Breaking Through the Information Bubble: How Surprise Shapes Belief Updating Across Media Sources*

Egor Bronnikov Elias Tsakas Alexander Vostroknutov Kaj Thomsson

[Check the latest version of this paper.]

March, 2026

Abstract

Increasing political polarization in the United States has raised the question of whether meaningful belief updating across partisan lines is systematically constrained. While partisan-motivated reasoning provides one explanation, we examine a complementary driver of belief updating—information-theoretic *surprise*. In two experiments conducted shortly before the 2024 U.S. presidential election, participants first reported their prior beliefs about the election outcome, were then exposed to identical political forecasts within each study with the attributed media source randomly assigned to *The New York Times* or *Fox News*, and subsequently reported their posterior beliefs. We find that belief updating remains strongly anchored in prior beliefs, with limited direct media source effects, and is strongly shaped by the extent to which signals deviate from individuals' baseline beliefs, including when attributed to out-group sources. These findings suggest that expectation-violating information may facilitate belief updating, pointing to new avenues for policy interventions in polarized societies.

Keywords: Partisanship, information provision, media, surprise, polarization, belief updating.
Word count: 9291.

*We would like to thank Honorata Mazepus, Monika Nalepa, Brendan Nyhan, and Dani Sandu as well as all participants of the 10th Swiss Summer School in Democracy Studies 2024, held by the University of Zurich and the University of Fribourg and participants of the BEELab meeting at Maastricht University. This research was approved by Maastricht University IRB. The experiment was pre-registered in the AsPredicted registry (#196961). Financial support from Maastricht University is gratefully acknowledged. Bronnikov: School of Business and Economics, Maastricht University; Political Science Department and Law School, University of Chicago; School of Science, Brīvā Universitāte; egor.bronnikov@maastrichtuniversity.nl. Tsakas: School of Business and Economics, Maastricht University; e.tsakas@maastrichtuniversity.nl. Vostroknutov: School of Business and Economics, Maastricht University; a.vostroknutov@maastrichtuniversity.nl. Thomsson: School of Business and Economics, Maastricht University; k.thomsson@maastrichtuniversity.nl.

1 Introduction

The United States is more politically divided than it has been in decades (e.g., Poole and Rosenthal, 1984, 1985; DeSilver, 2022). The gap between Democrats and Republicans extends beyond policy disagreements to deeper divides in values, trust, and even perceptions of reality (e.g., Sweetser, 2014; Balliet et al., 2018; Carlin and Love, 2018). Surveys show that members of each party increasingly view the other side with suspicion, with trust across party lines steadily declining (e.g., Jurkowitz et al., 2020; Schedler, 2023; Boxell et al., 2024). Differences in economic priorities, social policies, and cultural attitudes have widened, and partisan identity has become a significant marker of social belonging. Given this high level of polarization, one might expect many people prefer information sources that align with their views and are more skeptical of media associated with the opposing side. Shaped by both psychological and social factors, these tendencies are reinforced by (social) media that amplify content aligned with their audiences' ideological preferences.[1] At the same time, recent evidence suggests that the extent of informational segregation is more mixed than (stereo)typical "echo chamber" accounts would imply: most citizens' media diets remain relatively moderate, while strongly homogeneous information environments appear to characterize a smaller subset of users and may be shaped not only by outlet choice but also by the curation of the informational environment (e.g., Guess, 2021; Nyhan et al., 2023; Budak et al., 2024; Green et al., 2025).

While this phenomenon aligns with a broad literature on partisan-motivated reasoning, which suggests that individuals selectively process information, preferring politically congenial sources, that reinforce ideological predispositions (e.g., Lodge and Hamill, 1986; Schaffner and Streb, 2002; Gerber et al., 2010; Stroud, 2010; Gunther et al., 2012; Jerit and Barabas, 2012; Kahan, 2013; Bolsen et al., 2014; Peterson, 2017; Donovan et al., 2020; Guay and Johnston, 2022), recent work also cautions against treating all partisan information gaps as evidence of directional motivated reasoning alone. In some settings, citizens appear to differ less in their desire for accuracy than in the manner the information they happen to encounter appears to them (e.g., Druckman and McGrath, 2019; Peterson and Iyengar, 2021).

Existing research has primarily focused on factors such as source credibility, partisan alignment, and ideological congruence in shaping belief updating (e.g., Peterson and Iyengar, 2021; Skytte, 2025). Despite extensive theoretical and observational work on

[1]Psychologically, exposure to conflicting viewpoints can create cognitive dissonance, making people more inclined to seek out information that reinforces their existing beliefs rather than challenges them (e.g., Knobloch-Westerwick and Hastall, 2010; Stroud, 2010; Lane et al., 2023). Socially, political identity functions as a form of group affiliation, and consuming media from the out-group can signal a lack of loyalty, sometimes leading to social discomfort or even exclusion.

media polarization, direct experimental evidence that cleanly isolates causal source effects while holding informational content constant remains comparatively limited (e.g., Berinsky, 2017; Dafoe et al., 2018; Guess, 2021; Nyhan et al., 2023). Existing research shows that source familiarity, source hostility, and source credibility matter for how political information is selected and processed, but the evidence is rather conditional and context-dependent and in many cases cannot be fully explained by an in-vs.-out-group account (e.g., Peterson and Kagalwala, 2021; Peterson and Allamong, 2022; Carey et al., 2025).

At the same time, a key dimension of information processing—one that is central to information theory (e.g., Shannon, 1948; Cover and Thomas, 1999)—has remained relatively overlooked[2]: the role of *surprise* in (political) information. Regardless of its partisan source, the extent to which new information deviates from an individual's prior beliefs may significantly influence how—or whether—it is incorporated into their worldview. Using the term in an information-theoretic sense, in this paper, we refer to surprise as the degree to which the informational content of a signal is unexpected relative to an individual's prior beliefs. We measure this informational distance between prior and posterior beliefs using the log-likelihood ratio (LLR), defined as the change in log-odds between posterior and prior beliefs (e.g., Tversky and Kahneman, 1981; Bullock, 2009; Hill, 2017; Benjamin, 2019; Tappin et al., 2020).[3] While this definition is distinct from emotionally experienced surprise (as in, e.g., Bronnikov and Drouvelis, 2026), our measure captures outcome-based informational surprise relative to prior beliefs, rather than self-reported affective reactions that may reflect post-hoc rationalizations or experimental demand effects.

In this paper, we investigate how the extent of surprise in political information (relative to prior beliefs) shapes belief updating across different media sources. Specifically, we examine whether information that is more surprising relative to prior beliefs is more likely to induce updating across partisan and media contexts, including when it is attributed to ideologically incongruent sources. By exploring this underexamined mechanism, we aim to shed light on the potential for surprising political information to act as

[2]While explicit empirical applications of information-theoretic surprise in this context remain limited, closely related mechanisms are implicitly present in a large body of related work. Correction and prebunking experiments, for example, show that misinformation can sometimes be reduced, but effects are often modest, heterogeneous, and sensitive to source credibility and context (e.g., Berinsky, 2017, 2023; Carey et al., 2025; Compton et al., 2021; Roozenbeek and Van der Linden, 2024). Remarkably, the same psychological mechanism can also be exploited in more malign ways: in contemporary autocratic contexts, state-controlled media often stage talk shows or debates that simulate open discourse, but these are carefully engineered to deliberately misrepresent opposition views, making them seem illogical, extreme, or out of touch (e.g., Guriev and Treisman, 2022). At the same time, research in cognitive neuroscience demonstrates that surprise plays a foundational role in communication itself (e.g., Loewenstein, 2019; Buidze et al., 2024), underscoring the power of unexpected information.

[3]The measure has several advantages as well as an intuitive interpretation, on which we elaborate in Section 2.

a catalyst for belief updating, even in deeply polarized media environments.

We conducted two experimental studies (A and B) in which participants first reported their prior beliefs about Donald Trump's chances of winning the 2024 U.S. presidential election[4], were then exposed to a study-specific signal—a *pro-Trump* signal in Study A and an *anti-Trump* signal in Study B—with source attribution randomly assigned to either *The New York Times* or *Fox News*[5], and subsequently reported their posterior beliefs[6]. This design allows us to measure belief updating and assess whether patterns of updating are consistent with an information-theoretic role for surprise across partisan and media contexts. Since informational content is held constant within each study while only the source attribution is randomized, we can isolate the causal effect of source labeling in a politically realistic environment.

Our results indicate that belief updating is strongly anchored in individuals' prior beliefs, with comparatively limited evidence of direct media source effects. At the same time, the observed patterns of updating are consistent with an information-theoretic interpretation in which the informational content of a signal plays an important role in shaping belief updating. In particular, the effect of the signal's source appears to vary with the extent to which the signal deviates from individuals' prior beliefs. Contrary to explanations based solely on source in-group or out-group status or partisan alignment, belief updating was not consistently determined by whether a signal came from *The New York Times* or *Fox News*. Instead, stronger updates sometimes occurred when signals were attributed to counter-stereotypical sources.[7] In Study A, the largest belief update is observed among Republicans receiving a pro-Trump signal attributed to *The New York Times*. In Study B, the strongest updating occurs among Democrats receiving an anti-Trump signal attributed to *Fox News*. These patterns are consistent with the idea that counter-stereotypical signals may be perceived as more informative than signals coming from ideologically expected sources.

With regard to partisan identity, we find mixed evidence. Positive partisanship (strong attachment to one's party) is significantly associated with belief updating in Study A, but this relationship disappears in Study B. This suggests that belief updating reflects not only the informational distance between signals and prior beliefs, but also how individuals interpret information through the lens of partisan identity and affect. We also

[4]Both studies were conducted four days before the 2024 U.S. presidential election.

[5]We use news articles from the official websites of *The New York Times* and *Fox News* because this reflects how political information is often encountered in contemporary media environments—particularly through links shared on social media platforms. These links typically direct users to full articles hosted on the outlets' websites, making them a natural and ecologically valid stimulus format.

[6]Both prior and posterior beliefs were elicited using incentive-compatible scoring rules (see Section 3 for details).

[7]We use *counter-stereotypical* to refer to signals delivered by sources whose perceived ideological stance would normally imply the opposite message; such signals may therefore appear more surprising relative to prior expectations.

observe asymmetries in responsiveness to information. In particular, Republicans appear more resistant to anti-Trump signals than Democrats are to pro-Trump signals.

This paper contributes to several strands of literature. We add to the literature on the limitations of motivated (partisan) reasoning. While there is substantial evidence—both in formal theory (e.g., Little, 2025) and empirical research (e.g., Taber and Lodge, 2006; Nyhan and Reifler, 2010; Jerit and Barabas, 2012; Kahan, 2013; Flynn et al., 2017; Druckman and McGrath, 2019; Guay and Johnston, 2022; Little et al., 2022)—on the role and dynamics of motivated reasoning, there is also growing evidence that under certain conditions, individuals may incorporate information from ideologically incongruent sources (e.g., Melnikoff and Strohminger, 2024; Lois et al., 2025). Our results show that belief updating is not solely governed by partisan reasoning, but systematically varies with the relationship between prior beliefs and the informational content of the signal, even when information originates from ideologically incongruent sources.

We also contribute to the literature on the information landscape and (selective) exposure (e.g., Stroud, 2008; Iyengar and Hahn, 2009; Stroud, 2011; Arceneaux and Johnson, 2013; Messing and Westwood, 2014; Levendusky, 2013; Bail et al., 2018; Lazer et al., 2018; Guess et al., 2020; Broockman and Kalla, 2025). Traditional accounts of selective exposure emphasize the tendency of individuals to seek out information that aligns with their pre-existing beliefs and to avoid dissonant content (e.g., Stroud, 2010; Arceneaux and Johnson, 2013). At the same time, recent evidence suggests that politically homogeneous media environments are neither universal nor reducible to outlet choice alone: most users appear to consume relatively mixed media diets, while more polarized informational environments may emerge through curation, selective sharing, and selective engagement with specific stories (e.g., Guess, 2021; Green et al., 2025). Our findings suggest that the extent to which new information is incorporated depends on how its content relates to prior beliefs, helping to explain variation in responsiveness.

Finally, our study contributes to the literature on polarization (Poole and Rosenthal, 1984, 1985; Arceneaux and Johnson, 2013; Iyengar et al., 2012; Druckman et al., 2013; Iyengar and Westwood, 2015; McCarty et al., 2016; Druckman et al., 2021). While existing work has emphasized affective polarization and the increasing tendency of partisans to inhabit distinct informational and social worlds, our findings suggest that even in these fragmented environments, belief updating is possible when the element of surprise is present.[8] Patterns consistent with informational surprise appear in both partisan groups, suggesting that polarization does not eliminate the capacity for belief updating, but instead conditions when such updating occurs. Thus, rather than viewing polariza-

[8]This effect presumes exposure to such surprising content, which may be limited by selective avoidance. However, prior work has shown that incidental or unavoidable encounters with counter-attitudinal information do occur, particularly in social media or interpersonal settings (e.g., Garrett et al., 2013; Bakshy et al., 2015; Guess et al., 2020).

tion as an environment where information incorporation is excessively constrained, we highlight how surprising, expectation-violating content, particularly when embedded in ideologically incongruent sources—can loosen otherwise persistent patterns of partisan information processing.

