

When Information Affects Attitudes: The Effectiveness of Targeting Attitude-Relevant Beliefs

Yamil Velez, Patrick Liu*, and Scott Clifford†

March 29, 2025

Abstract

Do citizens update strongly held beliefs when presented with belief-incongruent information, and does such updating affect downstream attitudes? Though fact-checking studies find that corrections reliably influence beliefs, attitudinal effects are negligible. We argue that such findings may reflect belief relevance – the extent to which specific beliefs bear on attitudes. Using large language models (LLMs), we elicit deeply held issue attitudes and “focal beliefs” that are described as central to those attitudes. We then randomly assign participants to receive either an LLM-generated factual argument targeting (1) their focal belief, (2) an attitude-relevant but unmentioned belief (“distal belief”), or (3) a placebo. In experiments with two large online convenience samples ($N_1 = 2,468$; $N_2 = 1,910$), we show that counterarguments successfully decrease both focal and distal belief strength, with effects persisting after one week. More importantly, focal belief counterarguments produce larger and more durable attitude change than distal counterarguments. These findings suggest that corrective information can successfully shift political attitudes when targeting “focal beliefs,” providing new evidence for the role of belief relevance in persuasion.

*Columbia University Department of Political Science

†Texas A&M University Department of Political Science

Introduction

The duties of democratic citizenship place a premium not only on voter learning but on the assimilation of accurate information into reasoned political judgments. Under traditional theories of democratic accountability, voters must be able to convert information about performance into judgments of incumbent governments and cast their votes accordingly (Fearon, 1999). Likewise, the cause for widespread concern about misinformation is not merely that citizens confidently hold false beliefs, but that policy and candidate preferences are formed on their basis (Kuklinski et al., 2000; Hochschild and Einstein, 2015). Fact-checking campaigns may supply citizens with accurate beliefs, but their efficacy hinges on citizens translating belief change into attitude change.

Recent studies document that, across varied issues and contexts, interventions that improve citizens' knowledge fail to consistently or substantially change supposedly downstream attitudes—in spite of mounting evidence that people generally update their beliefs in an unbiased fashion (Coppock, 2023; Arias et al., 2022; Gerber and Green, 1999). Improving knowledge of government spending and inequality produces inconsistent effects on economic policy preferences (Barnes et al., 2018; Cook, Jacobs and Kim, 2010; Kuziemko et al., 2015). Factual corrections eliminate false beliefs about vaccines without affecting intention to vaccinate (Porter, Velez and Wood, 2022; Helfers and Ebersbach, 2023) and counteract misinformation propagated by political candidates while exhibiting minimal influence on candidate support (Nyhan et al., 2020; Coppock et al., 2023). Efforts to reduce hostility toward immigrants and out-partisans by correcting misperceptions also recover mixed results (Hopkins, Sides and Citrin, 2019; Grigorieff, Roth and Ubfal, 2020; Dias et al., 2024).

A review of the literature reveals little consensus for what this empirical pattern implies about mass political behavior. These findings at least seem to reaffirm the intuition that political attitudes are the sum of citizens' prior knowledge and evaluations, and are less elastic than their constituent parts (Barnes et al., 2018). Other popular interpretations

are less sanguine about the benefits of a more informed citizenry. Evidence that learning aggregate economic information facilitates retrospective voting only among politically sophisticated voters suggests some may lack the requisite cognitive skills to make the link between belief and attitude change (Alt, Lassen and Marshall, 2016). Worse still, misperceptions may be the consequence of political attitudes rather than the cause (Hopkins, Sides and Citrin, 2019; Lawrence and Sides, 2014). In this view, beliefs are little more than post-hoc rationalizations of attitudes conceived in order to shield one’s prior commitments from incongruent evidence, an idea consistent with theories of cognitive dissonance and motivated reasoning (Festinger, 1957; Kunda, 1990; Lodge and Taber, 2013).

Why do interventions so often shift beliefs but fail to move attitudes? Answering these questions, we argue, requires paying more careful attention to the relevance that people themselves assign to beliefs. Factual interventions are ordinarily designed to target beliefs that academics or policy analysts perceive *a priori* as focal to a policy debate or momentarily pertinent due to widely circulating misinformation. Yet to the mass public, a wholly different set of beliefs may matter. For one, the facts citizens deem relevant to a policy debate are shaped by how politicians frame the debate, and thus may differ from what experts prescribe as relevant (Kuklinski et al., 1998). This concern bears resemblance to Converse’s (1964, 8-10) familiar argument that mass ideology is more likely to exhibit “social” than “logical” constraint: voters’ ideas about “what goes with what” tend not to emerge organically as logical wholes, but rather are diffused from elites.¹ Moreover, because relevance depends on politics as well as people’s personal values—which shape the criteria used to determine what information matters (Gerber and Green, 1999, 193)—relevant beliefs may differ with partisan identity and group membership, leaving the false impression that prior attitudes bias subsequent attitude change. Even if scholars identify and account for the

¹Importantly, the argument and evidence in this paper do not speak directly to Converse’s notion of constraint between idea-elements, which concerns the bundle of attitudes a person holds more so than the set of factual beliefs held about an attitude object. Our findings about belief change facilitating attitude change should not be extrapolated as evidence for dynamic constraint, which describes how change in one attitude might beget change in another.

“modal salient beliefs” within a population or subgroup, doing so may mask substantial heterogeneity between subjects (Fishbein and Ajzen, 2010, 102), which can mute downstream attitude change. The total effect is that belief change may facilitate little to no attitude change when interventions target beliefs that respondents consider only weakly relevant.

Our argument proceeds in two steps. First, we show that relevance has been central to influential theories across psychology and political science which treat beliefs as “the basic building blocks of attitudes” (Eagly and Chaiken, 1993, 103). We offer conceptual clarity surrounding the distinction between beliefs and attitudes and summarize how attitudes may be modeled as the weighted sum of beliefs about an attitude object, where the influence of each belief is in direct proportion to its relevance. In doing so, this section illustrates why the concept of relevance enables scholars to reconcile widespread evidence of inconsistent or muted effects of information on attitudes with the standard intuition of belief-attitude models—“a reasonable and consistent link between information about an object and attitude toward that object” (Fishbein and Ajzen, 2010, 99).

To address the longstanding methodological challenge of identifying and targeting a respondent’s personally relevant beliefs, we developed a respondent-tailored design enabled by the large language model (LLM) GPT-4o. In the first stage, respondents engage in dialogue with an AI chatbot about a political attitude they consider deeply important and their reasoning behind that view. GPT then synthesizes a summary attitude, a “focal” belief underlying that attitude, and a second “distal” belief generated without reference to chat input. In the second stage, respondents are randomly exposed to an intervention that targets a focal belief, distal belief, or placebo topic, in each case displaying a paragraph-length counterargument incorporating substantial factual information. This design furthers the development of tailored experimental methods to advance long-standing debates in political persuasion

Across the two studies, we provide consistent evidence that belief relevance is critical

to attitude change. Tailored counterarguments were equally effective at shifting focal and distal beliefs. However, the attitudinal effects of focal belief counterarguments were twice as large, as well as significantly different based on a random effects meta-analysis of twelve attitudinal outcomes. Additionally, in follow-up surveys approximately ten days after initial exposure, we find some persistence of attitudinal effects of the focal belief counterarguments but no significant effects of the distal belief counterarguments. These differences in attitude change despite comparable belief change testify to the importance of accounting for belief relevance at the respondent level. We conclude by exploring additional mechanisms and discussing implications for the studies of persuasion and democratic politics.

Distinguishing Beliefs From Attitudes

The growing literature on the effectiveness of fact-checking has found that fact checks are largely successful in improving factual accuracy (Walter et al., 2020). Across settings (Porter, Velez and Wood, 2023) and time (Coppock et al., 2023), fact checks tend to have modest effects on factual beliefs, with little risk of backfire. Yet, despite their apparent effectiveness on epistemic grounds, the downstream consequences of fact checks on attitudes and behaviors are modest, if not negligible. Across diverse outcomes such as candidate evaluations (Nyhan et al., 2020) and vaccination intentions (Porter, Velez and Wood, 2022), corrective information rarely has discernible effects, even in well-powered studies.

In a typical design, beliefs and attitudes related to the factual corrections are measured in tandem (see Porter and Wood (2024a) for a review). Doing so overlooks fine-grained but theoretically important distinctions between the two, rendering it difficult to explain divergent effects on beliefs versus attitudes. Whereas beliefs pertain to perceptions of facts or states of the world, attitudes reflect opinions or evaluations that lack inherent truth value, instead representing personal preferences, values, or normative judgments about what ought to be (Petty, 2024, 1276). For example, one might hold the belief that inflation rates are

rising (a factual claim), while simultaneously expressing the attitude that the government should take aggressive action to combat inflation (an evaluative position).

