respond have meaningful beliefs.

middling levels of confidence. The partial success of DK options means that even though they serve their intended purpose to some extent, it should not be taken for granted that everyone who passes up the opportunity to say DK "believes" their answer in any strong sense.

To make the model of DK as a permeable barrier more concrete, consider the stylized example in Figure 1, which depicts how two underlying distributions of beliefs (top boxes) translate into a distribution of responses (bottom boxes) under some simplifying assumptions. Suppose that all respondents either have a belief (solid people) or do not have one (hollow people). Everyone who has a belief answers the survey question. Half of those without a belief say DK, while the other half answer the question. On question A, 1/3 of the population lacks a belief. Since DK is a permeable filter, half of those (1/6 of the total) say DK and 5/6 of respondents answer the question. Among those who provide an answer, 1/5 lack a belief. On question B, 2/3 of the population lacks a belief. Since only half say DK, two-thirds provide an answer. Among those who answer, half lack a belief—more than doubling relative to the scenario in which nonbelief was twice as common. The implied relationship between the proportion of DK responses and the level of credence among those who do not say DK bears out in data: a higher percentage DK corresponds to lower levels of response stability and confidence (Dodd and Svalastoga 1952; Graham 2021*b*).

In sum, though DK response options do some of what they are supposed to do, they do not accomplish their whole purpose. A DK response option is best thought of as a permeable barrier that provides a signal about the nature of responses across the entire sample. A high rate of DK is a signal to researchers and audiences that many of those who provided a substantive response also had difficulty connecting the question to their underlying beliefs.

Credence Scales

Dissatisfied with the low threshold imposed by DK responses, researchers often turn to measures of credence to identify those who truly believe their answers. In the study of attitudes (in the sense of subjective or evaluative beliefs), researchers refer to the measure of credence as strength or importance (Howe and Krosnick 2017). In the study of beliefs about matters that are not subjective or evaluative, researchers refer to the credence measure as confidence or certainty (Kuklinski et al. 1998, 2000).⁵ Although it is intuitive that higher-credence attitudes are in some sense more meaningful, evidence on their performance is mixed. In an earlier summary of this evidence, I wrote that "[t]he few previously published tests of the strength-stability relationship find that strong attitudes are only modestly more stable than weak attitudes, with little focus on exactly how strong the strongest attitudes are" (Graham 2023, 87).

Measures of credence have seen a revival of interest in recent years (e.g., Pasek, Sood and Krosnick 2015; Lee and Matsuo 2018; Graham 2020; Peterson and Iyengar 2021) alongside rising public concern about the spread and impact of misinformation. A battery of "misinformation

⁵This is related to the notion of epistemic confidence (van Leeuwen 2022).

items" was added to the ANES in 2020. Each question began by asking respondents to choose which of two statements was more likely to be true, consistent with research showing that "acquiescence bias" can inflate the apparent prevalence of beliefs that are measured with agree/disagree or true/false items (Schuman and Presser 1981; Hill and Roberts 2023). A follow-up question asked about confidence. For example, to measure the prevalence of belief in the efficacy of an ineffective but widely promoted cure for COVID, the ANES asked:

Which of these two statements do you think is most likely to be true?

- There is clear scientific evidence that the anti-malarial drug hydroxychloroquine is a safe and effective treatment for COVID-19
- There is not clear scientific evidence that the anti-malarial drug hydroxychloroquine is a safe and effective treatment for COVID-19

How confident are you about that?

- Not at all
- A little
- Moderately
- Very
- Completely

Following Kuklinski et al. (2000), standard practice is classify those who select one of the top two scale points on the confidence scale as having a belief, and to classify the remaining respondents as making a guess. As with DK options, this embodies a threshold conception of belief.