The rest of the paper is structured as follows. In Section 2, we introduce a simple model to capture the extent and direction of belief updating using the Log-Likelihood Ratio and develop the hypotheses. Sections 3 and 4 outline the design of each study and present the results of studies A and B respectively. In Section 5, we discuss the results and conclude.

2 Theoretical Framework and Hypotheses

In this section, we introduce a theoretical framework and present hypotheses.

2.1 Belief updating

We start with the binary state space, where each state corresponds to a winner of the upcoming presidential election, i.e.,

$$\Omega = \{\text{Trump}, \text{Harris}\}\,.$$

A participant holds a subjective *prior belief* that assigns probability $\mathbb{P}(\text{Trump})$ to Trump winning, and probability $\mathbb{P}(\text{Harris}) = 1 - \mathbb{P}(\text{Trump})$ to Harris winning, both of which we elicit in the experiment (in an incentivized manner).

After elicitation of prior beliefs, participants receive a signal—an opinion of one of the two public figures. In study A, it is Nate Silver's signal suggesting that Trump will win, and in study B it is Allan Lichtman's signal suggesting that Trump will lose. In both studies, we vary the source of the signal—which will become treatments *within* studies—as quoted by either *Fox News* or *The New York Times*.

Mathematical information theory (e.g., Shannon, 1948; Cover and Thomas, 1999) formalizes this setting as participants receiving a signal S (a piece of relevant information shown to the participant). The two main characteristics of a signal, in our case, are the *opinion* of the expert (whether Trump will win [in Study A] or lose [in Study B]) and the *identity* of the source that reproduces this opinion (published either in *Fox News* or *The New York Times*).

While S can in principle take four potential values, we only vary the source within each study (see Figure 1). Put differently, our experimental design does not constitute a full cross-design experiment, as the content and wording of the signals differ across studies. This limitation reflects a deliberate trade-off aimed at preserving ecological validity, since real-world expert forecasts typically convey a single directional prediction rather than experimentally balanced signals.

	NY Times	Fox News
Trump wins (Study A)	$S_{T,H}$	$S_{T,T}$
Trump loses (Study B)	$S_{H,H}$	$S_{H,T}$

Figure 1: Summary of Signals

Note: This paper does not provide a full cross-design experiment; comparisons of participants' reactions to signals are made within each study.

Upon receiving one of the four possible signals, the participant updates to a subjective posterior belief that assigns an updated probability to

$$\mathbb{P}(\text{Trump}|S) = \frac{\mathbb{P}(S|\text{Trump}) \cdot \mathbb{P}(\text{Trump})}{\mathbb{P}(S|\text{Trump}) \cdot \mathbb{P}(\text{Trump}) + \mathbb{P}(S|\text{Harris}) \cdot \mathbb{P}(\text{Harris})} \tag{1}$$

to Trump winning. The conditional probabilities $\mathbb{P}(S|\text{Trump})$ and $\mathbb{P}(S|\text{Harris})$ are the *likelihoods* that the participant subjectively assigns to receiving signal S assuming that Trump wins and respectively assuming that Harris wins.

It is important to stress that, by using the Bayes formula in eq. (1), we are not suggesting that participants are Bayesian agents who update their beliefs rationally. This is because the likelihoods, $\mathbb{P}(S|\text{Trump})$ and $\mathbb{P}(S|\text{Harris})$, are not objectively given, as they typically are in the experimental literature on belief updating biases (see Benjamin, 2019, and references therein). Instead, in our paper, the Bayes formula is used purely as a structural model that enables us to quantify the extent to which participants incorporate information into their beliefs, as further discussed later in this section.[9]

As it is commonly done (e.g., Grether, 1980; Tversky and Kahneman, 1981; Rabin, 1998; Bullock, 2009; Hill, 2017; Benjamin, 2019; Tappin et al., 2020), we measure the extent to which participants update their beliefs in response to a signal, using the log-likelihood ratio (LLR):

$$\underbrace{\log\left[\frac{\mathbb{P}(S|\text{Trump})}{\mathbb{P}(S|\text{Harris})}\right]}_{\text{LLR}} = \underbrace{\log\left[\frac{\mathbb{P}(\text{Trump}|S)}{\mathbb{P}(\text{Harris}|S)}\right]}_{\text{Log-posterior odds}} - \underbrace{\log\left[\frac{\mathbb{P}(\text{Trump})}{\mathbb{P}(\text{Harris})}\right]}_{\text{Log-prior odds}}. \tag{2}$$

This eq. (2) follows directly from dividing eq. (1) with the corresponding posterior belief

[9]For a more elaborate discussion of this interpretation of the Bayes formula, we refer to Rabin (2013) and Lois et al. (2023).

for Harris, and subsequently taking logarithms. That is, the LLR is inferred from the difference between the log-posterior odds and the log-prior odds (see Appendix A), which are directly elicited, by asking people to report the probability they attach to Trump winning before and after they receive signal S.

Conceptually, the LLR quantifies the amount of information that the participant incorporates into their beliefs upon receiving the signal. In Study A, where the signal suggests that Trump will win, a positive LLR implies that the participant interprets the received signal S as evidence in favor of Trump winning, whereas a negative LLR implies that the participant interprets S as evidence in favor of Harris winning. In Study B, where the signal suggests that Trump will lose, the interpretation is reversed: a negative LLR implies that the participant interprets the signal as evidence in favor of Harris winning, whereas a positive LLR implies that it is interpreted as evidence in favor of Trump winning. Naturally, an LLR equal to 0 implies that the participant has not taken the signal into account and has therefore not updated their prior belief (i.e., posterior belief equals prior belief). Furthermore, the absolute value of the LLR reflects how strongly the evidence is perceived by the participant: the further the LLR is from zero (in either direction), the stronger the perceived informational content of the signal and the greater the degree of belief updating.

Note that using the LLR, rather than the absolute difference between posterior and prior beliefs, allows us to control for the role of prior beliefs. For example, the LLR obtained when beliefs are updated from 80% to 90% is much larger than the LLR obtained when beliefs are updated from 50% to 60%. Although the absolute change in beliefs is the same in both cases (10 percentage points), the informational content of the update differs: in the former case the participant incorporates considerably more information (LLR = 0.811) than in the latter case (LLR = 0.405). That is, the signal is interpreted as stronger evidence in the former case than in the latter (see Appendix A for more details). This interpretation is consistent with information theory, in the sense that the LLR reflects the extent to which the evidence is surprising relative to the participant's prior beliefs.

Finally, it is worth noticing that we do not make any exogenous assumption on how participants interpret and subsequently incorporate the received signal S into their beliefs. In particular, we do not postulate any specific form of updating, nor are we comparing their observed updating with a Bayesian benchmark. In fact, given that LLR is not exogenously given, but rather inferred from the prior and posterior odds, it is not even possible to define what the objective Bayesian benchmark is. This flexibility is particularly advantageous in political contexts where it is unclear how each participant interprets the likelihood of a signal, which contributes to the ecological validity of the study.

Overall, we measure belief updating within the standard framework of decision theory and information theory that has been widely used in both economics (e.g., Grether, 1980; Tversky and Kahneman, 1981; Rabin, 1998; Benjamin, 2019; Ortoleva, 2022) and political science (e.g., Bullock, 2009; Hill, 2017; Tappin et al., 2020). For the rest of the paper, we will use LLR as the dependent variable, and we will refer to it as our measure of belief updating.

2.2 Hypotheses

Now, we turn to hypotheses. For the purpose of this study, we develop five hypotheses—three associative and two causal—all of which were pre-registered before the experiment.[10]

Since there is substantial evidence that individuals' political identities strongly influence how they interpret and incorporate new information (e.g., Lodge and Hamill, 1986; Dalton et al., 1998; Schaffner and Streb, 2002; Gerber et al., 2010; Stroud, 2010; Gunther et al., 2012; Jerit and Barabas, 2012; Peterson, 2017), it is natural to expect belief updating to be correlated with party affiliation.

Hypothesis 1 (Association with party affiliation)**.** *Belief updating is correlated with party affiliation.*

For instance, Republicans and Democrats often interpret the same piece of information differently based on their pre-existing ideological positions in general and party affiliation in particular (e.g., Miller et al., 2016; Garrett and Bond, 2021; Prike et al., 2023) Partisan motivated reasoning may lead individuals to update their beliefs selectively, aligning with party agenda rather than objective truth (e.g., Kahan, 2013; Petersen et al., 2013; Bolsen et al., 2014; Donovan et al., 2020; Guay and Johnston, 2022).

Second, belief updating being correlated with the source of the political signal is also particularly relevant in a polarized U.S. media landscape (e.g., Levendusky and Malhotra, 2016; Druckman et al., 2019; Peterson and Iyengar, 2021).

Hypothesis 2 (Association with the source)**.** *Belief updating is correlated with the source of the political signal.*

For example, previous studies showed that Republicans are more likely to trust conservative sources (e.g.,*Fox News*), while Democrats are more likely to trust liberal sources (e.g.,*The New York Times*) (e.g., Taber and Lodge, 2006; DellaVigna and Kaplan, 2007; Nyhan and Reifler, 2010; Nyhan et al., 2013; Schroeder and Stone, 2015; Jurkowitz et

[10]The experiment was pre-registered in the AsPredicted registry (#196961).

al., 2020; Ash et al., 2024). Mistrust of opposing-party-affiliated sources is documented to lead to (higher) skepticism (e.g., Goldberg et al., 2021; Merkley and Stecula, 2021).

Third, the level of partisanship, both positive and negative, significantly influences belief updating (e.g., Van Bavel and Pereira, 2018; Li and Wagner, 2020; Bankert, 2021; Lee et al., 2022).

Hypothesis 3 (Association with partisanship)**.** *Belief updating is correlated with the level of (positive and/or negative) partisanship.*

Positive partisanship (i.e., strong identification with one's own party) can lead to more biased assimilation of information that aligns with one's views, disregarding conflicting evidence (e.g., Taber and Lodge, 2006; Lodge and Taber, 2013; Iyengar et al., 2019). Negative partisanship (i.e., strong animosity toward the opposing party) may result in outright rejection of information perceived as coming from the opposing side, even when the source is credible (e.g., Taber and Lodge, 2006; Abramowitz and Webster, 2018; Iyengar et al., 2019).

Now we turn to the causal hypotheses. We begin by examining the effect of the source on belief updating, which is critical for understanding the role of the media sources through which the signal is conveyed. To process complex political information, individuals often rely on heuristic cues such as source identity (e.g., Cohen, 2003; Taber and Lodge, 2006). In polarized contexts, partisan identity strongly conditions both the exposure to and acceptance of political information, amplifying source effects (e.g., Levendusky, 2013; Guess et al., 2020). However, a competing view, grounded in information theory, suggests that belief updating depends less on ideological congruence and more on the extent to which new information—or its source—violates prior expectations. In this framework, surprise plays a central role in belief updating. Recent work in behavioral economics supports this perspective, showing that unexpected information can drive belief updating even when it originates from identity incongruent sources (e.g., Bronnikov and Drouvelis, 2026). Moreover, research in cognitive neuroscience demonstrates that surprise plays a foundational role in communication itself (e.g., Loewenstein, 2019; Buidze et al., 2024), underscoring the universal and domain-general power of surprise. These contrasting views motivate two competing hypotheses.

Hypothesis 4a (Motivated reasoning perspective)**.** *Individuals will update their beliefs* more (less) *when the signal comes from an ideologically* congruent (incongruent) *source.*

Hypothesis 4b (Surprise-based perspective)**.** *Individuals will update their beliefs more when the signal or its source is* surprising *relative to their prior beliefs, regardless of ideological congruence.*

On the one hand, consistent with H(4a), prior experimental research has shown that varying the source of identical information can significantly influence how individuals update their beliefs, particularly when the source aligns or conflicts with partisan affiliation (e.g., Thaler, 2021, 2024). This supports the idea that people tend to incorporate information from ideologically congruent sources while discounting incongruent ones. On the other hand, in line with H(4b), classical information theory (e.g., Shannon, 1948; Cover and Thomas, 1999) and recent epistemic models of belief updating (e.g., Benjamin, 2019; Bronnikov and Drouvelis, 2026) suggest that belief updating is shaped by the degree to which new information deviates from prior expectations. From this perspective, even signals from out-group sources can prompt substantial updating when they violate expectations in a salient way.

Finally, the interaction effects of party affiliation, source, and partisanship are likely to amplify belief polarization (e.g., Druckman, 2001; Taber and Lodge, 2006; Bullock, 2011).

Hypothesis 5 (Interaction effects)**.** *The interaction effects of the above-mentioned factors (namely party affiliation, the source, positive and negative partisanship) will have a significant effect on belief updating.*

For instance, a Democrat receiving information from a Republican-affiliated source (e.g.,Fox News) may exhibit even stronger resistance to updating due to a combination of party affiliation, mistrust of the source, and negative partisanship. Conversely, a Republican receiving information aligned with their party's stance but from an unfamiliar or neutral source may update their beliefs more cautiously. These interactions reflect the complexity of belief formation, which cannot be fully explained by one variable in isolation.