To illustrate how belief and attitude change may nevertheless be linked, it helps to represent the attitude formation process in an algebraic form, drawing from a long tradition in psychology (e.g., Fishbein and Ajzen, 1975; Anderson, 1981). The basic building blocks for this model come by way of Bayesian updating, one of the dominant frameworks for understanding the effectiveness of corrective interventions and to political information more generally (Gerber and Green, 1999). Bayesian updating is decidedly about *beliefs*, not attitudes. According to these models, individuals possess prior beliefs about a state of the world (e.g., rising inflation rates) and update these beliefs when presented with new information, weighing both the strength of their prior beliefs and the credibility of the new evidence. In the simplest model, the posterior probability that one’s view of a state of the world, b , is true given some evidence, e , depends on the prior probability of b (i.e., $\Pr(b)$) and the likelihood of e given b :

$$\Pr(b | e) = \frac{\Pr(e | b) \Pr(b)}{\Pr(e | b) \Pr(b) + \Pr(e | \neg b) \Pr(\neg b)}.$$

Whereas beliefs map onto objective states of the world, attitudes are evaluative judgments, reflecting how positively or negatively one feels about an attitude object (e.g., candidate, policy). As attitudes involve subjective evaluations of worth, desirability, or preference, updating the “probability” of an attitude being correct makes little sense. No truth value can be assigned to “liking” a political candidate or “favoring” a policy. Probabilities in a Bayesian framework solely measure degrees of certainty about factual propositions.

It is nonetheless possible to connect Bayesian updating models to attitudes by specifying a *belief–attitude function*. Attitudes, as evaluative judgments, do not have a truth value *per se*, but they often rest on a foundation of beliefs or considerations (Fishbein, 1979). Consider an individual forming an attitude A toward an object (e.g., a policy proposal). Suppose this person holds a set of n beliefs, b_i , about various attributes of the policy, such

as whether the policy is effective, necessary, or costly. Each belief b_i can be represented as the perceived probability that the policy possesses attribute i . Using the Bayesian model, $\Pr(b_i | e)$ can be updated in the light of new evidence e . However, to translate those updated beliefs into an overall evaluative judgment, an *attitude function* is necessary:

$$A = f(\Pr(b_1), \Pr(b_2), \dots, \Pr(b_n)).$$

Here, $f(\cdot)$ is an aggregation rule that maps the updated belief probabilities into a summary evaluation.

A key question concerns how an individual *weights* each belief in forming an attitude. One might assume that all beliefs contribute equally. However, a more psychologically realistic approach allows *relevance weights*, whereby some beliefs matter more than others. Formally, we might write:

$$A = \sum_{i=1}^n w_i g(\Pr(b_i)),$$

where w_i is the weight assigned to belief b_i , and $g(\cdot)$ is a function translating $\Pr(b_i)$ into a value for the attitude calculation. Beliefs that are highly relevant to the attitude would have higher w_i . New evidence can shift $\Pr(b_i)$ via Bayesian updating and the corresponding attitude so long as the belief is heavily weighted. Alternatively, within this framework, attitude polarization may occur if an informational intervention *increases* the probability of holding a congenial belief via “belief backfire” (Nyhan and Reifler, 2010), thereby strengthening the individual’s existing attitude by reinforcing its justification.

The model above bears close resemblance to the classic expectancy value (EV) model of Fishbein and Ajzen (1975, 2010), wherein beliefs are similarly depicted as the basis of attitudes. Beliefs are characterized as the subjective probability that an attitude object has

a particular positive or negative attribute (2010, 96). When, for instance, a voter learns that a candidate for office holds progressive policy commitments, they form a belief (i.e., “that candidate is a progressive”) which links their candidate attitude to the specific attribute and preexisting evaluation e of that attribute. An overall attitude toward the candidate object is produced by summing across all relevant attributes, indexed by i , in a process modeled symbolically as $A \propto \sum e_i Pr(b_i)$. Variations of EV in political science have expanded the theoretical primitives to include not only beliefs but also “considerations,” which blend cognitive and evaluative components (Chong and Druckman, 2007) while using the language of *weights* to characterize the importance of certain considerations in determining attitudes.² Regardless of the precise notation, the EV model and its variants underscore a crucial theoretical point: the extent to which changes in beliefs result in shifts in attitudes depends fundamentally on the relevance of those beliefs to individuals’ judgments. Beliefs deemed irrelevant or peripheral are less likely to alter attitudes, even if convincingly corrected.

To illustrate why understanding the variable attitudinal effects of information requires accounting for relevance at the respondent level, consider attitudes toward immigrants. Motivated by theories of power threat (e.g., Key, 1949; Oliver and Mendelberg, 2000), many have hypothesized that immigration attitudes hinge upon beliefs about the size of the immigrant population, and therefore that correcting persistent overestimates of immigrants’ numeracy should reduce hostility (Citrin and Sides, 2008; Wong, 2007). A substantial body of evidence now suggests that correcting misperceptions about numeracy is ineffective at altering anti-immigrant attitudes, leading a recent paper to conclude that misperceptions may be the consequence rather than the cause of hostility toward immigrants (Hopkins, Sides and Citrin, 2019). More broadly, and more worryingly, Hopkins and colleagues interpret the evidence as suggesting “correct information can change factual perceptions but not necessarily related attitudes” (316).

²In this model, beliefs and attribute evaluations are subsumed under a singular variable v_i , which is an evaluative belief about the object on attribute i . w_i is the subjective weight of the belief, capturing its perceived *relevance* or *importance* to the attitude.

This conclusion serves to bolster theories of motivated reasoning, a rival framework to Bayesian updating that describes how psychological motives may bias the processing of information. While there are multiple interpretations of the framework, the canonical essay by Kunda (1990) identifies two key types of motivation. Accuracy motives drive individuals to seek the most accurate and “unbiased” information to reach an appropriate conclusion. Directional motives, however, lead individuals to collect and interpret evidence that aligns with their preexisting beliefs, preferences, or desired outcomes. This is said to result in the selective acceptance of evidence that supports their worldview and the dismissal or devaluation of conflicting information—such as when corrective information about immigrants’ numeracy fails to alter anti-immigrant sentiments.

Yet, the motivated reasoning interpretation of these results assumes that people heavily weight immigrants’ numeracy when formulating immigration attitudes. Recent studies challenge this notion. Grigorieff, Roth and Ubfal (2020) offer evidence that whereas a bundle of information about immigrants’ characteristics and numeracy alters immigration policy preferences, a lighter treatment containing only information about numeracy does not. Likewise, experimental interventions that supply knowledge about immigration’s impact on crime, jobs, and taxes (Abascal, Huang and Tran, 2021) and about the difficulty of the legal immigration process (Kustov and Landgrave, 2025) have substantially affected immigration attitudes, suggesting that the limited downstream effects of past interventions may be “the result of not providing relevant information” (Kustov and Landgrave, 2025, 1). In Appendix B, we present evidence that respondents who favor restricting immigration are indeed far more likely to justify their attitude using beliefs about immigration’s effects on crime, public services, or economic opportunity than using arguments about immigrants’ numeracy.

Focusing on the potential role of belief relevance helps reconcile some of the seemingly contradictory findings in the factual corrections literature. Though there is ample evidence that corrections reliably improve belief accuracy with little risk of backfire, evi-

dence of attitudinal shifts has been decidedly more mixed. But, if factual correction studies target beliefs that are salient or viral but largely distal to the formation of relevant attitudes for much of the population, this could help explain the consistent observation of belief change without corresponding shifts in attitudes.

Challenges in Identifying Belief Relevance Weights

Researchers use various strategies to identify beliefs for correction, each with distinct theoretical implications. Misinformation scholars typically target beliefs based on their prevalence, drawing on fact-checking databases to identify widely circulating false claims. This approach prioritizes reach over psychological centrality. Other researchers rely on media analysis or researcher intuition about common misperceptions, potentially missing beliefs that are crucial but less visible in public discourse. A more systematic approach, common in attitude research but rare in misinformation studies, uses pilot surveys to identify “modal salient beliefs” within a population before designing interventions. Most recently, network methods have emerged as a tool for mapping belief-attitude relationships, offering a more structured way to identify beliefs that may be particularly consequential for attitude change. Yet these methods often identify belief networks in the population, potentially obscuring important individual heterogeneity in linkages between beliefs and attitudes.

Population-level approaches to identifying influential beliefs—whether based on expert judgment, viral spread, or aggregate network patterns—may miss crucial individual variation in belief relevance. For instance, while concerns about immigration’s fiscal impact may strongly predict immigration attitudes in aggregate data, this belief may be peripheral or irrelevant for many individuals. Even sophisticated network methods that map belief-attitude connections typically do so at the group level (e.g., Fishman and Davis, 2022), potentially obscuring individual-level heterogeneity in belief centrality (Brandt and Morgan, 2022). Moreover, recent studies that attempt to construct belief networks at the individual

level assess relationships between political issues determined *a priori*, which may obscure idiosyncratic connections between beliefs and attitudes (Brandt, 2022). For example, a belief about “universal basic income” may be central to an individual’s views about politics, but failure to include relevant survey questions may produce only a partially observable network.