Measures of credence are an unreliable means of identifying people who believe their answers in a high-threshold or outright sense. Studying the ANES battery and a range of other questions designed to measure belief in misinformation, Graham (2023) finds that even respondents who report 100 percent certain belief in falsehoods may not firmly believe them. Respondents were asked identical questions in two surveys conducted about one to two months apart. Those who were initially confident in the correct answers stuck to them at high rates in the follow-up survey, suggesting that they really believed the answers. Those who initially claimed to be confident in falsehoods were much less likely to endorse them with equal confidence again (with some variation across the items). Graham calls such beliefs "miseducated guesses": there is clearly something in peoples' heads that leads to semi-consistent endorsement of false claims about hydroxychloroquine or the size of an electron, but it is not outright belief of the claim based on some prior exposure to it. However, measurement properties vary from question to question. Using the same techniques, Graham and Yair (2024*a*) find greater commitment among Republican respondents who endorse the belief that the 2020 election was stolen.

Like DK options, credence scales accomplish some of what they are supposed to. However, whether high levels of credence indicate decided beliefs or educated guesses appears to vary from question to question. This is more likely to be a problematic for attempts to infer the prevalence of beliefs, as is common in research on belief in misinformation and conspiracy theories (Pasek et al. 2015). It is not as likely to be problematic for research on how people react to information they consume, such as research on truth discernment (Pennycook and Rand 2019).

Expressive Responding

Expressive responding is another reason why survey responses might not accurately reflect respondents' underlying beliefs. According to accounts of expressive responding, survey responses are not just reflections of respondents' underlying beliefs about the content of the question. Instead, responses are contaminated by respondents' motivation to provide responses that are favorable to their side (Bullock et al. 2015; Prior et al. 2015). For example, as of this writing, the President of the United States is Joe Biden, a Democrat. If asked about the state of the economy, respondents who like Biden or Democrats might say the economy is doing well even if they think it is doing poorly. Similarly, respondents who dislike Biden or Democrats might say the economy is doing belief.

Techniques for mitigating expressive responding are distinct from the issue of credence.⁶

⁶Although some research using pay-for-correct treatments has attempted to integrate the issue of credence, these

If successful, the techniques discussed below encourage respondents to sincerely state which response option comes closest to their underlying beliefs or attitudes. They provide no assurance that respondents held the belief before the survey, that they strongly believe their response, or anything of the sort. For example, in a review, Bullock and Lenz (2019) consider the example of a two-option multiple choice question. In their terms, any respondent who thinks that there is at least a 51 percent chance that one of the two options is correct — the lowest possible threshold without involving decimal places — "believes" their response.

Researchers use a variety of techniques to study expressive responding. The most common is *payment for correct answers*. In these experiments, randomly selected respondents are offered incentives between 10 cents and 2 dollars per question that is answered correctly. Partisan differences are smaller in the incentive condition, suggesting that respondents were either less likely to misreport their beliefs ("partisan cheerleading"; Bullock et al. 2015) or less likely to draw on a biased sample of their underlying beliefs as they made their on-the-spot inferences (Prior et al. 2015). A variant of the pay-for-correct strategy offers randomly selected respondents a financial incentive for making a correct prediction about the future (Allcott et al. 2020; Graham and Yair 2024*a*). An obvious limitation of the pay-for-correct is that it can only be applied to factual questions with correct answers. More subtly, there must also be a widely accepted point of reference for determining the correct answers. Otherwise, respondents may try to anticipate what the researcher believes to be true (Berinsky 2018).

The theory of response substitution gives rise to another paradigm for studying expressive responding: *asking the unasked question*. Response substitution is the idea that people sometimes want to answer a different question than the researcher has asked. Gal and Rucker (2011) give the example of a restaurant with terrible food and great service. If asked at the end of the meal about the quality of the service, one might say "terrible" in order to express their disapproval of the food.

attempts have variously failed to be robust to risk aversion and genuine uncertainty (Bullock et al. 2015), failed to maintain measurement equivalence (Graham 2021*a*, Chapter 6), or do not attempt to reduce expressive responding in the measure of credence (Peterson and Iyengar 2021).