3 Study A: "Trump Wins"

In Study A, we investigate how participants update their beliefs about Trump's chances of winning in response to a signal favorable to the candidate (i.e., an expert opinion that Trump will win). Participants were randomly assigned to a treatment in which favorable information about Trump was delivered either by *The New York Times* or *Fox News*. The between-subjects experiment was carried out on the Prolific platform on November 1, 2024, with a sample of $N_A = 277$, limited to US nationals and excluding participants from Study B (see details in Appendix B).

3.1 Experimental Design

Belief updating: priors, signals, and posteriors. First, participants were asked to state their initial (prior) beliefs about Trump's chances of winning the upcoming 2024 Presidential election. Using a scale from 0 to 100, they responded to the question: *What do you think the probability (in %) is that Donald Trump will win the next Presidential election?* (see Figure 3 in Appendix C). The distribution of prior and posterior beliefs are presented in Appendix D.

Following this, participants were shown a signal that presented the opinion of Nate Silver, who predicted that Donald Trump would win the election (upcoming at the time of the experiment). Each participant in Study A was randomly assigned to a treatment which provided the signal either from *The New York Times* or from *Fox News*. Table 1 shows the precise wording (see Figures 4 and 5 in Appendix C for the operationalization in the experiment). The content of the information provided was identical except for the source. Participants clearly distinguished between the ideological orientation of the two outlets (see Appendix G).

After receiving the signal, participants provided updated (posterior) beliefs about Trump's chances of winning (using the same 0 to 100 scale). The format of this elicitation mirrored that of the prior beliefs stage. To incentivize accurate responses, both prior and posterior beliefs elicitation were monetarily incentivized using a quadratic scoring rule. This method ensures incentive compatibility by encouraging participants to truthfully report their beliefs, as their earnings increased the closer their stated probabilities aligned with the actual outcome of the election.[11]

Additional measures and controls. We also collected additional variables to gain a better picture of the participants' characteristics and how they might affect belief updating. These additional variables include data that was used to construct indices for positive and negative partisanship, measures of cognitive reflection ability, and sociodemographic details (a full description of each is provided in Appendix H).

[11]Following a study by Danz et al. (2022), which addresses the most optimal way to communicate how the quadratic scoring rule works, as well as a recent detailed review by Haaland et al. (2023), we provided participants with a one-line explanation: *It is optimal for you to report your estimate as precisely as possible.* Since the realization of the payments required the truthful state of the world to be known, the payments were made two weeks after the Presidential Election. This timing aligns with the policies of Prolific, where the experiment was conducted.

Table 1: Formulations of signals used in both treatments in Study A.

Signal	Treatment (Source)	Particular Formulation
Trump wins	*Fox News*	*According to Fox News, Nate Silver stated that his gut tells him that Donald Trump will win.*
Trump wins	*The New York Times*	*According to The New York Times, Nate Silver stated that his gut tells him that Donald Trump will win.*

3.2 Main Results

We begin by examining the first three correlational hypotheses. We find only one statistically significant correlation with the LLR.

Result A1. *We find no significant correlation between LLR and party affiliation; LLR and the source; LLR and negative partisanship. The correlation between LLR and positive partisanship is* 0.126 *and significant* ($p = 0.040$)*.*

Thus, we do not find evidence to support hypotheses H(1) and H(2), though we do find some weak partial support for H(3). That is, there is (some) evidence to support the idea that belief updating is positively correlated with positive partisanship. Although the correlation is weak, it is statistically significant, suggesting that individuals with higher positive partisanship are slightly more likely to update their beliefs.

Next, we proceed to our main hypotheses H(4a) and H(4b). We summarize our findings as a result.

Result A2. *On average, the source of the signal did not have a direct effect on belief updating after accounting for other factors such as party affiliation and prior beliefs. Under signals predicting Trump's victory, Democrats exhibit no significant belief updating regardless of the source. This limited response suggests that, although the signal content runs against their partisan priors, it is not incorporated as sufficiently credible or informative to induce meaningful updating. For Republicans, belief updating is stronger, particularly when the signal is attributed to* The New York Times *rather than* Fox News*. This pattern is consistent with the idea that a pro-Trump signal coming from a counter-stereotypical source is perceived as more informative and therefore generates stronger updating.*

The evidence for these results is summarized in Table 2 that explores average belief-updating behavior (average LLR) among Democrats and Republicans in response to a signal from Nate Silver predicting Trump's victory. Positive LLR values indicate increased belief in Trump's victory after receiving the signal.

Table 2: Average LLRs in Study A (signal on Trump's High Chances of Winning).

Study A	Democrats	Republicans
The New York Times treatment	0.096 (0.077)	**0.177** (0.052)
Fox News treatment	0.018 (0.031)	0.109 (0.055)

Note: The table presents the results for Study A, showing the average Log-Likelihood Ratios and standard errors in parenthesis. The rows indicate the sources of the signal (*The New York Times* vs. *Fox News*), i.e., the treatment; the columns represent the receivers of the signal (Democrats vs. Republicans). The LLR values that are significantly different from zero (at $\alpha = 0.05$) are highlighted in bold.

When the signal was attributed to *The New York Times*, Democrats showed a modest positive belief update with an LLR of 0.096 (SE=0.077, 95% CI: [–0.058, 0.249]). In terms of information theory, the modest positive belief update is consistent with the signal containing limited informational surprise for Democrats. The source, *The New York Times*, aligns with their general expectations about the source but conflicts with their partisan priors that Trump is unlikely to win. The positive LLR reflects that Democrats showed only limited belief updating, suggesting that the signal contained little informational content relative to their prior beliefs, as it deviated from their expectations about Trump's chances while coming from a source they do not outright discount. However, the relatively small magnitude of the LLR and the lack of statistical significance indicate that the signal's information content—or surprise value—was limited. That is, the signal appears to have been only modestly unexpected relative to their prior beliefs, and not sufficiently so to substantially shift them.

When the signal was attributed to *Fox News*, Democrats exhibited an even smaller extent of belief updating, with an LLR of 0.018 (SE=0.031, 95% CI: [–0.043, 0.079]). Here, the negligible belief update is consistent with the signal from *Fox News* carrying minimal informational surprise. The alignment of the signal with Republican-leaning expectations, combined with Democrats' likely skepticism of *Fox News* as a credible source, resulted in very low information content. From an information theory perspective, the signal did not deviate substantially from Democrats' prior beliefs about the election outcome. As a result, the LLR indicates that the signal had almost no effect on their posterior beliefs, as it provided little to no surprising information.

When the signal was attributed to *The New York Times*, Republicans showed a statistically significant belief update, with an LLR of 0.177 (SE=0.052, 95% CI: [0.073, 0.281]). This positive update is consistent with high (informational) surprise in information-theoretic terms. For Republicans, the signal predicting Trump's victory coming from *The New York Times*—a source typically perceived as ideologically opposed to Trump—represents a counter-stereotypical signal. The counter-stereotypical nature of the signal is consistent with higher perceived informational content. As a result, Republicans may have interpreted the signal as more informative and incorporated it more strongly into their posterior beliefs. The higher LLR suggests that participants treated the signal as a surprisingly aligned piece of evidence and updated more strongly in response.

When the signal was attributed to *Fox News*, Republicans exhibited a smaller belief update, with an LLR of 0.109 (SE=0.055, 95% CI: [–0.001, 0.220]). The smaller belief update here suggests that the signal from *Fox News* is consistent with lower informational surprise for Republicans. Since *Fox News* is a source commonly perceived as aligning with their priors (that Trump has a high chance of winning), the signal likely deviated less from what they anticipated. Consequently, the information content of the signal was limited, as it served more as a reinforcement of existing beliefs rather than a surprising piece of evidence. The moderate LLR indicates that while the signal was somewhat informative, it did not provide much novelty and thus had a weaker impact on posterior beliefs compared to the signal from *The New York Times*.

Now we turn to two types of comparisons of the results presented in Table 2, namely comparison (i) over parties, and (ii) over sources. When the source is *The New York Times*, there is a statistically significant difference in how Democrats and Republicans update their beliefs under the pro-Trump signal, with Republicans exhibiting a stronger positive response than Democrats. When the source is *Fox News*, the result is not statistically significant (see full details in Appendix E). The comparison over sources does not show any significant results (see Appendix F for details).

Finally, we report the results for hypothesis H(5).

Result A3. *On average, the interaction between the signal source and party affiliation does not significantly influence belief updating. Similarly, the interactions between party affiliation and partisanship measures (both positive and negative) are not statistically significant.*

To examine the direct and interaction effects of partisan identity and partisanship on belief updating, we estimate three OLS regression models reported in Table 3. [12] All specifications include demographic controls (age, education, employment status, gender,

[12]In the baseline analyses, observations with prior or posterior beliefs equal to 0 or 100 are omitted because the LLR is undefined at those values; in robustness checks, these responses are retained by recoding 0 and 100 to 0.1 and 99.9, respectively.

Table 3: Direct and Interaction Effects on the LLR in Study A

	Dependent variable:		
	LLR		
	(1)	(2)	(3)
Prior		-0.005^{*}	-0.006^{**}
		(0.003)	(0.003)
Source = Fox News	-0.057	-0.043	-0.028
	(0.053)	(0.049)	(0.065)
Party = Republicans	0.075	0.182^{***}	-0.141
	(0.056)	(0.069)	(0.231)
Positive Partisanship	0.060^{*}	0.062^{**}	0.056
	(0.031)	(0.031)	(0.040)
Negative Partisanship	-0.023	-0.025	-0.057^{*}
	(0.025)	(0.025)	(0.030)
Fox News × Republicans			-0.027
			(0.105)
Republicans × Positive Partisanship			0.007
			(0.057)
Republicans × Negative Partisanship			0.075
			(0.049)
Controls	✓	✓	✓
Observations	268	268	268
R^2	0.067	0.097	0.106

Note: This table presents regression results examining direct and interaction effects on the LLR. Standard errors are shown in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. See the description of the variables in Appendix I.

ethnicity, place of birth, and student status), though these coefficients are suppressed in the table for brevity. The Appendix reports robustness checks using recoded extreme prior and posterior beliefs (see Appendix J) as well as an alternative model specification (see Appendix K).

In the baseline specification (Column 1), the model includes only direct effects and excludes both prior beliefs and interaction terms. In this model, positive partisanship is positively associated with belief updating ($\beta = 0.060$, $p < 0.05$), indicating that individuals with stronger affective attachment to their party exhibit larger belief updates in response to the signal. The coefficient on Republican identification is positive but not statistically significant ($\beta = 0.075$). Negative partisanship and the *Fox News* source indicator are also statistically insignificant.

Column 2 introduces prior beliefs while maintaining a purely additive structure. The coefficient on Prior is negative and statistically significant ($\beta = -0.005$, $p < 0.05$), consistent with a surprise interpretation: individuals whose initial beliefs are already aligned with the signal update less, reflecting lower informational surprise. Controlling for priors substantially strengthens the effect of Republican identification, which becomes positive and highly significant ($\beta = 0.182$, $p < 0.001$); positive partisanship also remains statistically significant ($\beta = 0.062$, $p < 0.01$).

The full specification in Column 3 incorporates interaction terms between party affiliation and both the signal source and partisanship measures. The coefficient on Prior remains negative and statistically significant ($\beta = -0.006$, $p < 0.01$), reinforcing the conclusion that higher prior alignment attenuates belief updating. Once interactions are introduced, the main effect of Republican identification becomes negative ($\beta = -0.141$) and statistically insignificant, indicating that partisan differences are now primarily captured through interaction channels. None of the interaction terms—including *Fox News* × Republicans and interactions between Republican identity and positive or negative partisanship—reach conventional significance levels. Negative partisanship becomes marginally significant in this specification ($\beta = -0.057$, $p < 0.10$), suggesting that stronger out-party animus may slightly dampen belief updating once partisan heterogeneity is fully modeled.

Taken together, these results indicate that prior beliefs and positive partisanship are the most robust factors of belief updating in Study A. While partisan identity matters once priors are accounted for, there is limited evidence that the effect of the signal source or partisan affect varies systematically across party lines. Overall, belief updating is primarily anchored in prior beliefs, while the magnitude of updating is consistent with an information-theoretic interpretation in which signals that deviate more strongly from prior beliefs generate larger updates.

4 Study B: "Trump Loses"

In Study B, we examine how individuals update their beliefs about Trump's chances of winning in response to a signal unfavorable to him. As in Study A, participants were randomly assigned to receive this signal—indicating low chances of his victory—from either *The New York Times* or *Fox News*. Study B was conducted on the Prolific platform on the same day as Study A, November 1, 2024. The sample of $N_B = 275$ was restricted to individuals with US nationality and those who had not participated in Study A (see details in Appendix B).

4.1 Experimental Design

The experimental design in Study B is identical to that in Study A, except for the signal received by participants. In Study B, participants received a signal conveying Allan Lichtman's prediction that Trump would lose the election. This signal was presented as reported by either *The New York Times* or *Fox News* (see Table 4). The distributions of prior and posterior beliefs are presented in Appendix D.

Table 4: Formulations of signals used in both treatments in Study B.