We propose an alternative method: using large language models (LLMs) to identify personally salient beliefs through direct dialogue with participants. Rather than assuming beliefs relevant at the population level will matter for all individuals, this approach elicits each participant’s own stated reasons for holding their attitudes. By engaging participants in an open-ended discussion about their views, we can identify the specific beliefs they spontaneously invoked to justify their positions—beliefs that are, on the respondent’s own terms, identified as central to their individual attitude structures. Upon identifying important attitude-relevant beliefs, we construct counterarguments targeting those beliefs to assess how this affects downstream attitudes. Moreover, we use a baseline of attitude-relevant beliefs that were not mentioned in the dialogue as a way of assessing how belief relevance affects attitudes while holding topical relevance fixed.

Using our design, we elicit issue attitudes (e.g., supporting DACA, opposing restrictions on automatic firearms) and focal justifications (e.g., improving the educational attainment of all Americans, protecting Second Amendment rights) for each individual in our study, which we use to produce tailored measures of attitude strength and belief strength. Moreover, we instruct an LLM—OpenAI’s GPT-4o—to create personalized corrective messages that target those focal justifications (hereafter, focal beliefs) and more distal justifications reflecting beliefs not disclosed in the open-ended dialogue that nonetheless support the issue attitude (e.g., economic contributions of undocumented immigrants and their children).³ This allows us to directly test whether heavily weighted considerations produce more attitude change than those that are less relevant to the participant’s attitude.

³Specifically, we used the GPT-4o-2024-08-06 model for generating counterarguments and GPT-4o-mini for carrying out the semi-structured chat.

Our primary hypotheses build on prior work in factual corrections, motivated reasoning, and attitude formation. We propose the following sequential expectations:

Assessing the “First Stage” of Belief Change

Hypothesis 1a (Parallel Updating). Studies of factual corrections have increasingly found that people consistently update their beliefs in the direction of evidence, irrespective of their prior beliefs (e.g., Porter, Velez and Wood, 2022; Coppock, 2023). These studies generate an expectation that *focal* and *distal* belief counterarguments will reduce focal and distal belief strength (respectively) relative to control conditions, and to similar degrees.

Individuals presented with a factual argument (either focal or distal) will update their beliefs toward the corrective information to similar degrees.

Hypothesis 1b (Differential Updating). Drawing on theories of motivated reasoning, individuals may process factual arguments differently depending on how central the challenged beliefs are to their broader belief system. While a strong form of motivated reasoning suggests complete resistance or backfire effects (Kunda, 1990; Taber and Lodge, 2006), people may also act as “cautious Bayesians” who update their beliefs to varying degrees based on the centrality and strength of their prior beliefs (Hill, 2017). Under this account:

The magnitude of belief updating will vary based on belief relevance, with focal beliefs showing either no movement or significantly smaller updates compared to distal beliefs. Specifically: (i) For focal beliefs, individuals will either maintain their prior beliefs or exhibit minimal updating in response to factual arguments; (ii) For distal beliefs, which are less central to attitudes, individuals will show larger shifts toward the corrective information.

Assessing the Belief-Attitude Function

Hypothesis 2a (Equal Effectiveness). It may be that *any* counterargument—whether targeting a focal or distal belief—has a comparable effect on attitudes, provided the individual updates and sees the corrected information as relevant enough to reshape their broader evaluation.

Counterarguments to focal and distal beliefs will be equally effective at shifting downstream attitudes.

Hypothesis 2b (Focal Belief Dominance). Alternatively, if focal beliefs are highly weighted in the formation of attitudes, changing these beliefs should be more influential in shifting attitudes than changing more peripheral, distal beliefs.

Factual arguments against focal beliefs (i.e., those identified by participants as central to their opinions) will produce greater changes in downstream attitudes than counterarguments for distal beliefs.

Research Design

Our experimental design uses LLMs at various stages of the process to recover key theoretical quantities and generate experimental stimuli (see Figure 1 for a diagram). Text classification, summarization, and generation are among the strengths of LLMs (Mellon et al., 2022), and previous work has shown that they can be productively integrated into surveys to measure deeply-held beliefs and attitudes (Velez and Liu, 2024) and produce compelling arguments (Bai et al., 2023). We leverage these strengths to (1) elicit issue attitudes; (2) capture justifications for those attitudes (“focal beliefs”); and (3) generate counterarguments that randomly target those “focal beliefs” versus unmentioned, but potentially relevant, beliefs that might support those attitudes (“distal beliefs”).

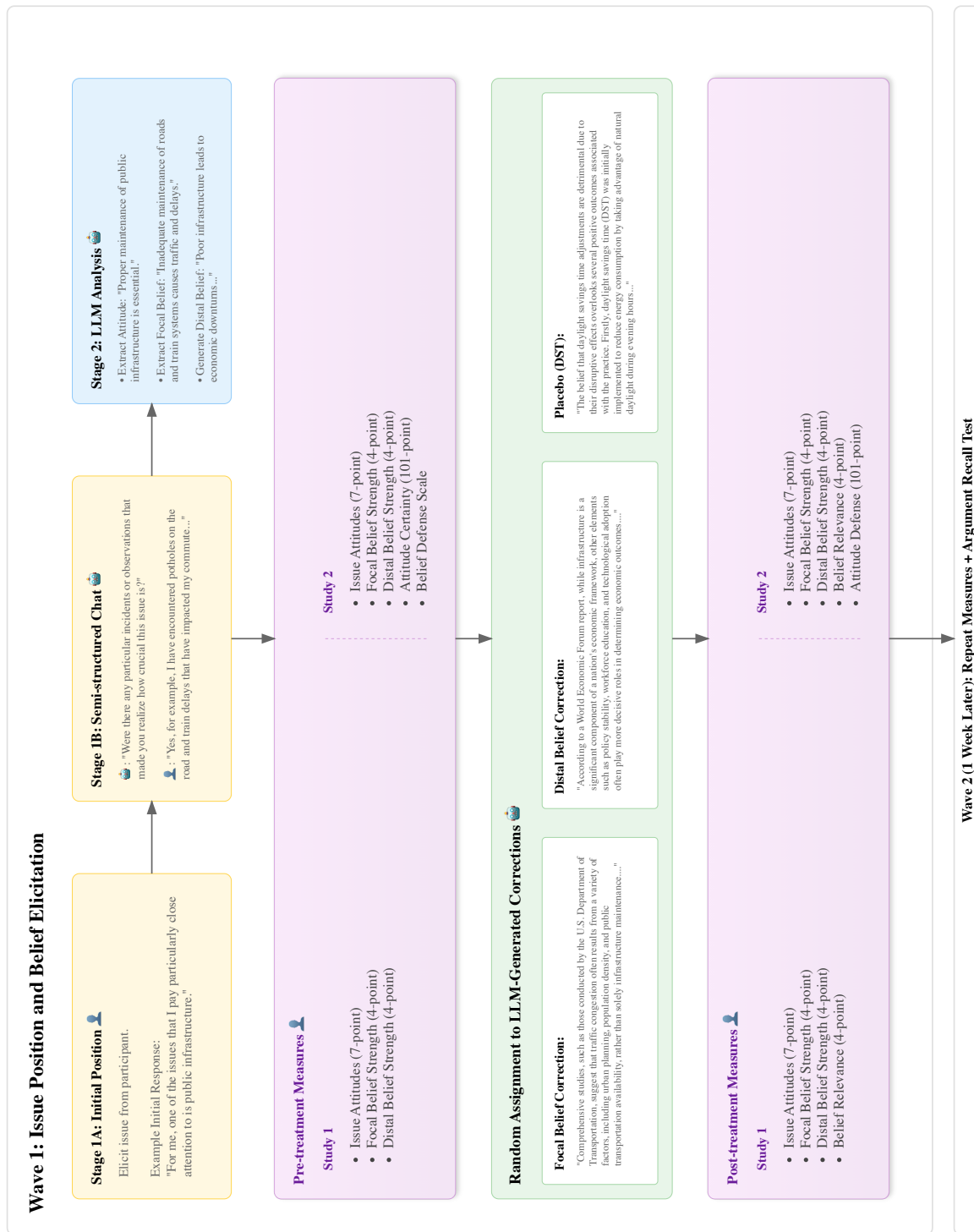


Figure 1: Overview of the Two-Wave Design and Experimental Procedures. In *Wave 1*, participants first provide an issue they care deeply about (*Stage 1A*), then engage in a semi-structured dialogue with an LLM to clarify their justifications (*Stage 1B*). The LLM identifies participants' "focal" beliefs, generates "distal" beliefs, and constructs tailored counterarguments (*Stage 2*). Pre-treatment measures are taken and participants are then randomly assigned to receive one of three messages (focal belief counterargument, distal belief counterargument, or placebo) before completing post-treatment measures. In *Wave 2*, approximately one week later, participants repeat the attitude and belief measures to assess the durability of belief and attitude change.