But if given a chance to rate the food first, one might rate the service more honestly. Researchers who apply this paradigm to the study of politics seek to anticipate and ask the "unasked question" that the respondent wants to answer (Yair and Huber 2020). For example, in studies of media literacy, Democrats and Republicans often claim that opinions they agree with are verifiable facts. This tendency is reduced when partisans are first given the opportunity to state their agreement or disagreement with the statement (Graham and Yair 2024*b*). A key limitation is the need to anticipate the specific sentiment that that the respondent wants to express.

List experiments, also known as the item count technique, were originally developed to study sensitive traits like illegal behavior (Miller 1984). Instead of directly asking respondents about the belief of interest, list experiments present respondents with a list of propositions and ask how many the respondent agrees with or how many apply to them. Beginning with Kuklinski et al. (1997), many studies have applied list experiments to the study of political beliefs. An important limitation of this strategy is that it produces large standard errors. A range of modifications can at least partly address this limitation (a good summary appears in Blair et al. 2020, 1305). Another limitation is the need to format the proposition of interest and the corresponding direct question in a way that works with the list format. This largely limits list experiments to binary agree/disagree items, with no information about credence (e.g., no DK option).

Implausible beliefs create another opportunity to observe expressive responding. This strategy takes advantage of situations in which one belief is so unlikely to be sincerely held that it can be ruled out by assumption. Schaffner and Luks (2018) developed this strategy to study a controversy over the relative size of the crowd at Barack Obama and Donald Trump's presidential inaugurations. It rained the day of Trump's inauguration, dampening attendance. Trump claimed that his crowd had been larger. Schaffner and Luks detected expressive responding by showing respondents photos of the two crowds and asking which was larger. About 15 percent of Republicans picked the obviously smaller crowd, defying the plain visual evidence to express an implausible but party-congenial belief. Five years later, Ross and Levy (2023) replicated these findings but at smaller magnitudes, suggesting that the tendency toward expressive responding decays as the relevant events fade from memory or lose salience. The central limitation of this strategy is that it can only be applied when the belief in question is so implausible as to be ruled out by assumption.

Despite the variety of techniques that have been developed to study expressive responding, the limitations attached to each technique means that in practice, researchers who suspect expressive responding in a particular case may not have a good option, or at least not an obvious one. For example, research on expressive responding about the economy does not examine the most commonly used measures of economic perceptions, which are subjective (e.g. from the ANES, "What do you think about the state of the economy these days in the United States? Would you say the state of the economy is good, neither good nor bad, bad, or very bad?"). Instead, in order to fit the pay-for-correct paradigm, researchers focus on questions about specific economic statistics like the unemployment rate or inflation rate (Bullock et al. 2015; Prior et al. 2015), which tend to register lower levels of partisan bias than the subjective items (Roush and Sood 2022).

Available techniques also provide little assurance as to their efficacy. When a technique for reducing expressive responding works, researchers can reasonably infer that some expressive responding existed. However, this provides no assurance that more expressive responding could not have been eliminated. Similarly, if a technique does not work, it is not clear if this means that expressive responding was absent or if the treatment was simply too weak (e.g., too small a payment) or mistargeted (e.g., asking the wrong unasked question).⁷ Moreover, a central step in most studies of expressive responding is to code questions or responses in terms of a variable that measures the hypothesized source of bias (namely, partisanship). If expressive responding occurs for another reason,⁸ the empirical analysis will not detect it even if the technique has successfully

⁷Though it is intuitive that some treatments could be too weak to eliminate expressive responding, evidence on this question is limited. Though the second study in Bullock et al. (2015) is often interpreted as showing that larger monetary incentives are more effective, the evidence does not support this claim. The article does not include a statistical test for the difference in effects between stronger and weaker treatments, and the difference in effect sizes between the smallest and largest payment amounts is about half a standard error (Table 4). Peterson and Iyengar (2022) find no difference in effect sizes between incentives of different amounts.