Signal	Treatment (Source)	Particular Formulation
Trump loses	*Fox News*	*According to Fox News, Allan Lichtman predicted that Kamala Harris will be the next president of the United States.*
Trump loses	*The New York Times*	*According to the New York Times, Allan Lichtman predicted that Kamala Harris will be the next president of the United States.*

4.2 Main Results

As in Study A, we first proceed with correlational hypotheses.

Result B1. *We find no significant correlation between LLR and party affiliation; LLR and the source; LLR and negative or positive partisanship.*

Although the simple correlations are not statistically significant, the regression analyses below reveal positive conditional associations between negative partisanship and belief updating. Accordingly, Study B provides no support for H(1) and only conditional evidence for H(3).

Next, we turn to hypotheses H(4a) and H(4b), which concern the primary result of Study B.

Result B2. *On average, the source of the signal does not exhibit a statistically significant direct effect on belief updating once other factors, including party affiliation and prior beliefs, are taken into account. At the descriptive level, however, Democrats exhibit stronger belief updating when the anti-Trump signal is attributed to* Fox News *than to* The New York Times*, consistent with the idea that a counter-stereotypical source increases the perceived* *informational content of the signal. Republicans significantly update their beliefs when the same signal is attributed to* The New York Times*, but show no significant updating when it is attributed to* Fox News*. This pattern suggests that the informational value of the signal depends on the interaction between source identity and partisan expectations.*

When the source is *Fox News*, Democrats update significantly more than Republicans, suggesting that an anti-Trump signal from a counter-stereotypical source carries greater informational content for Democrats than for Republicans.

Table 5 shows average LLR among Democrats and Republicans. Negative LLR values signify that participants changed their beliefs in the direction of the signal (Trump loses).

Table 5: Average LLRs in Study B (signal on Trump's Low Chances of Winning).

Study B	Democrats	Republicans
The New York Times treatment	**–0.218**	**–0.259**
	(0.065)	(0.085)
Fox News treatment	**–0.423**	–0.093
	(0.096)	(0.068)

Note: The table presents the results for Study B, showing the average Log-Likelihood Ratios and standard errors in parenthesis. The rows indicate the sources of the signal (*The New York Times* vs. *Fox News*), i.e., the treatment; the columns represent the receivers of the signal (Democrats vs. Republicans). The LLR values that are significantly different from zero (at $\alpha = 0.05$) are highlighted in bold.

When the signal was attributed to *The New York Times*, Democrats showed a significant decrease in their belief in Trump's chances of winning, with an LLR of –0.218 (SE=0.065, 95% CI: [–0.348, –0.087]). From an information theory perspective, the significant negative LLR indicates that Democrats found the signal about Trump's low chances of winning from *The New York Times* to be informative. This signal aligns with their prior beliefs (that Trump is unlikely to win), and its delivery by *The New York Times* may have made it easier to incorporate. The moderate extent of surprise stems from the fact that while the signal is consistent with their expectations about Trump's chances, it still provides new information. The decrease in belief (negative LLR) reflects that the signal confirmed and reinforced Democrats' priors while offering enough informational content to drive a belief update.

When the signal was attributed to *Fox News*, Democrats exhibited an even stronger belief update, with an LLR of –0.423 (SE=0.096, 95% CI: [–0.615, –0.231]). Here, the stronger negative LLR is consistent with the signal from *Fox News* carrying high informational surprise for Democrats. This result aligns with information theory principles:

the signal deviates from Democrats' expectations about the source (*Fox News* is typically associated with pro-Republican bias). The counter-stereotypical nature of the signal (a Republican-aligned source predicting Trump's low chances) is consistent with the signal being perceived as more informative, leading to a significant belief update. The greater magnitude of the LLR compared to *The New York Times* suggests that this pattern is consistent with higher informational content of the signal, amplifying belief updating.

When the signal came from *The New York Times*, Republicans showed a statistically significant belief update, with an LLR of –0.259 (SE=0.085, 95% CI: [–0.428, –0.089]). For Republicans, the significant negative LLR reflects a substantial belief update in response to the signal about Trump's low chances of winning from *The New York Times*. In terms of information theory, the signal is consistent with moderate informational surprise: the content conflicted with their partisan priors (that Trump has a high chance of winning), but the source is expected to present anti-Trump evidence. This combination of prior beliefs and source attribution is consistent with the signal carrying sufficient informational content to induce belief updating. While the signal was not entirely unexpected, it still represented enough of a deviation from Republicans' priors to result in a meaningful shift in posterior beliefs.

In contrast, when the signal came from *Fox News*, Republicans exhibited a smaller and statistically insignificant belief update, with an LLR of –0.093 (SE=0.068, 95% CI: [–0.229, 0.043]). The small and statistically insignificant negative LLR suggests that Republicans' response to the signal from *Fox News* is consistent with the signal carrying low informational content relative to their prior beliefs. While the signal contradicted Republicans' prior beliefs about Trump's chances, it came from *Fox News* (a source they generally perceive as aligned with their worldview). Because of this alignment, the signal may have been interpreted as less surprising, reducing its perceived informativeness and resulting in minimal belief updating. From an information-theoretic perspective, this pattern is consistent with the signal carrying relatively low informational content.

Now we turn to the comparisons of the results presented in Table 5 over parties and sources. When the source is *Fox News*, there is a statistically significant difference in how Democrats and Republicans update their beliefs under the anti-Trump signal, with Democrats exhibiting a substantially stronger update than Republicans. This pattern is consistent with the idea that an anti-Trump signal from a counter-stereotypical source carries greater perceived informational content for Democrats (see Appendix E). When the source is *The New York Times*, the result is not statistically significant. The comparison over sources did not bring any significant results (see Appendix F for details).

Finally, we evaluate hypothesis H(5) using evidence from Study B.

Result B3. *Belief updating is partially shaped by the interaction between signal source and* *party affiliation. In particular, the effect of a* Fox News *source differs between Democrats and Republicans. By contrast, the interactions between party affiliation and positive or negative partisanship are not statistically significant.*

The positive and significant *Fox News* × Republicans coefficient indicates that, relative to Democrats, Republicans react less negatively to the anti-Trump signal when it is attributed to *Fox News*. In other words, the impact of the *Fox News* source is conditional on party affiliation.

To investigate belief updating under an unfavorable signal, we estimate OLS regression models analogous to those in Study A. The results are reported in Table 6.[13] All specifications include demographic controls (age, education, employment status, gender, ethnicity, place of birth, and student status), though these coefficients are suppressed in the table for brevity. The Appendix presents robustness analyses that incorporate recoded extreme prior and posterior beliefs (see Appendix J) and an alternative model specification (see Appendix K).

In the baseline specification (Column 1), we exclude prior beliefs and all interaction terms. The model therefore captures only direct associations between the treatment source, partisan identity, partisanship measures, and the controls. In this specification, Republican identification is positive and statistically significant ($\beta = 0.172$, $p < 0.05$), while the *Fox News* indicator is small and statistically indistinguishable from zero ($\beta = -0.006$). Negative partisanship is positively associated with belief updating ($\beta = 0.084$, $p < 0.01$), whereas positive partisanship is not statistically significant ($\beta = -0.052$).

Column 2 introduces the prior belief measure while maintaining an additive structure. The coefficient on Prior is negative and highly significant ($\beta = -0.009$, $p < 0.001$), indicating that higher prior beliefs (i.e., beliefs closer to the signal and therefore less surprising) are associated with smaller belief updates. Controlling for priors strengthens the estimated role of partisan identity: the Republican coefficient increases to 0.321 and is highly significant ($p < 0.001$). Negative partisanship remains positive and statistically significant ($\beta = 0.068$, $p < 0.05$), while the *Fox News* coefficient remains statistically insignificant ($\beta = -0.019$).

The full specification (Column 3) adds interaction terms that allow the effect of the signal source and partisanship to vary by partisan identity. Prior beliefs remain negative and highly significant ($\beta = -0.010$, $p < 0.001$), again consistent with an information-theoretic interpretation. Although the direct *Fox News* coefficient becomes more negative ($\beta = -0.165$) and the main effect of Republican identification becomes statistically insignificant ($\beta = 0.159$), the interaction between *Fox News* and Republican identification is positive and statistically significant ($\beta = 0.356$, $p < 0.01$). This indicates that the

[13]The baseline LLR analyses exclude 0/100 prior or posterior responses because log-odds are undefined at those values; robustness checks retain them by recoding 0 and 100 to 0.1 and 99.9.

Table 6: Direct and Interaction Effects on the LLR in Study B

	Dependent variable:		
	(1)	(2)	(3)
Prior		−0.009***	−0.010***
		(0.003)	(0.003)
Source = Fox News	−0.006	−0.019	−0.165
	(0.088)	(0.085)	(0.113)
Party = Republicans	0.172*	0.321***	0.159
	(0.093)	(0.099)	(0.361)
Positive Partisanship	−0.052	−0.030	−0.067
	(0.033)	(0.031)	(0.042)
Negative Partisanship	0.084**	0.068*	0.108**
	(0.039)	(0.036)	(0.051)
Fox News × Republicans			0.356**
			(0.155)
Republicans × Positive Partisanship			0.099
			(0.060)
Republicans × Negative Partisanship			−0.118
			(0.075)
Controls	✓	✓	✓
Observations	260	260	260
R^2	0.076	0.115	0.143

Note: This table presents regression results examining direct and interaction effects on the LLR. Standard errors are shown in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. See the description of the variables in Appendix I

effect of a *Fox News* source differs significantly between Republicans and Democrats. In particular, relative to Democrats, Republicans react less negatively to the anti-Trump signal when it is attributed to *Fox News*. The remaining interaction terms with positive and negative partisanship are not statistically significant.

Taken together, the results in Study B highlight the robust role of prior beliefs and indicate that negative partisanship is positively associated with belief updating. Moreover, the significant *Fox News* × Republican interaction provides evidence that source effects differ by partisan identity in this setting: the impact of a *Fox News* source on belief updating is conditional on whether the receiver is a Democrat or a Republican.

5 Discussion

We now turn to a discussion of the study's limitations, the robustness of the findings, and their implications for future research and policy.

Strengths and limitations. Several limitations warrant discussion. First, throughout the paper, we use the term *surprise* to refer exclusively to outcome-based belief updating relative to prior beliefs, operationalized as changes in log-odds between posterior and prior beliefs. We do not directly elicit participants' expectations about source behavior or their subjective sense of unexpectedness upon receiving the signal. Accordingly, our findings speak to how beliefs adjust conditional on baseline outcome expectations, rather than to psychologically experienced surprise or subjective expectancy violations per se. Second, Studies A and B do not constitute a fully crossed factorial design, as signal valence is bundled with messenger identity and wording. Importantly, however, our causal estimates rely exclusively on within-study comparisons, where informational content is held constant and only source attribution is randomized. We therefore do not interpret differences across Studies A and B. Rather, each study provides an internally valid test of source attribution within a distinct real-world informational context. The absence of a fully crossed design thus reflects a deliberate tradeoff between strict experimental symmetry and ecological realism. Third, while our sample size provides adequate power for detecting main effects, interaction estimates are necessarily less precise, and statistically insignificant coefficients should not be interpreted as evidence of zero effects.

At the same time, the study design offers an important strength: it embeds randomized information exposure within a real electoral context at a moment of high political salience. By randomizing source attribution while holding informational content constant within each study, we isolate causal effects of source labeling under conditions that approximate naturally occurring political communication. This design therefore provides rare experimental leverage on belief updating in an ecologically valid, high-stakes setting.

Robustness. Our conclusions are supported by two complementary robustness checks. First, recoding extreme prior and posterior beliefs to retain respondents with absolute expectations yields substantively similar patterns: partisan identity and affective polarization remain relevant predictors of belief updating, while direct source effects remain limited (see Appendix J). Although the negative association between prior beliefs and updating attenuates when extreme responses are included—particularly in Study A—this shift suggests that individuals holding absolute convictions may respond to political information differently, with updating less tightly related to informational distance from prior beliefs. Second, ANCOVA specifications conditioning directly on baseline log-odds beliefs replicate the main qualitative findings across both studies, confirming that posterior beliefs are strongly anchored in prior beliefs, with more modest contributions from partisanship and limited evidence of systematic source-based persuasion (see Appendix K) .

Taken together, while interaction effects are estimated with limited precision, the robustness checks consistently point to the same conclusion: belief updating is anchored primarily in baseline beliefs, with partisan affect playing a secondary role and direct source effects remaining comparatively weak.

Implications. Much of the existing literature on polarization has focused on diagnosing its origins, tracing, among other things, how increasing electoral competition within and across parties, media fragmentation, and identity-driven reasoning entrench attitudes and deepen political divides. Although this work has illuminated the mechanisms by which polarization persists and deepens, its implications for intervention remain limited, since understanding *why* polarization occurs is not necessarily identical to identifying *how* it can be mitigated. Our study addresses this gap by shifting the focus from descriptive accounts of partisan entrenchment to an information-theoretic mechanism—surprise—that offers a testable interpretation of when belief updating is more likely to occur.

From a policy perspective, our findings open up new possibilities: rather than attempting to reduce polarization by merely correcting misinformation—which often proves limited or context-dependent—strategies that leverage expectation violations may, in some contexts, prove more effective in facilitating belief updating. Designing interventions that embed counter-stereotypical signals may help catalyze belief updating more effectively than conventional appeals to neutrality or balance.