Across two studies, we carry out a pre-post experimental design comprised of three stages (Clifford, Sheagley and Piston, 2021). First, we elicit issue positions from participants. In Study 1, we randomly assigned participants to either a closed-ended slate of issue positions or open-ended question asking them to report a deeply-held issue position. In Study 2, we solely rely on the latter question. Second, in both studies, we have all participants engage in a semi-structured discussion about their issue position and their justifications for holding the issue position. Once the conversation is complete, these data are passed to an LLM that summarizes the participant’s issue attitude, summarizes their primary justification for holding that attitude (“focal belief”), generates a related, but unmentioned, justification for holding the attitude (“distal belief”), and produces three factual argument vignettes: (1) a vignette countering the “focal belief”; (2) a vignette countering the “distal belief”; and (3) a placebo vignette producing an argument about Daylight Savings Time. Participants were informed of their interactions with AI models in the consent form and debriefing materials.

Following this phase, we measure issue attitudes on a seven-point Likert item ranging from “strongly disagree” to “strongly agree,” focal belief strength on a four-point scale ranging from “not at all accurate” to “very accurate,” and distal belief strength using the same scale. In Study 2, we also measure pre-treatment scores on a “belief defense” scale developed in Velez and Liu (2024). After the pre-treatment measurement step, participants are randomly assigned to see one of the tailored vignettes, with participants equally likely to be assigned to a “focal belief counterargument,” “distal belief counterargument,” or “placebo” condition. These vignettes involved approximately paragraph-long arguments disputing the assigned belief. The LLM was instructed to incorporate logical arguments and data when available.

In the post-treatment phase, we measured attitude strength and belief strength for the focal and distal beliefs. In Study 1, we additionally measured a multi-item “attitude extremity” scale developed in Velez and Liu (2024), whereas in Study 2, we measured “attitude

extremity,” “attitude defense,” and certainty on a 101-point scale. In addition to attitudinal measures, we asked participants to report the relevance of their beliefs to their attitudes on a four-point scale ranging from “not at all relevant” to “very relevant” in Studies 1 and 2. In Study 2, we provided participants with a text entry box to indicate “omitted beliefs” that support their attitude, and a rating scale. The latter measure was included to capture whether participants selectively bolstered unmentioned beliefs as a defense mechanism in the face of corrective information.

In both studies, we carry out a second wave approximately one week after the first wave. In this study, we measure the full set of beliefs and attitudes measured in the initial wave, along with measures of argument recall. Specifically, we provide participants with two randomly selected sentences from the three conditions (i.e., “focal belief counterargument,” “distal belief counterargument,” and placebo) and ask them to identify the sentences that they recall seeing in the previous study. This serves as a test of selective memory, such that participants may forget corrective information that bears on their deeply-held beliefs or attitudes.

Can We Recover Strong Attitudes and Focal Beliefs?

We focus on identifying participants’ deeply-held attitudes. Though this may decrease our ability to detect belief and attitude change, we do so for three important reasons. First, many of the attitudes that are the focus in the corrections literature are often “hot button” issues such as vaccination and immigration. As previous work has shown, these attitudes are generally difficult to move (Kustov, Laaker and Reller, 2021). Second, relative to “non-attitudes” or those that are not very salient to participants, deeply-held attitudes ought to have a richer set of connected considerations (Zaller, 1992) that render it easier to observe how intervening on a belief affects attitudes. Third, focusing on deeply-held attitudes provides a “harder test” of our theoretical predictions, as any observed effects are likely to represent meaningful attitude change rather than temporary shifts in superficial opinions.

To what extent can we recover deeply-held attitudes? As shown in Table 1, we observe attitude strength scores of 4.85 (SE = .01) in Study 1 and 4.84 (SE = .02) in Study 2. Given that attitude strength is measured on a five-point scale, this means that the vast majority of participants select the maximum score, indicating strong attitudes in the aggregate. Indeed, we find that 89% of participants in Study 1 select the maximum score on the Likert item, whereas 88% do so in Study 2. This dovetails with Velez and Liu (2024) who find that LLMs can successfully summarize issue positions and recover deeply-held attitudes that persist over time.

Next, we turn to belief strength. Our theoretical tests hinge on the capability of LLMs to properly summarize focal justifications for attitudes. In Study 1, we find a mean belief strength score of 3.64 (SE = .01), with 72% of participants selecting the maximum score; Study 2 replicates this general pattern, with a mean belief strength score of 3.67 (SE = .02) and 72% of participants selecting the maximum score. Directly comparing focal beliefs to distal beliefs, we observe a difference in belief strength of .32 scale points in Study 1 and .31 scale points in Study 2, suggesting that the focal beliefs retrieved are stronger than distal beliefs.

Additionally, we can assess whether participants see the focal beliefs as relevant to their attitudes. In Study 1, we observe a mean belief relevance score of 3.72 with 76% selecting the maximum option. In Study 2, the mean belief relevance score is 3.84 with 85% selecting the maximum. This can be compared to the distal belief, where scores are approximately .27 scale points lower in Study 1 and .23 in Study 2. Thus, focal beliefs are generally seen as relevant to attitudes, and *more* relevant than distal beliefs.

Overall, the evidence suggests that we are successfully recovering deeply-held attitudes and relevant beliefs. Moreover, the strong performance of the distal beliefs suggests that the LLM is also producing plausible distal beliefs that may support participants' attitudes, even if they are not as central to their reasoning. This provides reassuring evidence

Table 1: Summary Statistics for Variables Across Studies.

Variable	Study 1	Study 2
Attitude Strength	4.85 (0.01); 89.4% selected max	4.84 (0.02); 88.4% selected max
Belief Relevance (Distal)	3.45 (0.02); 60.1% selected max	3.61 (0.03); 70.2% selected max
Belief Relevance (Focal)	3.72 (0.01); 75.7% selected max	3.84 (0.02); 84.6% selected max
Belief Strength (Distal)	3.32 (0.02); 55.3% selected max	3.36 (0.03); 57.3% selected max
Belief Strength (Focal)	3.64 (0.01); 71.9% selected max	3.67 (0.02); 71.6% selected max

Note: Mean (SE) and Percentage Selecting Maximum Option.

that our approach successfully identifies strong attitudes and focal beliefs.

Inducing Belief Change

We now examine whether and how these beliefs respond to corrective information. This analysis provides a possible test of competing expectations about belief change. According to the differential updating hypothesis (H1b), we should observe resistance or even backfire effects when correcting focal beliefs that participants use to justify their attitudes, whereas a parallel updating process (H1a) might hold that counterarguments have similar efficacy regardless of belief strength.

To assess the effects of arguments on focal and distal belief strength, we estimate OLS regression models with HC2 standard errors. The models take the following form:

$$Y_i = \beta_0 + \beta_1 \text{Focal Belief Counterarg}_i + \beta_2 \text{Distal Belief Counterarg}_i + \mathbf{X}_i' \boldsymbol{\gamma} + \epsilon_i \quad (1)$$

where Y_i is post-treatment belief strength for participant i , $\mathbf{X}_i$ is a vector of pre-treatment covariates, and the omitted category is the placebo condition. The coefficients of interest are β_1 and β_2 , which capture the average treatment effects of focal and distal belief arguments relative to the placebo. In Study 1, we adjust for pre-treatment attitude strength, focal belief strength, and distal belief strength. In Study 2, we adjust for the covariates, as well as pre-treatment scores on a multi-item “attitude defense” scale.

Table 2: Effect of Interventions on Belief Strength (Wave 1)

	Study 1	Study 1	Study 2	Study 2
	Focal Belief	Distal Belief	Focal Belief	Distal Belief
Focal Correction	−0.32*** (0.03)	−0.13*** (0.03)	−0.32*** (0.03)	−0.17*** (0.03)
Distal Correction	−0.12*** (0.02)	−0.31*** (0.03)	−0.12*** (0.03)	−0.28*** (0.03)
Num.Obs.	2480	2478	1898	1897
R2	0.406	0.585	0.417	0.544
Covariates Included	✓	✓	✓	✓

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We directly compare effect estimates across two outcomes: focal belief strength and distal belief strength. For each outcome, we estimate separate regression models where the dependent variable is either the focal or distal belief strength. This allows us to test our competing hypotheses in several ways:

First, to assess H1a, we compare β_1 when focal belief strength is the outcome to β_2 when distal belief strength is the outcome. Similar effect sizes would indicate that factual arguments are equally effective regardless of belief type—that is, focal belief arguments move focal beliefs just as much as distal belief arguments. Second, for H1b, we examine whether focal beliefs are more resistant to change than distal beliefs. Under this hypothesis, β_2 might be negative when distal belief strength is the outcome (indicating successful belief change), but β_1 could be positive or null when focal belief strength is the outcome (indicating resistance or backfire). Alternatively, β_1 could be negative but smaller in absolute magnitude than β_2 , consistent with a “cautious Bayesian” view.