⁸For example, Crowder-Meyer and Ferrín (2021) find that ideological gaps in fact-opinion discernment are larger (smaller) among respondents with aligned (cross-cutting) racial identities.

reduced it. For all of these reasons, it is easier to be sure that expressive responding is present than it is to be sure that it is absent. The best one can do is think carefully about what approaches are applicable to one's case and apply multiple approaches if feasible (e.g., Berinsky 2018; Graham and Yair 2024*a*).

Careless and Mischievous Responding

Researchers studying political beliefs must also contend with the possibility that respondents are not taking the survey in good faith (Mercer et al. 2024). Some respondents rush through the survey, carelessly clicking random buttons or failing to listen carefully to the interviewer (Westwood et al. 2022). Others are up to no good, misrepresenting their characteristics in order to gain eligibility for the survey or for the fun of it (Robinson-Cimpian 2014). Such respondents can significantly inflate the apparent prevalence of beliefs and the apparent correlation between beliefs and other personal traits, particularly when the beliefs or personal traits in question are only held by a small portion of the population (Huang et al. 2015).

The belief that political violence is justified is a notable example of the threat from careless responding. One popular measure of support for political violence asks respondents "How much do you feel it is justified for [respondent's own party] to use violence in advancing their political goals these days?" with five response options: "Not at all," "A little," "A moderate amount," "A lot," and "A great deal." Respondents who choose the first option are scored as not supporting violence, while the other four options are scored as supporting violence. Consequently, a respondent choosing randomly would have an 80 percent chance of being scored as supporting violence. Across three studies, Westwood et al. (2022) find that about two-thirds of endorsements of political violence come from the roughly one-fourth of respondents who failed engagement checks (Table 1). Westwood and colleagues also find that respondents who genuinely endorse violence generally have much less severe forms of violence in mind than commentators: think vandalism, not murder.

Careless responding can also inflate correlations between traits (Fan et al. 2006; Robinson-

Cimpian 2014; Huang et al. 2015). This is because someone choosing randomly has a higher than average probability of selecting *any* response option that is uncommon. When two traits are uncommon or rare, careless responding will generally inflate the correlation between them, and can result in a larger correlation than is present in either subgroup. Sulik et al. (2023) find that careless respondents inflate the gap between cognitive biases and delusion-like beliefs. Because extreme levels of cognitive bias and delusion-like beliefs are both relatively uncommon, careless or inattentive respondents are disproportionately likely to have both traits, thus inflating the association between the two.

Another notable case pertains to the apparent behavioral consequences of belief in misinformation about COVID-19. In June 2020, a team of researchers from the Centers for Disease Control and Prevention (CDC) published an estimate that 4 percent of Americans had drank or gargled bleach in order to cure COVID (Gharpure et al. 2020). Seemingly adding to the credibility of these reports, respondents who reported gargling bleach were also more likely to report plausibly related adverse health effects during the past month (e.g., nose, sinus, eye, or skin irritation). In two replication studies in the same sample, Litman et al. (2023) show that all of the alleged bleach-garglers were careless or mischievous respondents. A large majority reported other implausible traits, such as eating concrete for its iron content, having had a fatal heart attack, or being able to name every U.S. Senator who has ever lived. Careless respondents were also responsible for the apparent correlation between bleach gargling and associated symptoms; since most people do not experience throat irritation in any given 30-day period, careless respondents made up a disproportionate share of this group as well.