At the same time, the use of surprise as a mechanism for belief updating is not without limits. Our results suggest that while surprise can facilitate engagement with information that might otherwise be dismissed, its effectiveness likely depends on its relative rarity. When used sparingly, expectation-violating information can provoke cognitive engagement and reappraisal. However, as with many informational interventions, the effect of surprise is likely subject to diminishing returns: with repeated exposure, even signals that initially violate expectations may lose their disruptive power and become assimilated into what individuals come to anticipate.

6 Conclusion

This paper examines how belief updating in polarized political contexts depends on the relationship between new information and individuals' prior beliefs. We operationalize *surprise* as outcome-based updating relative to baseline expectations and show that belief updating is strongly anchored in prior beliefs, with partisan affect playing a secondary but meaningful role.

Across both studies, participants tended to update their beliefs more strongly when new information diverged more from their baseline expectations, rather than simply when it originated from ideologically aligned or opposing media sources. In descriptive terms, Republicans exhibited stronger updating in response to favorable information attributed to a traditionally Democratic outlet, while Democrats updated more strongly in response to unfavorable information attributed to a traditionally Republican outlet. At the same time, direct source effects were modest, and belief updating remained tightly constrained by prior convictions and affective polarization.

Taken together, these findings suggest that political information exerts its greatest influence when it meaningfully diverges from individuals' baseline expectations, while also highlighting the limits of source-based belief updating in highly polarized environments. Thus, even in polarized settings, belief updating may not be foreclosed, but instead contingent on whether new information meaningfully violates what individuals have come to expect, pointing to new avenues for policy design.

References

Abramowitz, Alan I and Steven W Webster, "Negative partisanship: Why Americans dislike parties but behave like rabid partisans," *Political Psychology*, 2018, *39*, 119–135.

Arceneaux, Kevin and Martin Johnson, *Changing minds or changing channels?: Partisan news in an age of choice*, University of Chicago Press, 2013.

Ash, Elliott, Sergio Galletta, Dominik Hangartner, Yotam Margalit, and Matteo Pinna, "The effect of Fox News on health behavior during COVID-19," *Political Analysis*, 2024, *32* (2), 275–284.

Bail, Christopher A, Lisa P Argyle, Taylor W Brown, John P Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky, "Exposure to opposing views on social media can increase political polarization," *Proceedings of the National Academy of Sciences*, 2018, *115* (37), 9216–9221.

Bakshy, Eytan, Solomon Messing, and Lada A Adamic, "Exposure to ideologically diverse news and opinion on Facebook," *Science*, 2015, *348* (6239), 1130–1132.

Balliet, Daniel, Joshua M Tybur, Junhui Wu, Christian Antonellis, and Paul AM Van Lange, "Political ideology, trust, and cooperation: In-group favoritism among Republicans and Democrats during a US national election," *Journal of Conflict Resolution*, 2018, *62* (4), 797–818.

Bankert, Alexa, "Negative and positive partisanship in the 2016 US presidential elections," *Political Behavior*, 2021, *43* (4), 1467–1485.

Bavel, Jay J Van and Andrea Pereira, "The partisan brain: An identity-based model of political belief," *Trends in cognitive sciences*, 2018, *22* (3), 213–224.

Benjamin, Daniel J, "Errors in probabilistic reasoning and judgment biases," *Handbook of Behavioral Economics: Applications and Foundations 1*, 2019, *2*, 69–186.

Berinsky, Adam J, "Rumors and health care reform: Experiments in political misinformation," *British Journal of Political Science*, 2017, *47* (2), 241–262.

_ , *Political rumors: Why we accept misinformation and how to fight it*, Princeton University Press, 2023.

Bolsen, Toby, James N Druckman, and Fay Lomax Cook, "The influence of partisan motivated reasoning on public opinion," *Political Behavior*, 2014, *36*, 235–262.

Boxell, Levi, Matthew Gentzkow, and Jesse M Shapiro, "Cross-country trends in affective polarization," *Review of Economics and Statistics*, 2024, *106* (2), 557–565.

Bronnikov, Egor, "Beyond (Distorted) Posteriors: How Beliefs Spill Over into Action," *Working paper*, 2026.

_ **and Michalis Drouvelis**, "Feeling the Evidence: How Emotions Shape Asymmetric Belief Updating," *Working paper*, 2026.

Broockman, David E and Joshua L Kalla, "Consuming cross-cutting media causes learning and moderates attitudes: A field experiment with Fox News viewers," *The Journal of Politics*, 2025, *87* (1), 000–000.

Budak, Ceren, Brendan Nyhan, David M Rothschild, Emily Thorson, and Duncan J Watts, "Misunderstanding the harms of online misinformation," *Nature*, 2024, *630* (8015), 45–53.

Buidze, Tatia, Tobias Sommer, Ke Zhao, Xiaolan Fu, and Jan Gläscher, "Communication with Surprise–Computational Principles of Goal Signaling in Novel Human Interactions," *Working paper*, 2024.

Bullock, John G, "Partisan bias and the Bayesian ideal in the study of public opinion," *The Journal of Politics*, 2009, *71* (3), 1109–1124.

_ , "Elite influence on public opinion in an informed electorate," *American Political Science Review*, 2011, *105* (3), 496–515.

Carey, John M, Brian Fogarty, Marília Gehrke, Brendan Nyhan, and Jason Reifler, "Prebunking and credible source corrections increase election credibility: Evidence from the US and Brazil," *Science Advances*, 2025, *11* (35), eadv3758.

Carlin, Ryan E and Gregory J Love, "Political competition, partisanship and interpersonal trust in electoral democracies," *British Journal of Political Science*, 2018, *48* (1), 115–139.

Cohen, Geoffrey L, "Party over policy: The dominating impact of group influence on political beliefs.," *Journal of personality and social psychology*, 2003, *85* (5), 808.

Compton, Josh, Sander van der Linden, John Cook, and Melisa Basol, "Inoculation theory in the post-truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories," *Social and Personality Psychology Compass*, 2021, *15* (6), e12602.

Cover, Thomas M and Joy A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1999.

Dafoe, Allan, Baobao Zhang, and Devin Caughey, "Information equivalence in survey experiments," *Political Analysis*, 2018, *26* (4), 399–416.

Dalton, Russell J, Paul A Beck, and Robert Huckfeldt, "Partisan cues and the media: Information flows in the 1992 presidential election," *American Political Science Review*, 1998, *92* (1), 111–126.

Danz, David, Lise Vesterlund, and Alistair J Wilson, "Belief elicitation and behavioral incentive compatibility," *American Economic Review*, 2022, *112* (9), 2851–2883.

DellaVigna, Stefano and Ethan Kaplan, "The Fox News effect: Media bias and voting," *The Quarterly Journal of Economics*, 2007, *122* (3), 1187–1234.

DeSilver, Drew, "The polarization in today's Congress has roots that go back decades," Technical Report, Pew Research Center 2022.

Donovan, Kathleen, Paul M Kellstedt, Ellen M Key, and Matthew J Lebo, "Motivated reasoning, public opinion, and presidential approval," *Political Behavior*, 2020, *42*, 1201–1221.

Druckman, James N, "On the limits of framing effects: Who can frame?," *The journal of politics*, 2001, *63* (4), 1041–1066.

_ and Mary C McGrath, "The evidence for motivated reasoning in climate change preference formation," *Nature Climate Change*, 2019, *9* (2), 111–119.

_ , Erik Peterson, and Rune Slothuus, "How elite partisan polarization affects public opinion formation," *American Political Science Review*, 2013, *107* (1), 57–79.

_ , Samara Klar, Yanna Krupnikov, Matthew Levendusky, and John Barry Ryan, "Affective polarization, local contexts and public opinion in America," *Nature human behaviour*, 2021, *5* (1), 28–38.

_ , SR Gubitz, Ashley M Lloyd, and Matthew S Levendusky, "How incivility on partisan media (de) polarizes the electorate," *The Journal of Politics*, 2019, *81* (1), 291–295.

Flynn, Daniel J, Brendan Nyhan, and Jason Reifler, "The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics," *Political Psychology*, 2017, *38*, 127–150.

Frederick, Shane, "Cognitive reflection and decision making," *Journal of Economic Perspectives*, 2005, *19* (4), 25–42.

Garrett, R Kelly and Robert M Bond, "Conservatives' susceptibility to political misperceptions," *Science Advances*, 2021, *7* (23), eabf1234.

_, Dustin Carnahan, and Emily K Lynch, "A turn toward avoidance? Selective exposure to online political information, 2004–2008," *Political Behavior*, 2013, *35*, 113–134.

Gerber, Alan S, Gregory A Huber, and Ebonya Washington, "Party affiliation, partisanship, and political beliefs: A field experiment," *American Political Science Review*, 2010, *104* (4), 720–744.

Goldberg, Matthew H, Abel Gustafson, Seth A Rosenthal, and Anthony Leiserowitz, "Shifting Republican views on climate change through targeted advertising," *Nature Climate Change*, 2021, *11* (7), 573–577.

Green, Jon, Stefan McCabe, Sarah Shugars, Hanyu Chwe, Luke Horgan, Shuyang Cao, and David Lazer, "Curation bubbles," *American Political Science Review*, 2025, *119* (4), 1704–1722.

Grether, David M, "Bayes rule as a descriptive model: The representativeness heuristic," *The Quarterly journal of economics*, 1980, *95* (3), 537–557.

Guay, Brian and Christopher D Johnston, "Ideological asymmetries and the determinants of politically motivated reasoning," *American Journal of Political Science*, 2022, *66* (2), 285–301.

Guess, Andrew M, "(Almost) everything in moderation: new evidence on Americans' online media diets," *American Journal of Political Science*, 2021, *65* (4), 1007–1022.

_, Brendan Nyhan, and Jason Reifler, "Exposure to untrustworthy websites in the 2016 US election," *Nature Human Behaviour*, 2020, *4* (5), 472–480.

Gunther, Albert C, Stephanie Edgerly, Heather Akin, and James A Broesch, "Partisan evaluation of partisan information," *Communication Research*, 2012, *39* (4), 439–457.

Guriev, Sergei and Daniel Treisman, *Spin dictators: The changing face of tyranny in the 21st century*, Princeton University Press, 2022.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart, "Designing information provision experiments," *Journal of Economic Literature*, 2023, *61* (1), 3–40.

Hill, Seth J, "Learning together slowly: Bayesian learning about political facts," *The Journal of Politics*, 2017, *79* (4), 1403–1418.

Iyengar, Shanto and Kyu S Hahn, "Red media, blue media: Evidence of ideological selectivity in media use," *Journal of communication*, 2009, *59* (1), 19–39.

_ **and Sean J Westwood**, "Fear and loathing across party lines: New evidence on group polarization," *American journal of political science*, 2015, *59* (3), 690–707.

_ **, Gaurav Sood, and Yphtach Lelkes**, "Affect, not ideology: A social identity perspective on polarization," *Public opinion quarterly*, 2012, *76* (3), 405–431.

_ **, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J Westwood**, "The origins and consequences of affective polarization in the United States," *Annual review of political science*, 2019, *22* (1), 129–146.

Jerit, Jennifer and Jason Barabas, "Partisan perceptual bias and the information environment," *The Journal of Politics*, 2012, *74* (3), 672–684.

Jurkowitz, Mark, Amy Mitchell, Elisa Shearer, and Mason Walker, "US media polarization and the 2020 election: A nation divided," *Pew Research Center*, 2020, *24.*

Kahan, Dan M, "Ideology, motivated reasoning, and cognitive reflection," *Judgment and Decision Making*, 2013, *8* (4), 407–424.

Knobloch-Westerwick, Silvia and Matthias R Hastall, "Please your self: Social identity effects on selective exposure to news about in-and out-groups," *Journal of Communication*, 2010, *60* (3), 515–535.

Lane, Daniel S, Cassandra M Moxley, and Cynthia McLeod, "The group roots of social media politics: Social sorting predicts perceptions of and engagement in politics on social media," *Communication research*, 2023, *50* (7), 904–932.

Lazer, David MJ, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild et al., "The science of fake news," *Science*, 2018, *359* (6380), 1094–1096.

Lee, Amber Hye-Yon, Yphtach Lelkes, Carlee B Hawkins, and Alexander G Theodoridis, "Negative partisanship is not more prevalent than positive partisanship," *Nature human behaviour*, 2022, *6* (7), 951–963.

Levendusky, Matthew and Neil Malhotra, "Does media coverage of partisan polarization affect political attitudes?," *Political Communication*, 2016, *33* (2), 283–301.

Levendusky, Matthew S, "Why do partisan media polarize viewers?," *American journal of political science*, 2013, *57* (3), 611–623.

Li, Jianing and Michael W Wagner, "The value of not knowing: Partisan cue-taking and belief updating of the uninformed, the ambiguous, and the misinformed," *Journal of Communication*, 2020, *70* (5), 646–669.

Little, Andrew T, "How to Distinguish Motivated Reasoning from Bayesian Updating," *Political Behavior*, 2025, pp. 1–25.