Turning to Table 2, we find evidence that is more consistent with H1a. For example, in the first wave of Study 1, we find that focal belief arguments reduce focal belief strength by .32 scale points and distal belief strength by .12 scale points relative to control (SE = .03; $p < .001$). Turning to distal belief strength, we observe similar effect sizes for the distal

Table 3: Effect of Interventions on Belief Strength (Wave 2)

	Study 1	Study 1	Study 2	Study 2
	Focal Belief	Distal Belief	Focal Belief	Distal Belief
Focal Correction	−0.24*** (0.04)	−0.10** (0.04)	−0.26*** (0.04)	−0.13** (0.04)
Distal Correction	−0.07* (0.03)	−0.15*** (0.04)	−0.08* (0.03)	−0.23*** (0.04)
Num.Obs.	2053	2051	1679	1679
R2	0.200	0.343	0.264	0.355
Covariates Included	✓	✓	✓	✓

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

argument, with reductions in distal belief strength of .31 scale points ($SE = .03$; $p < .001$) and .13 scale points ($SE = .03$; $p < .001$) in focal belief strength. In the first wave of Study 2, we find comparable effect sizes, with focal belief counterarguments reducing focal belief strength by .26 scale points ($SE = .04$; $p < .001$) and distal belief arguments reducing distal belief strength by .23 scale points ($SE = .04$; $p < .001$). Thus, focal arguments are approximately just as effective in moving focal beliefs as distal arguments are in moving distal beliefs.

Importantly, these effects persist into the second wave of each study, approximately one week after the initial exposure. In Study 1, focal belief counterarguments continue to reduce focal belief strength by .24 scale points ($SE = .04$; $p < .001$). Similarly, distal belief arguments maintain their impact on distal belief strength though the effect is somewhat smaller in magnitude ($\beta = -.15$, $SE = .04$; $p < .001$). Study 2 replicates this pattern of persistence: focal belief arguments continue to reduce focal belief strength by .26 scale points ($SE = .04$; $p < .001$), while distal belief arguments maintain their effect on distal belief strength ($\beta = -.23$, $SE = .04$; $p < .001$). The durability of these effects across both studies suggests that factual arguments produce lasting changes in belief strength, rather than reflecting momentary shifts arising due to other processes such as demand effects.

The similarity in effect sizes between focal and distal belief counterarguments suggests that both types of beliefs are movable and that arguments targeting either type of belief can have spillover effects on the other, which is inconsistent with H1b. Moreover, our finding that focal argument effects exhibit more persistence than distal argument effects suggests that interventions targeting beliefs central to attitudes produce *more* durable belief change than those targeting peripheral beliefs—perhaps because they are more relevant to deeply important attitudes. This provides additional evidence against motivational accounts that predict greater resistance to updating focal beliefs, while also underscoring an important pattern: the apparent equivalence between focal and distal belief arguments may only hold in the short term.

Inducing Attitude Change

Demonstrating the effectiveness of factual arguments on beliefs, we assess whether those targeting focal beliefs exhibit larger effects on attitudes relative to those targeting distal beliefs. In the first wave of Study 1, the focal counterargument reduces attitude strength by 0.12 scale points ($SE = 0.02$, $p < 0.001$), while the distal argument has approximately half the effect, decreasing strength by 0.06 scale points ($SE = 0.02$, $p < 0.01$). Turning to attitude extremity, we find a .05 scale point reduction ($SE = .07$; $p = \text{n.s.}$) in scores for the focal argument and .01 scale point increase for the distal argument ($SE = .07$; $p = \text{n.s.}$). Though neither are significant, the focal argument moves estimates in the expected direction. In the first wave of Study 2, we largely replicate the findings from Study 1. The focal argument reduces attitude strength by .13 scale points ($SE = .03$; $p < .001$) relative to the distal argument, which produces an effect estimate of .09 scale points ($SE = .03$; $p < .001$). Turning to attitude extremity, we find modest evidence that focal arguments reduce extremity by .06 scale points ($SE = .04$, $p < .10$), while distal arguments have no discernible effect with estimates hovering around zero ($SE = .04$, $p = \text{n.s.}$). Focusing on our additional attitudinal measures, we detect a -3.7 scale point reduction on a 101-point attitude certainty

Table 4: Effect of Interventions on Attitudes (Wave 1)

	Study 1 Strength	Study 1 Extremity	Study 2 Strength	Study 2 Extremity	Study 2 Certainty	Study 2 Defense
Focal Correction	-0.12*** (0.02)	-0.05 (0.07)	-0.13*** (0.03)	-0.06+ (0.04)	-3.73*** (0.68)	-0.13** (0.05)
Distal Correction	-0.06** (0.02)	0.01 (0.07)	-0.09*** (0.03)	0.001 (0.04)	-2.16*** (0.65)	-0.09+ (0.05)
Num.Obs.	2479	2482	1900	1902	1897	1901
R2	0.366	0.026	0.271	0.117	0.279	0.695
Covariates Included	✓	✓	✓	✓	✓	✓

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

scale ($SE = .68$; $p < .001$) among those exposed to the focal argument, whereas this estimate is -2.16 ($SE = .65$; $p < .001$) among those exposed to the distal argument. Finally, we observe a -.13 scale point reduction in “attitude defense” ($SE = .06$; $p < .01$) among those exposed to the focal argument, which is .04 scale points larger than the effect we observe for the distal argument ($\beta_2 = -.09$; $SE = .05$; $p < .10$).

Overall, effect sizes tend to be small, reflecting the choice to target deeply-held attitudes. Most importantly, however, we find consistent evidence that focal belief counterarguments produce larger immediate attitude change than distal belief counterarguments across multiple measures and studies.

Turning now to our second wave, we find that effects on attitudes generally weaken. In Study 1, while the immediate effects on attitude strength dissipate ($\beta = -.02$, $SE = .03$, $p = \text{n.s.}$ for focal arguments; $\beta = .03$, $SE = .03$, $p = \text{n.s.}$ for distal arguments), we continue to observe modest reductions in attitude extremity, though these effects do not reach conventional levels of statistical significance ($\beta = -.10$, $SE = .07$, $p = \text{n.s.}$ for focal arguments; $\beta = -.11$, $SE = .07$, $p = \text{n.s.}$ for distal arguments). Study 2 reveals a similar pattern of attenuation, but with some notable exceptions. While effects of the focal argument on attitude strength and extremity weaken ($\beta_1 = -.06$, $SE = .04$, $p = \text{n.s.}$ for

Table 5: Effect of Interventions on Attitudes (Wave 2)

	Study 1 Strength	Study 1 Extremity	Study 2 Strength	Study 2 Extremity	Study 2 Certainty	Study 2 Defense
Focal Correction	-0.02 (0.03)	-0.10 (0.07)	-0.06 (0.04)	-0.14+ (0.08)	-2.57*** (0.71)	-0.10 (0.07)
Distal Correction	0.03 (0.03)	-0.11 (0.07)	0.02 (0.03)	-0.003 (0.08)	-1.00 (0.72)	-0.04 (0.07)
Num.Obs.	2056	2055	1687	1680	1678	1680
R2	0.104	0.026	0.112	0.079	0.241	0.447
Covariates Included	✓	✓	✓	✓	✓	✓

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

attitude strength; $\beta_1 = -.14$, $SE = .08$, $p < .10$ for extremity), focal arguments continue to reduce attitude certainty by 2.57 points on our 101-point scale ($SE = .71$, $p < .001$). In contrast, the effect of distal belief arguments on certainty is no longer significant ($\beta_2 = -1.00$, $SE = .72$, $p = \text{n.s.}$). Effects on attitude defense show a similar pattern of persistence for focal arguments ($\beta_1 = -.10$, $SE = .07$, $p = \text{n.s.}$) compared to distal arguments ($\beta_2 = -.04$, $SE = .07$, $p = \text{n.s.}$), though neither reaches conventional levels of statistical significance. Taken together, we observe mixed evidence of persistence, but generally observe larger effect estimates in Wave 2 for focal counterarguments relative to distal counterarguments.

Assessing Differences in Attitude Change

We now conduct an exploratory analysis to formally assess differences between focal and distal counterarguments on attitudes. While focal arguments generally produce larger effect estimates, examining each attitudinal measure separately may obscure the overall pattern of effects and reduce our ability to detect meaningful differences between argument types. A meta-analytic approach allows us to systematically compare the relative effectiveness of focal versus distal arguments while accounting for both within- and between-measure heterogeneity in effect sizes.

We employ a random effects meta-analysis for several reasons. First, our attitudinal measures—though related—capture distinct theoretical constructs ranging from strength to extremity to defense. We should therefore expect some variation in the true effect sizes across measures. Second, these measures operate on different scales (e.g., 4-point versus 101-point), which could produce mechanical differences in effect magnitudes. Third, we collected these measures across two independent studies and waves, introducing another potential source of effect heterogeneity.

Pooling across all attitudinal measures, we find that focal arguments produce significantly larger effects than distal belief arguments. In 11 out of 12 comparisons, we observe larger effect sizes for the focal argument relative to the distal argument. The average difference in treatment effects is $-.06$ control-group standard deviation units ($SE = .02$, $p < .001$). We fail to reject the null hypothesis of effect homogeneity ($Q(11) = 8.79$, $p = .64$), suggesting that the effect of focal arguments is comparable across different attitudinal measures, waves, and studies. Looking at individual comparisons, the largest differences appear in attitude certainty, where focal arguments outperform distal arguments by roughly $.10$ standard deviation units in both Wave 1 ($d = -.104$, $SE = .050$) and Wave 2 ($d = -.105$, $SE = .052$) of Study 2. The consistency of positive differences between the focal and distal arguments provides support for H2b (Focal Belief Dominance) while suggesting that larger attitudinal effects due to targeting focal beliefs generalizes across different measures.