To deal with careless responses, researchers first use screening questions to identify respondents who are likely to respond carelessly. A wide variety of screening questions have been developed. Some researchers use attention checks that can only be answered correctly by respondents who are reading carefully. For example, Berinsky et al. (2014) field a question that starts out asking about the respondent's favorite color, then directs careful readers to choose a particular color rather than their favorite. Those who choose the requested color are likely to be paying attention. Other indicators of carelessness include inconsistent reporting of demographic information like age, self-reporting other implausible traits, changing their answer when asked the same question again, or indicating in follow-up questions that they chose their original response by mistake (Lit-man et al. 2023). An unusually high prevalence of traits that are widely understood to be rare, such as opposition to abortion among young people, can also signal the likely presence of low-quality respondents (Mercer et al. 2024). Respondents who fail these checks are often rendered ineligible for the survey or later removed from the analysis. It is also possible to impute or bound their attitudes (Tyler et al. 2024).

Conclusion

Techniques for measuring political beliefs are dominated by multiple choice questions consisting of propositions selected by researchers. The centrality of this technique is driven by what researchers want to do with the data: quantify the prevalence of various beliefs and their statistical associations with other beliefs or traits. Methods of assuring that survey measures reflect people's beliefs are appended based on various conceptions of belief. Existing techniques are pretty good at capturing people's general dispositions toward the words on the page, especially when a multiitem scale is deployed. Existing techniques are inconsistent at, and sometimes bad at, measuring beliefs in specific propositions. The success or failure of a given technique varies from question to question depending on the underlying content of people's beliefs.

The weak sense in which prevailing techniques reliably measure political beliefs can be dissatisfying. Those who want to measure belief in a stronger or more specific sense should look for evidence that specific survey questions measure belief in the specific sense of belief that one has in mind. For example, to interrogate the sense in which Republicans who say the 2020 election was decided by fraud believe their answers, Graham and Yair (2024a) apply a range of techniques for detecting expressive responding and probing the reliability of measured credence. Though

the resulting evidence is not unassailable, it provides some assurance that reports of this belief are genuine in two specific senses: (a) they are sincere and (b) they often reflect a high level of credence.

Regardless of one's conception of belief, researchers need to be attentive to data quality, especially when the belief in question is uncommon or is observed to be more prevalent among a small subgroup of the population. Although the foundational cases in the literature on careless and mischievous respondents came from pencil-and-paper surveys (Fan et al. 2002), these problems may be more severe in online convenience samples, which have grown in popularity in recent years. Even samples recruited by vendors with good reputations are vulnerable. For example, Ansolabehere et al. (2015) and Mercer et al. (2024) find that widely-circulated statistics based on YouGov samples were likely to have been products of a collision between careless or fraudulent responding, rare traits, and small population subgroups.

Despite the reasons to be pessimistic about the current state of practice, there are also reasons for optimism. The survey research community has a strong commitment to improvement, as illustrated by its forthright and detailed responses to well-known forecasting errors in presidential polling. Indeed, the critiques levied here are only possible because of this culture of self-interrogation. At the end of the day, the fundamental reason that our expectations for surveys outstrip their capabilities is the lack of a better option for learning what we want to know. Continuing the rich tradition of careful examination of the relationship between theory and practice will help the study of political beliefs continue to improve as a cognitive science.