_ , Keith E Schnakenberg, and Ian R Turner, "Motivated reasoning and democratic accountability," *American Political Science Review*, 2022, *116* (2), 751–767.

Lodge, Milton and Charles S Taber, *The rationalizing voter*, Cambridge University Press, 2013.

_ and Ruth Hamill, "A partisan schema for political information processing," *American political science review*, 1986, *80* (2), 505–519.

Loewenstein, Jeffrey, "Surprise, recipes for surprise, and social influence," *Topics in Cognitive Science*, 2019, *11* (1), 178–193.

Lois, Giannis, Elias Tsakas, and Arno Riedl, "Motivated reasoning or biased prior impressions: Updating trust towards partisan sources based on evidence," *Working paper*, 2023.

_ , _ , and _ , "Facts over partisanship: Evidence-based updating of trust in partisan sources," *Working Paper*, 2025.

McCarty, Nolan, Keith T Poole, and Howard Rosenthal, *Polarized America: The dance of ideology and unequal riches*, MIT Press, 2016.

Melnikoff, David E and Nina Strohminger, "Bayesianism and wishful thinking are compatible," *Nature Human Behaviour*, 2024, *8* (4), 692–701.

Merkley, Eric and Dominik A Stecula, "Party cues in the news: Democratic elites, Republican backlash, and the dynamics of climate skepticism," *British Journal of Political Science*, 2021, *51* (4), 1439–1456.

Messing, Solomon and Sean J Westwood, "Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online," *Communication research*, 2014, *41* (8), 1042–1063.

Miller, Joanne M, Kyle L Saunders, and Christina E Farhart, "Conspiracy endorsement as motivated reasoning: The moderating roles of political knowledge and trust," *American Journal of Political Science*, 2016, *60* (4), 824–844.

Nyhan, Brendan and Jason Reifler, "When corrections fail: The persistence of political misperceptions," *Political Behavior*, 2010, *32* (2), 303–330.

_ , Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Y Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery et al., "Like-minded sources on Facebook are prevalent but not polarizing," *Nature*, 2023, *620* (7972), 137–144.

_ , Jason Reifler, and Peter A Ubel, "The hazards of correcting myths about health care reform," *Medical care*, 2013, *51* (2), 127–132.

Ortoleva, Pietro, "Alternatives to bayesian updating," *Annual Review of Economics*, 2022, *16*.

Petersen, Michael Bang, Martin Skov, Søren Serritzlew, and Thomas Ramsøy, "Motivated reasoning and political parties: Evidence for increased processing in the face of party cues," *Political Behavior*, 2013, *35*, 831–854.

Peterson, Erik, "The role of the information environment in partisan voting," *The Journal of Politics*, 2017, *79* (4), 1191–1204.

_ and Ali Kagalwala, "When unfamiliarity breeds contempt: How partisan selective exposure sustains oppositional media hostility," *American Political Science Review*, 2021, *115* (2), 585–598.

_ and Maxwell B Allamong, "The influence of unknown media on public opinion: Evidence from local and foreign news sources," *American Political Science Review*, 2022, *116* (2), 719–733.

_ and Shanto Iyengar, "Partisan gaps in political information and information-seeking behavior: Motivated reasoning or cheerleading?," *American Journal of Political Science*, 2021, *65* (1), 133–147.

Poole, Keith T and Howard Rosenthal, "The polarization of American politics," *The Journal of Politics*, 1984, *46* (4), 1061–1079.

_ and _ , "A spatial model for legislative roll call analysis," *American Journal of Political Science*, 1985, pp. 357–384.

Prike, Toby, Robert Reason, Ullrich KH Ecker, Briony Swire-Thompson, and Stephan Lewandowsky, "Would I lie to you? Party affiliation is more important than Brexit in processing political misinformation," *Royal Society Open Science*, 2023, *10* (2), 220508.

Rabin, Matthew, "Psychology and economics," *Journal of economic literature*, 1998, *36* (1), 11–46.

— , "An approach to incorporating psychology into economics," *American Economic Review*, 2013, *103* (3), 617–622.

Roozenbeek, Jon and Sander Van der Linden, *The psychology of misinformation*, Cambridge University Press, 2024.

Schaffner, Brian F and Matthew J Streb, "The partisan heuristic in low-information elections," *Public Opinion Quarterly*, 2002, *66* (4), 559–581.

Schedler, Andreas, "Rethinking political polarization," *Political science quarterly*, 2023, *138* (3), 335–359.

Schroeder, Elizabeth and Daniel F Stone, "Fox news and political knowledge," *Journal of Public Economics*, 2015, *126*, 52–63.

Shannon, Claude E, "A mathematical theory of communication," *The Bell system technical journal*, 1948, *27* (3), 379–423.

Skytte, Rasmus, "The effect of real-news party cues," *American Journal of Political Science*, 2025.

Stroud, Natalie Jomini, "Media use and political predispositions: Revisiting the concept of selective exposure," *Political behavior*, 2008, *30*, 341–366.

— , "Polarization and partisan selective exposure," *Journal of communication*, 2010, *60* (3), 556–576.

— , *Niche news: The politics of news choice*, Oxford University Press, 2011.

Sweetser, Kaye D, "Partisan personality: The psychological differences between Democrats and Republicans, and independents somewhere in between," *American Behavioral Scientist*, 2014, *58* (9), 1183–1194.

Taber, Charles S and Milton Lodge, "Motivated skepticism in the evaluation of political beliefs," *American Journal of Political Science*, 2006, *50* (3), 755–769.

Tappin, Ben M, Gordon Pennycook, and David G Rand, "Bayesian or biased? Analytic thinking and political belief updating," *Cognition*, 2020, *204*, 104375.

Thaler, Michael, "Gender differences in motivated reasoning," *Journal of Economic Behavior & Organization*, 2021, *191*, 501–518.

— , "The fake news effect: Experimentally identifying motivated reasoning using trust in news," *American Economic Journal: Microeconomics*, 2024, *16* (2), 1–38.

Tversky, Amos and Daniel Kahneman, "The framing of decisions and the psychology of choice," *science*, 1981, *211* (4481), 453–458.

Appendix

Table of Contents

A Details on the Log-likelihood Ratio

Here we present the graph of the log-likelihood ratio (LLR) and elaborate on the example given in the text.

A.1 Derivation Details

The LLR follows directly from Bayes' rule. Recall from eq. (1) that the posterior probability of Trump winning is

$$\mathbb{P}(\text{Trump} \mid S) = \frac{\mathbb{P}(S \mid \text{Trump}) \cdot \mathbb{P}(\text{Trump})}{\mathbb{P}(S \mid \text{Trump}) \cdot \mathbb{P}(\text{Trump}) + \mathbb{P}(S \mid \text{Harris}) \cdot \mathbb{P}(\text{Harris})}. \tag{3}$$

Analogously,

$$\mathbb{P}(\text{Harris} \mid S) = \frac{\mathbb{P}(S \mid \text{Harris}) \cdot \mathbb{P}(\text{Harris})}{\mathbb{P}(S \mid \text{Trump}) \cdot \mathbb{P}(\text{Trump}) + \mathbb{P}(S \mid \text{Harris}) \cdot \mathbb{P}(\text{Harris})}. \tag{4}$$

Taking the ratio of posterior beliefs cancels the common denominator and yields

$$\frac{\mathbb{P}(\text{Trump} \mid S)}{\mathbb{P}(\text{Harris} \mid S)} = \frac{\mathbb{P}(S \mid \text{Trump})}{\mathbb{P}(S \mid \text{Harris})} \cdot \frac{\mathbb{P}(\text{Trump})}{\mathbb{P}(\text{Harris})}. \tag{5}$$

Taking logarithms gives

$$\log\left(\frac{\mathbb{P}(\text{Trump} \mid S)}{\mathbb{P}(\text{Harris} \mid S)}\right) = \log\left(\frac{\mathbb{P}(\text{Trump})}{\mathbb{P}(\text{Harris})}\right) + \log\left(\frac{\mathbb{P}(S \mid \text{Trump})}{\mathbb{P}(S \mid \text{Harris})}\right). \tag{6}$$

Rearranging yields eq. (2), so the LLR is the difference between log-posterior odds and log-prior odds. Since prior and posterior beliefs are directly elicited, the LLR can be computed from reported beliefs.

A.2 Example Explained

In Section 2, we provide the following example, emphasizing the non-linear nature of the LLR function: the LLR obtained when beliefs are updated from 80% to 90% is much larger than the LLR obtained when beliefs are updated from 50% to 60%, even though both represent a 10 percentage-point increase.

This is because the log-likelihood ratio (LLR) is a non-linear transformation of beliefs defined as the change in log-odds from prior to posterior beliefs. As prior beliefs become more extreme (closer to 0 or 1), the same absolute change in probability reflects a larger shift in log-odds space. More rigorously, we can see that the extent of belief updating

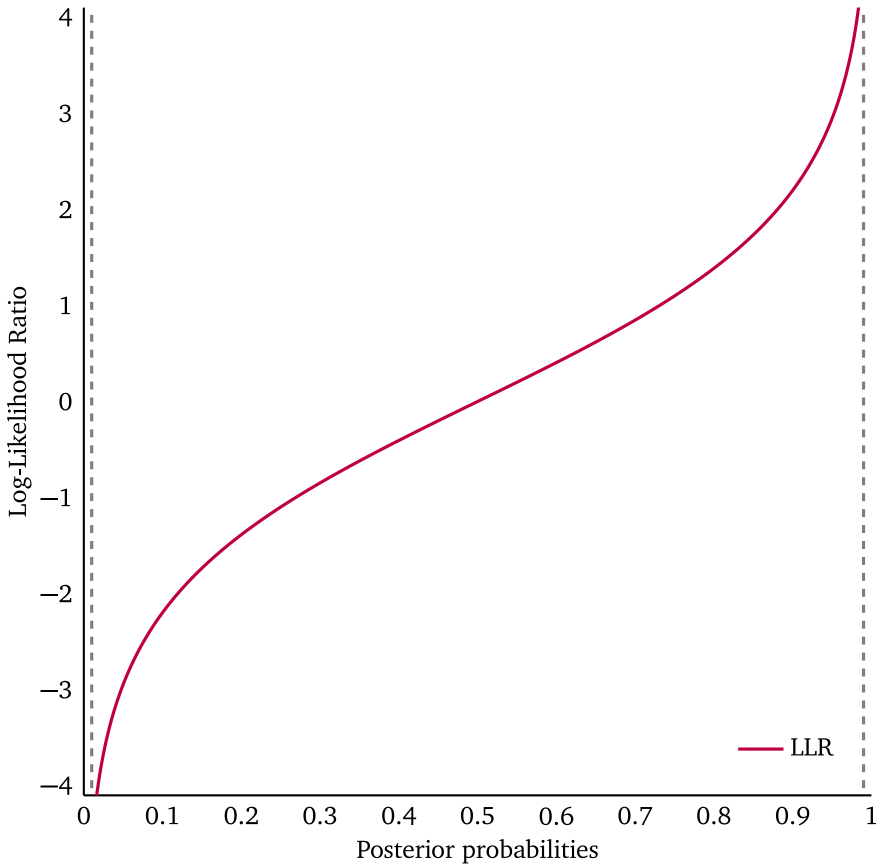


Figure 2: The Log-Likelihood Ratio for a binary Sample Space.

Note: Fixing prior beliefs, the purple curve shows how the log-likelihood ratio varies with posterior beliefs. The graph is undefined at $p = 0$ and $p = 1$ due to division by zero. Dashed gray lines represent the vertical asymptotes (as the posterior cannot be 0 or 1). If we change the posterior while keeping the priors fixed, the value of the LLR changes according to the purple graph. However, if we change the values of the priors while keeping the posteriors fixed, the LLR graph shifts vertically: if the priors become lower, the LLR graph shifts upwards; if the priors become higher, the LLR graph shifts downwards. *Source:* Bronnikov (2026).

from 50% to 60%

$$\mathrm{LLR}_{50\to 60} = \log\left(\frac{0.60}{1-0.60}\right) - \log\left(\frac{0.50}{1-0.50}\right) \tag{7}$$

$$= \log(1.5) - \log(1) = \log(1.5) \approx 0.405 \tag{8}$$

is (very) different from belief updating from 80% to 90%

$$\mathrm{LLR}_{80\to 90} = \log\left(\frac{0.90}{1-0.90}\right) - \log\left(\frac{0.80}{1-0.80}\right) \tag{9}$$

$$= \log(9) - \log(4) \approx 0.811 \tag{10}$$

Thus, an increase from 80% to 90% represents a much stronger signal (in terms of belief updating) than an increase from 50% to 60%, since it requires more weight of evidence to move already strong prior beliefs further in the same direction.

B Sample

B.1 Sample in Study A

Study A includes 277 respondents. Participants are 42.3 years old on average (SD = 14.4; range 18–78). The sample is 61.4% female and 37.9% male (0.7% unavailable). Education levels are relatively high: 56.7% hold at least a bachelor's degree (35.7% BA, 17.3% MA). A majority identify as White (72.9%) and were born in the United States (91.3%). In terms of employment, 48.0% are employed full-time, 9.4% are unemployed, and 13.7% identify as students.