Mechanisms of Attitude Change: Relevance and Recall

Belief relevance In addition to belief strength, we consider auxiliary outcomes that help clarify why focal arguments generally have larger and more persistent effects than distal arguments. First, we focus on belief relevance, which is measured on a four-point scale capturing the relevance of focal and distal beliefs to attitudes. We regress belief relevance for the two relevance outcomes on treatment indicators and pre-treatment covariates with HC2 robust standard errors.

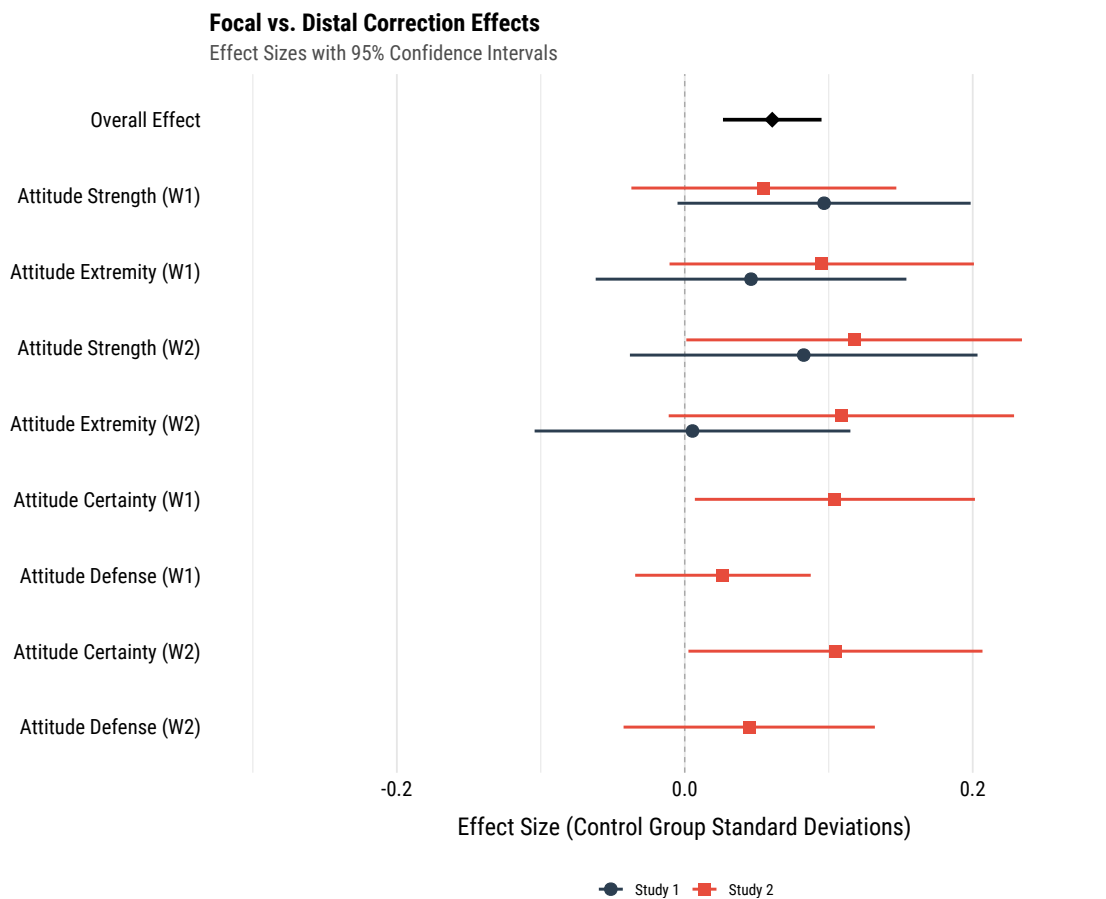


Figure 2: Results from a random effects meta-analysis with correlated sampling errors showing mean differences between the focal and distal argument conditions across two studies. Points show the estimated effects and horizontal lines represent 95% confidence intervals. The diamond shows the overall random effects meta-analytic estimate.

Table 6: Effect of Interventions on Belief Relevance

	Study 1 Focal Relevance	Study 1 Distal Relevance	Study 2 Focal Relevance	Study 2 Distal Relevance
Focal Correction	-0.17*** (0.02)	-0.12*** (0.03)	-0.09** (0.03)	-0.06+ (0.04)
Distal Correction	-0.04+ (0.02)	-0.20*** (0.03)	0.005 (0.03)	-0.13*** (0.04)
Num.Obs.	2473	2474	1679	1679
R2	0.202	0.366	0.160	0.227
Covariates Included	✓	✓	✓	✓

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Turning to Table 6, we find evidence that participants decrease belief relevance when exposed to the corresponding argument. In Wave 1 of Study 1, focal arguments decrease focal belief relevance by -.167 scale points ($SE = .025$; $p < .001$) and distal arguments decrease distal belief relevance by -.205 ($SE = .032$; $p < .001$), with some evidence of spillover on focal belief relevance ($\hat{ATE} = -.120$; $SE = .030$; $p < .001$). We replicate these findings in Study 2, with reductions of .09 (.13) scale points in focal (distal) relevance among those assigned to the focal (distal) argument.

On their face, these findings are consistent with a “bolstering” process (Jervis, 2006, 653), where once salient considerations are supplanted by alternative considerations, leading individuals to discount their previous reliance on these beliefs when reporting their attitudes. However, it is worth noting that the self-reports of relevance and belief strength are intertwined, with moderate correlations between the two.⁴ A possible interpretation is that *self-reported* belief relevance may simply be another proxy for belief strength, as it could be challenging for participants to faithfully return measures of internal belief weights.

To directly assess the bolstering mechanism, we included a text entry option in Study 2 when asking about belief relevance. In addition to presenting them with focal and

⁴In the initial wave of Studies 1 and 2, the correlation between focal belief strength and self-reported relevance is .40 and .46, respectively. The correlation between distal strength and relevance is .59 and .58, respectively.

distal beliefs, we told participants “if there are more important reasons why you hold this view that are not listed, please tell us using the *Other* text box.” If participants are simply substituting a corrected belief for another, we ought to see additional considerations being offered in the counterargument conditions compared to the control condition. This would provide direct evidence of the theorized process whereby individuals compensate for weakened beliefs by elevating the importance of alternative considerations. However, turning to Table 7, we fail to find evidence that this is what participants are doing, with small estimates hovering around zero for both experimental conditions ($p > .05$). This suggests that when participants’ beliefs are weakened by corrective information, they simultaneously downgrade both their strength in these beliefs and their perceived relevance, without necessarily “propping up” alternative beliefs to fill this explanatory void.

Table 7: Effect of Interventions on Stated Relevance of Alternative Beliefs (Study 2)

	Mentioned Other Reason	Other Reason Relevance
Focal Belief Treatment	0.01 (0.02)	−0.05 (0.07)
Distal Belief Treatment	0.03 (0.02)	−0.02 (0.07)
Num.Obs.	1902	1902
R2	0.020	0.017
Covariates Included	✓	✓

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Argument recall Previous research on motivated reasoning has stressed the importance of biased recall, such that people more easily retrieve pro-attitudinal considerations than counter-attitudinal considerations (Taber and Lodge, 2006). While our design did not directly ask participants to provide pro-attitudinal arguments, we can test whether participants exhibit selective memory when recalling arguments that bear on their focal beliefs. We assess this in the second waves of Study 1 and 2 by asking participants to indicate whether they had seen two arguments drawn from their assigned condition and two from the other

Table 8: Effect of Interventions on Argument Recall

	Study 1	Study 1	Study 1	Study 2	Study 2	Study 2
	Focal	Distal	Assigned	Focal	Distal	Assigned
	Recall	Recall	Recall	Recall	Recall	Recall
Focal Correction	0.15*** (0.02)	0.04* (0.02)	0.09*** (0.02)	0.14*** (0.02)	0.06* (0.02)	0.09*** (0.03)
Distal Correction	0.002 (0.02)	0.11*** (0.02)	0.05* (0.02)	0.008 (0.02)	0.15*** (0.02)	0.05* (0.02)
Num.Obs.	2072	2076	2072	1660	1654	1655
R2	0.030	0.013	0.009	0.032	0.028	0.009
Covariates Included				✓	✓	✓

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

conditions each in the last month.⁵ Our evidence suggests slightly higher rates of recall for the focal versus distal belief condition in both studies, such that participants are 4pp more likely to recall an assigned argument in the focal condition ($\hat{ATE} = .09$; SE = .02; p < .001 in Study 1; $\hat{ATE} = .09$; SE = .03; p < .001 in Study 2) relative to the distal condition ($\hat{ATE} = .05$; SE = .02; p < .05; $\hat{ATE} = .05$; SE = .02 in Study 1; p < .05 in Study 2). This pattern of enhanced recall for focal belief counterarguments may help explain their greater persistence in producing attitude change relative to distal counterarguments. The heightened memorability of focal arguments suggests deeper engagement with information that challenges core justifications for attitudes, rather than a reflexive defensiveness against attitude-incongruent information.