References

- Achen, C. H. (1975), 'Mass Political Attitudes and the Survey Response', *The American Political Science Review* 69/4: 1218–1231.
- Ahler, D. J. and Sood, G. (2018), 'The Parties in Our Heads: Misperceptions about Party Composition and Their Consequences', *The Journal of Politics* 80/3: 964–981.
- Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M. and Yang, D. (2020), 'Polarization and Public Health: Partisan Differences in Social Distancing during the Coronavirus Pandemic', *Journal of Public Economics* 191.
- Allen, M. S., Iliescu, D. and Greiff, S. (2022), 'Single Item Measures in Psychological Science', European Journal of Psychological Assessment 38/1: 1–5.
- Ansolabehere, S., Luks, S. and Schaffner, B. F. (2015), 'The perils of cherry picking low frequency events in large sample surveys', *Electoral Studies* 40: 409–410.
- Ansolabehere, S., Rodden, J. and Snyder, J. M. (2008), 'The strength of issues: Using multiple measures to gauge preference stability, ideological constraint, and issue voting', *American Political Science Review* 102/2: 215–232.
- Beatty, P. C. and Willis, G. B. (2007), 'Research synthesis: The practice of cognitive interviewing'.
- Berinsky, A. J. (2018), 'Telling the Truth about Believing the Lies? Evidence for the Limited Prevalence of Expressive Survey Responding', *The Journal of Politics* 80/1: 211–224.
- Berinsky, A. J., Margolis, M. F. and Sances, M. W. (2014), 'Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys', *American Journal of Political Science* 58/3: 739–753.
- Blair, G., Coppock, A. and Moor, M. (2020), 'When to Worry about Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments', *American Political Science Review* 114/4: 1297–1315.
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R. and Young, S. L. (2018), 'Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer', *Frontiers in Public Health* 6.
- Bullock, J. G. (2011), 'Elite influence on public opinion in an informed electorate', American Political Science Review 105/3: 496–515.
- Bullock, J. G., Gerber, A. S., Hill, S. J. and Huber, G. A. (2015), 'Partisan Bias in Factual Beliefs about Politics', *Quarterly Journal of Political Science* 10: 1–60.
- Bullock, J. G. and Lenz, G. (2019), 'Partisan Bias in Surveys', *Annual Review of Political Science* 22: 325–342.
- Converse, P. E. (1964), The nature of belief systems in mass publics, Free Press, pp. 206-61.
- Converse, P. E. (2000), 'Assessing the Capacity of Mass Electorates', *Annual Review of Political Science* 3: 331–53.

- Crowder-Meyer, M. and Ferrín, M. (2021), 'The Effects of Ideological and Ethnoracial Identity on Political (Mis)Information', *Public Opinion Quarterly* 85/3: 753–779.
- Dodd, S. C. and Svalastoga, K. (1952), 'On estimating latent from manifest undecidedness: the "don't know" percent as a warning of instability among the knowers', *Educational and Psychological Measurement* 12/3: 467–471.
- Eagly, A. H. and Chaiken, S. (1993), The Psychology of Attitudes.
- Erikson, R. S. (1979), 'The SRC Panel Data and Mass Political Attitudes', *British Journal of Political Science* 9/1: 89–114.
- Fan, X., Miller, B. C., Christensen, M., Park, K. E., Grotevant, H. D., van Dulmen, M., Dunbar, N. and Bayley, B. (2002), 'Questionnaire and interview inconsistencies exaggerated differences between adopted and non-adopted adolescents in a national sample', *Adoption Quarterly* 6/2: 7–27.
- Fan, X., Miller, B. C., Park, K.-E., Winward, B. W., Christensen, M., Grotevant, H. D. and Tai, R. H. (2006), 'An Exploratory Study about Inaccuracy and Invalidity in Adolescent Self-Report Surveys', *Field Methods* 18/3: 223–244.
- Foley, R. (1992), 'The Epistemology of Belief and the Epistemology of Degrees of Belief', American Philosophical Quarterly 29/2: 111–124.
- Gal, D. and Rucker, D. D. (2011), 'Answering the Unasked Question: Response Substitution in Consumer Surveys', *Journal of Marketing Research (JMR)* 48/1: 185–195.
- Gharpure, R., Hunter, C. M., Schnall, A. H., Barrett, C. E., Kirby, A. E., Kunz, J., Berling, K., Mercante, J. W., Murphy, J. L. and Garcia-Williams, A. G. (2020), 'Knowledge and Practices Regarding Safe Household Cleaning and Disinfection for COVID-19 Prevention — United States, May 2020', *MMWR*. *Morbidity and Mortality Weekly Report* 69/23: 705–709.
- Graham, M. H. (2020), 'Self-Awareness of Political Knowledge', Political Behavior 42/1: 305–326.
- Graham, M. H. (2021*a*), 'Mismeasuring Misperceptions: How Surveys Distort the Nature of Partisan Belief Differences'.
- Graham, M. H. (2021*b*), "We Don't Know' Means 'They're Not Sure", *Public Opinion Quarterly* 85/2: 571–93.
- Graham, M. H. (2023), 'Measuring Misperceptions?', American Political Science Review 117/1: 85-102.
- Graham, M. H. and Yair, O. (2024*a*), 'Expressive Responding and Belief in 2020 Election Fraud', *Political Behavior* 46: 1349–74.
- Graham, M. H. and Yair, O. (2024b), 'Less Partisan but No More Competent: Expressive Responding and Fact-Opinion Discernment', *Public Opinion Quarterly* /forthcoming.
- Hill, S. J. and Roberts, M. E. (2023), 'Acquiescence Bias Inflates Estimates of Conspiratorial Beliefs and Political Misperceptions', *Political Analysis* 31/4: 575–590.
- Holbrook, A. L. (2011), Attitude Change Experiments in Political Science, Cambridge University Press, pp. 141–154.