B.2 Sample in Study B

Study B includes 275 respondents. Participants are 40.9 years old on average (SD = 13.9; range 18–80). The sample is 58.5% female and 40.7% male (0.7% unavailable). Overall, 54.2% hold at least a bachelor's degree (31.6% BA, 18.5% MA). Most respondents identify as White (70.2%) and were born in the United States (88.4%). Regarding employment status, 53.1% are employed full-time, 6.9% are unemployed, and 14.9% report being students.

C Elements of Instructions: Priors/Posteriors and signals

C.1 How Priors/Posteriors Were Introduced

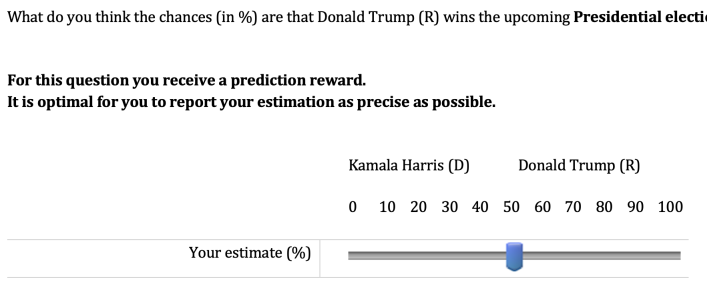


Figure 3: Beliefs Elicitation Screenshot.

Note: The screenshot captures how both prior and posterior beliefs were elicited during the study.

C.2 Study A: How Signal Was Introduced

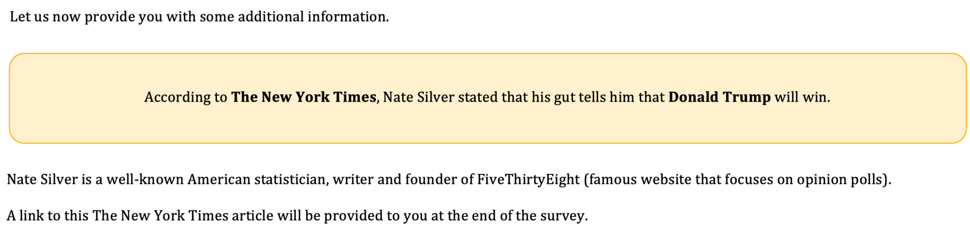


Figure 4: Signal communicating Trump's high chances of winning reported by the *New York Times*.

Note: The screenshot exemplifies how the signal (signal) was introduced to participants in one of the two treatments in Study A.

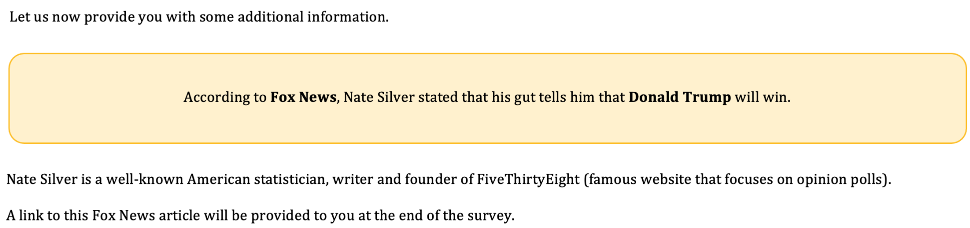


Figure 5: Signal communicating Trump's high chances of winning reported by *Fox News*.

Note: The screenshot exemplifies how the signal (signal) was introduced to participants in one of the two treatments in Study A.

C.3 Study B: How Signal Was Introduced

Let us now provide you with some additional information.

According to **The New York Times**, Allan Lichtman predicted that **Kamala Harris** will be the next president of the United States.

Allan Lichtman is a well-known American professor and forecaster who has predicted correctly 9 out of the last 10 elections. He was one of the very few people to correctly predict Donald Trump's win in 2016.

A link to this The New York Times article will be provided to you at the end of the survey.

Figure 6: Signal communicating Trump's low chances of winning reported by the New York Times.

Note: The screenshot exemplifies how the signal (signal) was introduced to participants in one of the two treatments in Study B.

Let us now provide you with some additional information.

According to **Fox News**, Allan Lichtman predicted that **Kamala Harris** will be the next president of the United States.

Allan Lichtman is a well-known American professor and forecaster who has predicted correctly 9 out of the last 10 elections. He was one of the very few people to correctly predict Donald Trump's win in 2016.

A link to this FOX News article will be provided to you at the end of the survey.

Figure 7: Signal communicating Trump's low chances of winning reported by *Fox News*.

Note: The screenshot exemplifies how the signal (signal) was introduced to participants in one of the two treatments in Study B.

D Beliefs Distributions

D.1 Study A

D.1.1 Prior Beliefs

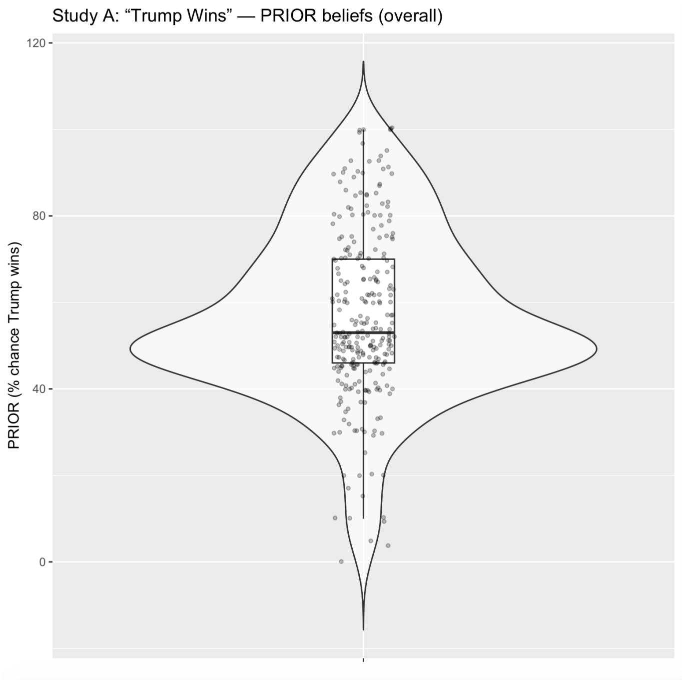


Figure 8: Overall distribution of prior beliefs in Study A.

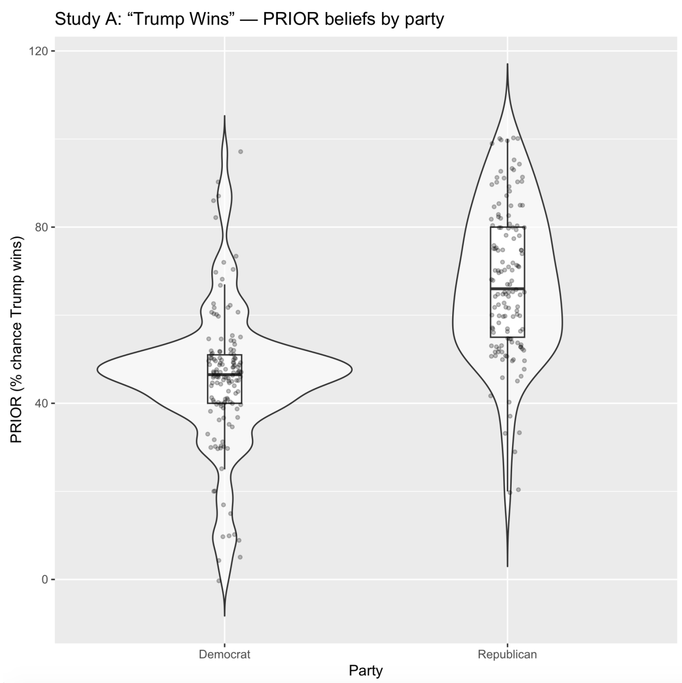


Figure 9: Distribution of prior beliefs in Study A by party.

D.1.2 Posterior Beliefs

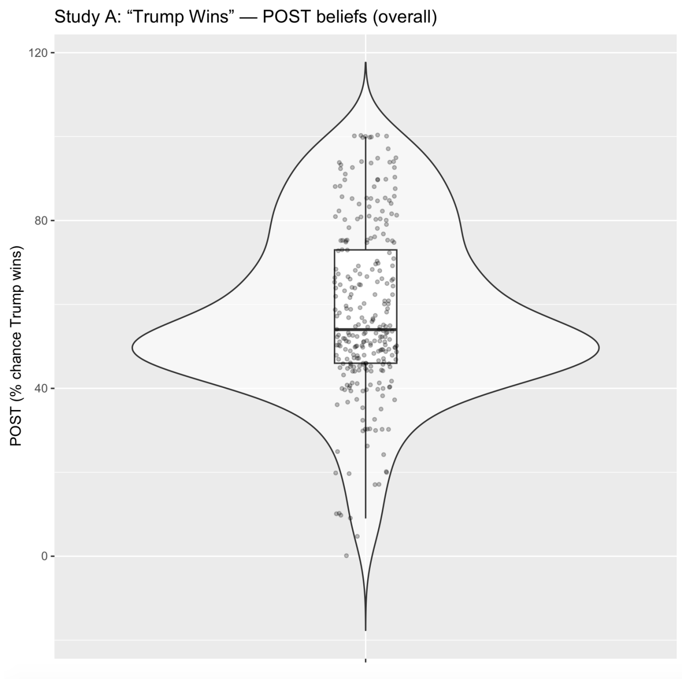


Figure 10: Overall distribution of posterior beliefs in Study A.

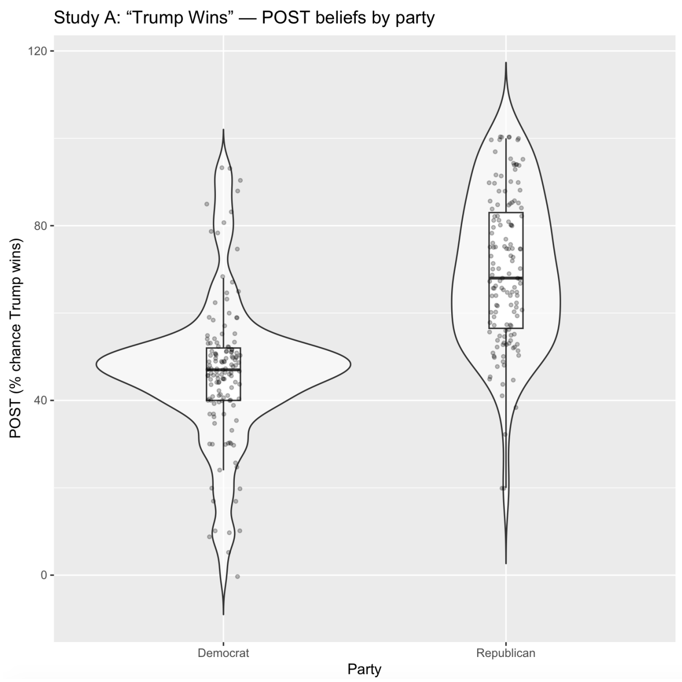


Figure 11: Distribution of posterior beliefs in Study A by party.

D.2 Study B

D.2.1 Prior Beliefs

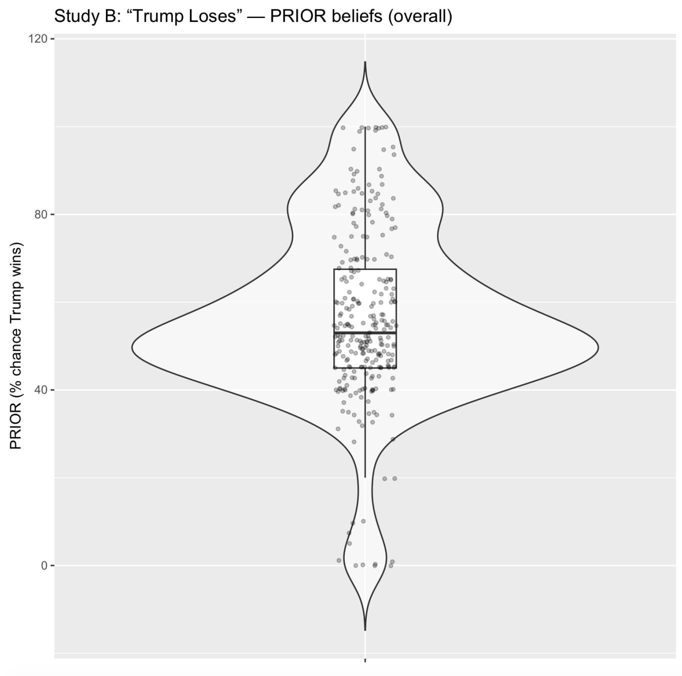


Figure 12: Overall distribution of prior beliefs in Study B.