Discussion

Our findings shed new light on the relationship between beliefs and attitudes. By using LLMs to recover and confront beliefs that participants identified as central to their policy views, we provide experimental evidence that both “focal” and “distal” beliefs are capable of being

⁵For example, those assigned to the focal counterargument condition would see two statements from focal counterargument, two statements from the distal argument, and two placebo argument statements.

corrected. Across two studies, participants updated their beliefs in the direction of the belief-incongruent information, running counter to expectations from approaches that strongly emphasize resistance or bias in the face of exposure to belief-incongruent information. The durability of these shifts – observed one week later – provides further support for a view of persuasion vis-a-vis beliefs that does not align neatly with approaches predicting extreme resistance to persuasion.

Focal counterarguments exerted larger and more durable effects on downstream attitudes than arguments targeting relatively peripheral beliefs. We observed stronger immediate reductions in attitudinal measures for participants who received a counterargument targeting beliefs they had explicitly described as key justifications for their position. In contrast, counterarguments to more distal beliefs that were unmentioned by the participant but nonetheless relevant to the attitude yielded smaller or inconsistent attitudinal effects. This dovetails with the notion that beliefs differ in their relevance to broader evaluations: changing a belief that participants see as important to their political views appears to produce more discernible evidence of persuasion.

The greater memorability of focal counterarguments may help explain their persistence. Participants assigned to focal counterarguments recognized the arguments at higher rates, suggesting deeper engagement with these arguments. Throughout our studies, we found little evidence of widespread defensiveness in our studies. Participants in focal counterargument conditions did not uniformly strengthen alternative beliefs or discount the credibility of the argument. While we observed decreases in belief relevance, we did not find evidence that alternative considerations were “bolstered” to compensate for the weakened belief.

These findings help reconcile an apparent puzzle in the fact-checking literature: why does successful belief change often yield minimal attitude change? Our results suggest this pattern may stem partly from informational interventions that target malleable but relatively

peripheral beliefs. Many popular fact-checks address viral claims on social media platforms but may not reflect the most relevant beliefs underlying individuals' political attitudes. The methodological approach developed here—using LLMs to identify personally salient beliefs—provides a framework for designing more effective arguments by targeting beliefs that individuals themselves identify as central to their political evaluations.

Several questions merit further investigation. First, while we focused on strongly held political attitudes to provide a stringent test of our theoretical expectations, future work might examine whether the relative effectiveness of focal versus distal counterarguments generalizes to less crystallized attitudes. Second, our finding that arguments reduce both belief strength and perceived relevance raises questions about the relationship between these constructs and their distinct roles in attitude formation. Finally, additional research could explore whether providing multiple arguments targeting different types of beliefs produces cumulative effects on attitudes, as argued by Porter and Wood (2024*b*).

Our findings extend beyond fact-checking to pivotal questions about political cognition and attitude change, suggesting that citizens can meaningfully update both peripheral and central political beliefs when presented with belief-incongruent information. However, the translation of these changes in beliefs into attitude change appears to depend critically on the centrality of the corrected beliefs. This highlights the importance of understanding how individuals weight different beliefs in forming political attitudes and points toward strategies for more effective belief-based persuasion.

References

- Abascal, Maria, Tiffany J. Huang and Van C. Tran. 2021. “Intervening in Anti-Immigrant Sentiments: The Causal Effects of Factual Information on Attitudes toward Immigration.” *The ANNALS of the American Academy of Political and Social Science* 697(1):174–191. Publisher: SAGE Publications Inc.
URL: <https://doi.org/10.1177/00027162211053987>
- Alt, James E., David D. Lassen and John Marshall. 2016. “Credible Sources and Sophisticated Voters: When Does New Information Induce Economic Voting?” *The Journal of Politics* 78(2):327–342. Publisher: The University of Chicago Press.
URL: <https://www.journals.uchicago.edu/doi/10.1086/683832>
- Anderson, Norman H. 1981. *Foundations of Information Integration Theory*. Academic Press. Google-Books-ID: 3XN9AAAAMAAJ.
- Arias, Eric, Horacio Larreguy, John Marshall and Pablo Querubín. 2022. “Priors Rule: When Do Malfeasance Revelations Help Or Hurt Incumbent Parties?” *Journal of the European Economic Association* 20(4):1433–1477.
URL: <https://doi.org/10.1093/jeea/jvac015>
- Bai, (Max) Hui, Jan G. Voelkel, johannes C. Eichstaedt and Robb Willer. 2023. “Artificial Intelligence Can Persuade Humans on Political Issues.”
URL: <https://osf.io/stakv/>
- Barnes, Lucy, Avi Feller, Jake Haselswerdt and Ethan Porter. 2018. “Information, Knowledge, and Attitudes: An Evaluation of the Taxpayer Receipt.” *The Journal of Politics* 80(2):701–706. Publisher: The University of Chicago Press.
URL: <https://www.journals.uchicago.edu/doi/abs/10.1086/695672>
- Brandt, Mark J. 2022. “Measuring the belief system of a person.” *Journal of Personality and Social Psychology* 123(4):830.
- Brandt, Mark J. and G. Scott Morgan. 2022. “Between-person methods provide limited insight about within-person belief systems.” *Journal of Personality and Social Psychology* 123(3):621–635.
- Chong, Dennis and James N. Druckman. 2007. “Framing Theory.” *Annual Review of Political Science* 10(Volume 10, 2007):103–126. Publisher: Annual Reviews.
URL: <https://www.annualreviews.org/content/journals/10.1146/annurev.polisci.10.072805.103054>
- Citrin, Jack and John Sides. 2008. “Immigration and the Imagined Community in Europe and the United States.” *Political Studies* 56(1):33–56. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9248.2007.00716.x>.
URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9248.2007.00716.x>
- Clifford, Scott, Geoffrey Sheagley and Spencer Piston. 2021. “Increasing Precision without Altering Treatment Effects: Repeated Measures Designs in Survey Experiments.” *American Political Science Review* 115(3):1048–1065.

URL: <https://www.cambridge.org/core/journals/american-political-science-review/article/increasing-precision-without-altering-treatment-effects-repeated-measures-designs-in-survey-experiments/2E1637FFE36DFD0BA4AD0822634D8337>

Converse, Philip E. 1964. The Nature of Belief Systems in Mass Publics. In *Ideology and Discontent*, ed. D. E. Apter. Ann Arbor: University of Michigan Press pp. 206–261.

URL: <http://www.tandfonline.com/doi/abs/10.1080/08913810608443650>

Cook, Fay Lomax, Lawrence R. Jacobs and Dukhong Kim. 2010. “Trusting What You Know: Information, Knowledge, and Confidence in Social Security.” *The Journal of Politics* 72(2):397–412. Publisher: [The University of Chicago Press, Southern Political Science Association].

URL: <https://www.jstor.org/stable/10.1017/s0022381610000034>

Coppock, Alexander. 2023. *Persuasion in Parallel: How Information Changes Minds about Politics*. Chicago Studies in American Politics Chicago, IL: University of Chicago Press.

URL: <https://press.uchicago.edu/ucp/books/book/chicago/P/bo181475008.html>

Coppock, Alexander, Kimberly Gross, Ethan Porter, Emily Thorson and Thomas J. Wood. 2023. “Conceptual Replication of Four Key Findings about Factual Corrections and Misinformation during the 2020 US Election: Evidence from Panel-Survey Experiments.” *British Journal of Political Science* 53(4):1328–1341.

URL: https://www.cambridge.org/core/product/identifier/S0007123422000631/type/journal_article

Dias, Nicholas C, Laurits F Aarslew, Kristian Vrede Skaaning Frederiksen, Yphtach Lelkes, Lea Pradella and Sean J Westwood. 2024. “Correcting misperceptions of partisan opponents is not effective at treating democratic ills.” *PNAS Nexus* 3(8):pgae304.

URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11333104/>

Eagly, Alice Hendrickson and Shelly Chaiken. 1993. *The Psychology of Attitudes*. Harcourt Brace Jovanovich College Publishers. Google-Books-ID: oPYMAQAAMAAJ.

Fearon, James D. 1999. Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance. In *Democracy, Accountability, and Representation*, ed. Adam Przeworski, Bernard Manin and Susan C. Stokes. Cambridge Studies in the Theory of Democracy Cambridge: Cambridge University Press pp. 55–97.

URL: <https://www.cambridge.org/core/books/democracy-accountability-and-representation/electoral-accountability-and-the-control-of-politicians-selecting-good-types-versus-sanctioning-poor-performance/BA4FE99774A89B68A3E307F887A7BB49>

Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Stanford University Press. Pages: xi, 291.

Fishbein, M. 1979. “A theory of reasoned action: Some applications and implications.” *Nebraska Symposium on Motivation* 27:65–116. Place: US Publisher: University of Nebraska Press.