Howe, L. C. and Krosnick, J. A. (2017), 'Attitude strength', Annual Review of Psychology pp. 237–260.

- Huang, J. L., Liu, M. and Bowling, N. A. (2015), 'Insufficient effort responding: Examining an insidious confound in survey data.', *Journal of Applied Psychology* 100/3: 828–845.
- Jefferson, H. (2024), 'The Curious Case of Black Conservatives: Assessing the Validity of the Liberal-Conservative Scale among Black Americans', *Political Behavior* /forthcoming: 1–20.
- King, G., Murray, C. J. L., Health, W. and a Salomon, J. (2004), 'Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research', *American Political Science Review* 98/1: 191–207.
- Kuklinski, J. H., Cobb, M. D. and Gilens, M. (1997), 'Racial Attitudes and the 'New South", *The Journal* of *Politics* 59/323-349.
- Kuklinski, J. H., Quirk, P. J., Jerit, J., Schwieder, D. and Rich, R. F. (2000), 'Misinformation and the currency of democratic citizenship', *The Journal of Politics* 62/3: 790–816.
- Kuklinski, J. H., Quirk, P. J., Schwieder, D. W. and Rich, R. F. (1998), "Just the Facts, Ma'am': Political Facts and Public Opinion Source', *The Annals of the American Academy of Political and Social Science* 560/November: 143–154.
- Landy, D., Guay, B. and Marghetis, T. (2018), 'Bias and ignorance in demographic perception', *Psycho-nomic Bulletin and Review* 25/5: 1606–1618.
- Lee, S. and Matsuo, A. (2018), 'Decomposing political knowledge: What is confidence in knowledge and why it matters', *Electoral Studies* 51/1: 1–13.
- Litman, L., Rosen, Z., Hartman, R., Rosenzweig, C., Weinberger-Litman, S. L., Moss, A. J. and Robinson, J. (2023), 'Did people really drink bleach to prevent COVID-19? A guide for protecting survey data against problematic respondents', *PLoS ONE* 18/7 July.
- Mercer, A., Kennedy, C. and Keeter, S. (2024), 'Online opt-in polls can produce misleading results, especially for young people and Hispanic adults', *Pew Research Center* /March 2024.
- Miller, J. D. (1984), 'A New Survey Technique for Studying Deviant Behavior'.
- Oliver, J. E. and Wood, T. J. (2014), 'Conspiracy Theories and the Paranoid Style of Mass Opinion', American Journal of Political Science 58/4: 952–966.
- Pasek, J., Sood, G. and Krosnick, J. A. (2015), 'Misinformed About the Affordable Care Act? Leveraging Certainty to Assess the Prevalence of Misperceptions', *Journal of Communication* 65/4: 660–673.
- Pennycook, G. and Rand, D. G. (2019), 'Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning', *Cognition* 188/September 2017: 39–50.
- Peterson, E. and Iyengar, S. (2021), 'Partisan Gaps in Political Information and Information-Seeking Behavior: Motivated Reasoning or Cheerleading', *The American Journal of Political Science* 65/1: 133–47.
- Peterson, E. and Iyengar, S. (2022), 'Partisan reasoning in a high stakes environment: Assessing partisan informational gaps on COVID-19', *Harvard Kennedy School Misinformation Review*.
- Peyton, K., Weiss, C. M. and Vaughn, P. E. (2022), 'Beliefs about minority representation in policing and support for diversification', *Proceedings of the National Academy of Sciences* 119/52.