Fishbein, Martin and Icek Ajzen. 1975. *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Addison-Wesley Publishing Company. Google-Books-ID: 8o0QAQAIAAJ.

- Fishbein, Martin and Icek Ajzen. 2010. *Predicting and Changing Behavior: The Reasoned Action Approach*. New York: Taylor & Francis.
- Fishman, Nic and Nicholas T. Davis. 2022. “Change We Can Believe In: Structural and Content Dynamics within Belief Networks.” *American Journal of Political Science* 66(3):648–663. [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12626](https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12626).
URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12626>
- Gerber, Alan and Donald Green. 1999. “Misperceptions about Perceptual Bias.” *Annual Review of Political Science* 2(Volume 2, 1999):189–210. Publisher: Annual Reviews.
URL: <https://www.annualreviews.org/content/journals/10.1146/annurev.polisci.2.1.189>
- Grigorieff, Alexis, Christopher Roth and Diego Ubfal. 2020. “Does Information Change Attitudes Toward Immigrants?” *Demography* 57(3):1117–1143. Publisher: [Population Association of America, Duke University Press].
URL: <https://www.jstor.org/stable/48681395>
- Helfers, Anna and Mirjam Ebersbach. 2023. “The differential effects of a governmental debunking campaign concerning COVID-19 vaccination misinformation.” *Journal of Communication in Healthcare* 16(1):113–121. Publisher: Taylor & Francis [_eprint: https://doi.org/10.1080/17538068.2022.2047497](https://doi.org/10.1080/17538068.2022.2047497).
URL: <https://doi.org/10.1080/17538068.2022.2047497>
- Hill, Seth J. 2017. “Learning Together Slowly: Bayesian Learning about Political Facts.” *The Journal of Politics* 79(4):1403–1418. Publisher: The University of Chicago Press.
URL: <https://doi.org/10.1086/692739>
- Hochschild, Jennifer L. and Katherine Levine Einstein. 2015. “Do Facts Matter? Information and Misinformation in American Politics.” *Political Science Quarterly* 130(4):585–624. Publisher: [The Academy of Political Science, Wiley].
URL: <https://www.jstor.org/stable/43828727>
- Hopkins, Daniel J., John Sides and Jack Citrin. 2019. “The Muted Consequences of Correct Information about Immigration.” *The Journal of Politics* 81(1):315–320. Publisher: The University of Chicago Press.
URL: <https://www.journals.uchicago.edu/doi/10.1086/699914>
- Jervis, Robert. 2006. “Understanding Beliefs.” *Political Psychology* 27(5):641–663.
- Key, V. O. 1949. *Southern Politics in State and Nation*. New York: A.A. Knopf.
URL: http://archive.org/details/southernpolitics0000keyv_d3i3
- Kuklinski, James H., Paul J. Quirk, David W. Schwieder and Robert F. Rich. 1998. ““Just the Facts, Ma’am”: Political Facts and Public Opinion.” *The Annals of the American Academy of Political and Social Science* 560:143–154. Publisher: [Sage Publications, Inc., American Academy of Political and Social Science].
URL: <https://www.jstor.org/stable/1048982>
- Kuklinski, James H., Paul J. Quirk, Jennifer Jerit, David Schwieder and Robert F. Rich. 2000. “Misinformation and the Currency of Democratic Citizenship.” *The Journal of*

- Politics* 62(3):790–816. Publisher: The University of Chicago Press.
URL: <https://doi.org/10.1111/0022-3816.00033>
- Kunda, Ziva. 1990. “The Case for Motivated Reasoning.” *Psychological bulletin* 108(3):480. Publisher: American Psychological Association.
- Kustov, Alexander, Dillon Laaker and Cassidy Reller. 2021. “The Stability of Immigration Attitudes: Evidence and Implications.” *The Journal of Politics* 83(4):1478–1494. Publisher: The University of Chicago Press.
URL: <https://www.journals.uchicago.edu/doi/10.1086/715061>
- Kustov, Alexander and Michelangelo Landgrave. 2025. “Immigration is difficult?! Informing voters about immigration policy fosters pro-immigration views.” *Journal of Experimental Political Science* pp. 1–13.
URL: <https://www.cambridge.org/core/journals/journal-of-experimental-political-science/article/immigration-is-difficult-informing-voters-about-immigration-policy-fosters-proimmigration-views/464D2A994E38A25EE49B1464C6729773>
- Kuziemko, Ilyana, Michael I. Norton, Emmanuel Saez and Stefanie Stantcheva. 2015. “How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments.” *American Economic Review* 105(4):1478–1508.
URL: <https://www.aeaweb.org/articles?id=10.1257/aer.20130360>
- Lawrence, Eric D and John Sides. 2014. “The consequences of political innumeracy.” *Research & Politics* 1(2):2053168014545414. Publisher: SAGE Publications Ltd.
URL: <https://doi.org/10.1177/2053168014545414>
- Lodge, Milton and Charles S. Taber. 2013. *The Rationalizing Voter*. Cambridge Studies in Public Opinion and Political Psychology Cambridge: Cambridge University Press.
URL: <https://www.cambridge.org/core/books/rationalizing-voter/9E4E27965B612D5DDB1091BD907DF492>
- Mellon, Jonathan, Jack Bailey, Ralph Scott, James Breckwoldt and Marta Miori. 2022. “Does GPT-3 know what the Most Important Issue is? Using Large Language Models to Code Open-Text Social Survey Responses At Scale.”
URL: <https://papers.ssrn.com/abstract=4310154>
- Nyhan, Brendan, Ethan Porter, Jason Reifler and Thomas J. Wood. 2020. “Taking Fact-Checks Literally But Not Seriously? The Effects of Journalistic Fact-Checking on Factual Beliefs and Candidate Favorability.” *Political Behavior* 42(3):939–960.
URL: <https://doi.org/10.1007/s11109-019-09528-x>
- Nyhan, Brendan and Jason Reifler. 2010. “When corrections fail: The persistence of political misperceptions.” *Political Behavior* 32(2):303–330.
- Oliver, J. Eric and Tali Mendelberg. 2000. “Reconsidering the Environmental Determinants of White Racial Attitudes.” *American Journal of Political Science* 44(3):574–589. Publisher: [Midwest Political Science Association, Wiley].
URL: <https://www.jstor.org/stable/2669265>

- Petty, Richard E. 2024. “What Is Unique About Acceptance and Correction of Misinformation? Insights from Work on Attitudes, Persuasion, and Beyond.” *American Psychologist* 79(9):1275–1288.
URL: <https://doi.apa.org/doi/10.1037/amp0001447>
- Porter, Ethan and Thomas J Wood. 2024a. “Factual corrections: Concerns and current evidence.” *Current Opinion in Psychology* 55:101715.
- Porter, Ethan and Thomas J. Wood. 2024b. “Factual corrections: Concerns and current evidence.” *Current Opinion in Psychology* 55:101715.
URL: <https://www.sciencedirect.com/science/article/pii/S2352250X23001604>
- Porter, Ethan, Yamil R. Velez and Thomas J. Wood. 2022. “Factual Corrections Eliminate False Beliefs About COVID-19 Vaccines.” *Public Opinion Quarterly* 86(3):762–773.
URL: <https://doi.org/10.1093/poq/nfac034>
- Porter, Ethan, Yamil Velez and Thomas J Wood. 2023. “Correcting COVID-19 vaccine misinformation in 10 countries.” *Royal Society open science* 10(3):221097.
- Taber, Charles S. and Milton Lodge. 2006. “Motivated Skepticism in the Evaluation of Political Beliefs.” *American Journal of Political Science* 50(3):755–769. Publisher: John Wiley & Sons, Ltd.
URL: <https://doi.org/10.1111/j.1540-5907.2006.00214.x>
- Velez, Yamil Ricardo and Patrick Liu. 2024. “Confronting Core Issues: A Critical Assessment of Attitude Polarization Using Tailored Experiments.” *American Political Science Review* pp. 1–18.
URL: <https://www.cambridge.org/core/journals/american-political-science-review/article/confronting-core-issues-a-critical-assessment-of-attitude-polarization-using-tailored-experiments/0753FAC7590FC8D7F99ED413B367E74B>
- Walter, Nathan, Jonathan Cohen, R Lance Holbert and Yasmin Morag. 2020. “Fact-checking: A meta-analysis of what works and for whom.” *Political communication* 37(3):350–375.
- Wong, Cara J. 2007. ““Little” and “Big” Pictures in Our Heads: Race, Local Context, and Innumeracy about Racial Groups in the United States.” *The Public Opinion Quarterly* 71(3):392–412. Publisher: [Oxford University Press, American Association for Public Opinion Research].
URL: <https://www.jstor.org/stable/4500384>
- Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. Cambridge Studies in Public Opinion and Political Psychology Cambridge: Cambridge University Press.
URL: <https://www.cambridge.org/core/books/nature-and-origins-of-mass-opinion/70B1485D3A9CFF55ADCCDD42FC7E926A>