- Pietryka, M. T. and MacIntosh, R. C. (2022), 'ANES Scales Often Do Not Measure What You Think They Measure', *The Journal of Politics* 84/2: 1074–1090.
- Prior, M., Sood, G. and Khanna, K. (2015), 'You Cannot be Serious: The Impact of Accuracy Incentives on Partisan Bias in Reports of Economic Perceptions', *Quarterly Journal of Political Science* 10: 489–518.
- Robinson-Cimpian, J. P. (2014), 'Inaccurate estimation of disparities due to mischievous responders: Several suggestions to assess conclusions', *Educational Researcher* 43/4: 171–185.
- Ross, R. M. and Levy, N. (2023), 'Expressive Responding in Support of Donald Trump: An Extended Replication of Schaffner and Luks (2018)', *Collabra: Psychology* 9/1.
- Roush, C. E. and Sood, G. (2022), 'An Unclear Gap: How Vague Response Options Produce Partisan Knowledge Gaps', Unpublished Manuscript /December 2022.
- Schaffner, B. F. and Luks, S. (2018), 'Misinformation or Expressive Responding?', Public Opinion Quarterly 82/1: 135–147.
- Schuman, H. and Presser, S. (1981), Questions and Answers in Attitude Surveys, SAGE Publications.
- Sturgis, P., Allum, N. and Smith, P. (2008), 'An experiment on the measurement of political knowledge in surveys', *Public Opinion Quarterly* 72/1: 90–102.
- Sudman, S., Bradburn, N. M. and Schwarz, N. (1996), *Thinking About Answers: The Application of Cogni tive Processes to Survey Methodology.*
- Sulik, J., Ross, R. M., Balzan, R. and McKay, R. (2023), 'Delusion-like beliefs and data quality: Are classic cognitive biases artifacts of carelessness?', *Journal of Psychopathology and Clinical Science* 132/6: 749– 760.
- Thorson, E. (2024), The Invented State, Oxford University Press.
- Tourangeau, R., Rips, L. J. and Rasinski, K. (2000), *The Psychology of Survey Response*, Cambridge University Press.
- Tyler, M., Grimmer, J. and Westwood, S. (2024), 'A Statistical Framework to Engage the Problem of Disengaged Survey Respondents: Measuring Public Support for Partisan Violence', *Working Paper* /Rice University.
- Uscinski, J. E., Klofstad, C. and Atkinson, M. D. (2016), 'What Drives Conspiratorial Beliefs? The Role of Informational Cues and Predispositions', *Political Research Quarterly* 69/1: 57–71.
- Westwood, S. J., Grimmer, J., Tyler, M. and Nall, C. (2022), 'Current research overstates American support for political violence', *Proceedings of the National Academy of Sciences* 119/12.
- Wilson, T. D. and Hodges, S. D. (1992), Attitudes as Temporary Constructions, Erlbaum, pp. 37–65.
- Yair, O. and Huber, G. A. (2020), 'How robust is evidence of perceptual partisan bias in survey responses? A new approach for studying expressive responding', *Public Opinion Quarterly* 84/2: 469–92.
- Zaller, J. (1992), The Nature and Origins of Mass Opinion, Cambridge University Press.