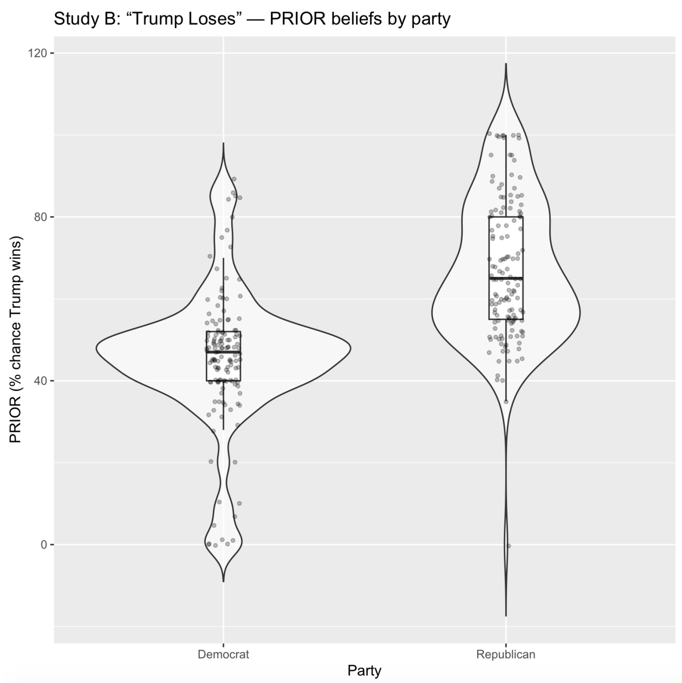


Figure 13: Distribution of prior beliefs in Study B by party.

D.2.2 Posterior Beliefs

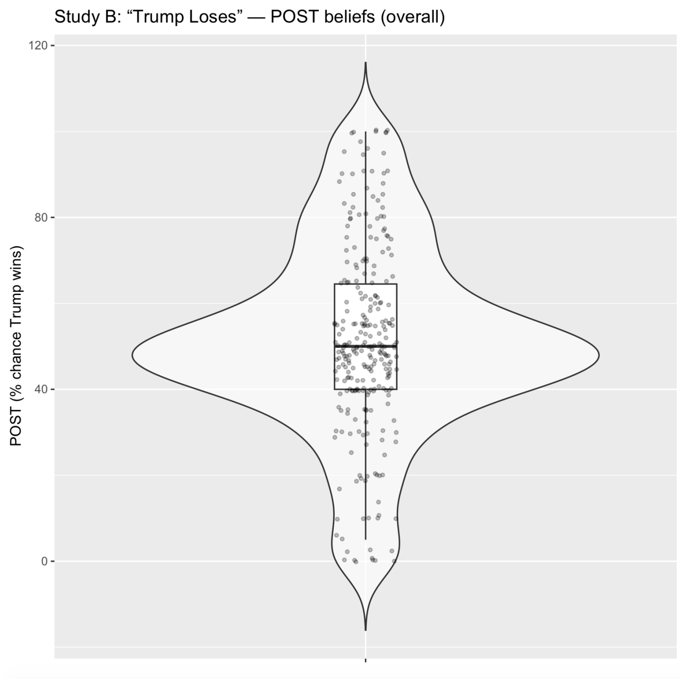


Figure 14: Overall distribution of posterior beliefs in Study B.

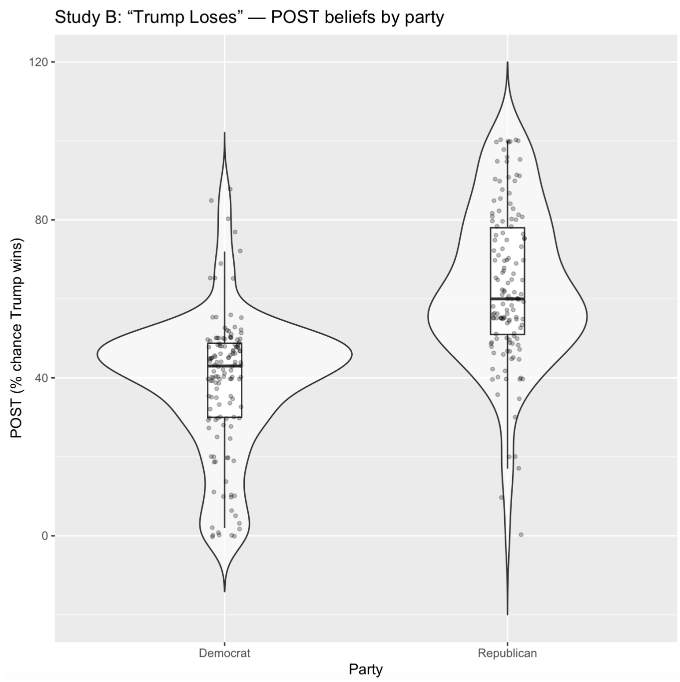


Figure 15: Distribution of posterior beliefs in Study B by party.

E Comparison Over Party

E.1 Study A

We compare belief updating between Democrats and Republicans using the Wilcoxon rank-sum (Mann–Whitney) test under two source conditions: *The New York Times* and *Fox News*.

First, we consider the condition in which the signal is attributed to *The New York Times* (Table 7). The difference in LLR distributions across parties is statistically significant ($p = 0.031$). Republicans have a higher rank sum than Democrats, indicating that Republicans updated their beliefs more positively (or less negatively) than Democrats in response to the signal.

Table 7: Wilcoxon Rank-Sum Test Results (SOURCE = 0, The New York Times)

PARTY	Obs	Rank Sum	Expected
Democrats (0)	68	4098.5	4556
Republicans (1)	65	4812.5	4355
Combined	133	8911	8911
Test Statistics:			
Unadjusted Variance: 49356.67			
Adjustment for Ties: -4531.78			
Adjusted Variance: 44824.88			
Test Statistic (z): -2.161			
p-value (Prob $> \lvert z \rvert$): 0.0307			
Exact p-value: 0.0305			

Next, we examine the condition in which the signal is attributed to *Fox News* (Table 8). The difference across parties is not statistically significant at conventional levels ($p = 0.055$), although the rank sums again suggest somewhat larger updates among Republicans.

Overall, the results for Study A indicate that Republicans tended to exhibit more positive belief updating than Democrats in response to the pro-Trump signal, particularly when the signal was attributed to *The New York Times*. This pattern is consistent with the interpretation that the signal aligned more closely with Republicans' prior beliefs, while conflicting with those of Democrats.

Table 8: Wilcoxon Rank-Sum Test Results (SOURCE = 1, Fox News)

PARTY	Obs	Rank Sum	Expected
Democrats (0)	69	4268	4692
Republicans (1)	66	4912	4488
Combined	135	9180	9180
Test Statistics:			
Unadjusted Variance: 51612.00			
Adjustment for Ties: -2950.79			
Adjusted Variance: 48661.21			
Test Statistic (z): -1.922			
p-value (Prob $>$ $\|z\|$): 0.0546			
Exact p-value: 0.0545			

E.2 Study B

We conduct the same comparison in Study B, again using the Wilcoxon rank-sum test to examine differences in belief updating between Democrats and Republicans under the two source conditions.

First, when the signal predicting Trump's loss is attributed to *The New York Times* (Table 9), the difference across parties is not statistically significant at the 5% level ($p =$ 0.069). However, Republicans again have a higher rank sum than Democrats, indicating somewhat less negative belief updating among Republicans.

Table 9: Wilcoxon Rank-Sum Test Results (SOURCE = 0, The New York Times)

PARTY	Obs	Rank Sum	Expected
Democrats (0)	64	3755	4128
Republicans (1)	64	4501	4128
Combined	128	8256	8256
Test Statistics:			
Unadjusted Variance: 44032.00			
Adjustment for Ties: -2045.10			
Adjusted Variance: 41986.90			
Test Statistic (z): -1.820			
p-value (Prob $>$ $\|z\|$): 0.0687			
Exact p-value: 0.0688			

Next, when the signal is attributed to *Fox News* (Table 10), the difference across parties is statistically significant ($p = 0.022$). Republicans have a higher rank sum than Democrats, indicating less negative belief updating in response to the signal.

Taken together, the results in Study B again suggest systematic differences in belief updating across partisan groups. Republicans tend to exhibit less negative updating than Democrats when confronted with information predicting Trump's electoral loss. This

Table 10: Wilcoxon Rank-Sum Test Results (SOURCE = 1, Fox News)

PARTY	**Obs**	**Rank Sum**	**Expected**
Democrats (0)	68	4031	4522
Republicans (1)	64	4747	4256
Combined	132	8778	8778
Test Statistics:			
Unadjusted Variance: 48234.67			
Adjustment for Ties: -2623.10			
Adjusted Variance: 45611.56			
Test Statistic (z): -2.299			
p-value (Prob $> \|z\|$): 0.0215			
Exact p-value: 0.0212			

pattern is consistent with the interpretation that the signal diverges more strongly from Republicans' prior beliefs than from those of Democrats.

F Comparison Over Source

F.1 Study A

We examine whether belief updating differs across the two sources of the signal, *The New York Times* and *Fox News*, separately for Democrats and Republicans using the Wilcoxon rank-sum (Mann–Whitney) test.

First, we consider Democrats (Table 11). The difference in LLR distributions across sources is not statistically significant ($p = 0.254$). Although the rank sum is slightly higher when the signal is attributed to *The New York Times*, the difference is small and statistically indistinguishable from zero. This indicates that Democrats updated their beliefs similarly regardless of whether the signal was attributed to *The New York Times* or *Fox News*. Consistent with the interpretation developed in the main text, the signal predicting Trump's victory likely generated limited informational surprise for Democrats, as it sharply conflicted with their prior beliefs.

Table 11: Wilcoxon Rank-Sum Test Results for Democrats (SOURCE Comparison, Study A)

SOURCE	**Obs**	**Rank Sum**	**Expected**
The New York Times (0)	68	4940.5	4692
Fox News (1)	69	4512.5	4761
Combined	137	9453	9453
Test Statistics:			
Unadjusted Variance: 53958.00			
Adjustment for Ties: -6602.22			
Adjusted Variance: 47355.78			
Test Statistic (z): 1.142			
p-value (Prob $> \|z\|$): 0.2535			
Exact p-value: 0.2551			

Next, we consider Republicans (Table 12). Again, the difference across sources is not statistically significant ($p = 0.4387$). Although the rank sum is somewhat higher when the signal is attributed to *The New York Times*, the difference is not statistically meaningful. Republicans therefore updated their beliefs similarly across the two source conditions. This pattern is consistent with the idea that the pro-Trump signal aligned with Republicans' priors and therefore contained relatively little informational surprise.

Overall, the results indicate no statistically significant differences in belief updating across sources in Study A. This suggests that the content of the signal played a more important role than the attributed source in shaping responses. In line with the paper's information-theoretic interpretation, belief updating appears primarily related to the informational distance between the signal and prior beliefs.

Table 12: Wilcoxon Rank-Sum Test Results for Republicans (SOURCE Comparison, Study A)

SOURCE	**Obs**	**Rank Sum**	**Expected**
The New York Times (0)	65	4455	4290
Fox News (1)	66	4191	4356
Combined	131	8646	8646
Test Statistics:			
Unadjusted Variance: 47190.00			
Adjustment for Ties: -1788.55			
Adjusted Variance: 45401.45			
Test Statistic (z): 0.774			
p-value (Prob $> \|z\|$): 0.4387			
Exact p-value: 0.4407			

F.2 Study B

We perform the same comparison in Study B, examining whether belief updating differs across sources for Democrats and Republicans using the Wilcoxon rank-sum test.

For Democrats (Table 13), the difference across sources is not statistically significant ($p = 0.469$). The rank sums for the two sources are very similar, indicating that Democrats updated their beliefs to a comparable extent whether the signal was attributed to *The New York Times* or *Fox News*. This pattern is consistent with the interpretation that the signal predicting Trump's loss largely confirmed Democrats' prior beliefs and therefore generated limited informational surprise.

Table 13: Wilcoxon Rank-Sum Test Results for Democrats (SOURCE Comparison, Study B)

SOURCE	**Obs**	**Rank Sum**	**Expected**
The New York Times (0)	64	4412.5	4256
Fox News (1)	68	4365.5	4522
Combined	132	8778	8778
Test Statistics:			
Unadjusted Variance: 48234.67			
Adjustment for Ties: -1554.86			
Adjusted Variance: 46679.81			
Test Statistic (z): 0.724			
p-value (Prob $> \|z\|$): 0.4688			
Exact p-value: 0.4708			

For Republicans (Table 14), the difference across sources is also statistically insignificant ($p = 0.8848$). The nearly identical rank sums indicate that Republicans responded similarly regardless of whether the signal was attributed to *The New York Times* or *Fox News*. This suggests that the anti-Trump signal did not produce systematically different belief updating across source conditions.

Table 14: Wilcoxon Rank-Sum Test Results for Republicans (SOURCE Comparison, Study B)

SOURCE	**Obs**	**Rank Sum**	**Expected**
The New York Times (0)	64	4157.5	4128
Fox News (1)	64	4098.5	4128
Combined	128	8256	8256
Test Statistics:			
Unadjusted Variance: 44032.00			
Adjustment for Ties: -3305.83			
Adjusted Variance: 40726.17			
Test Statistic (z): 0.146			
p-value (Prob $> \|z\|$): 0.8838			
Exact p-value: 0.8852			

Taken together, these results show no statistically significant source differences in Study B. As in Study A, the findings suggest that belief updating is primarily related to how the signal aligns with or deviates from prior beliefs, rather than to the identity of the attributed source alone.

G Source Classifications

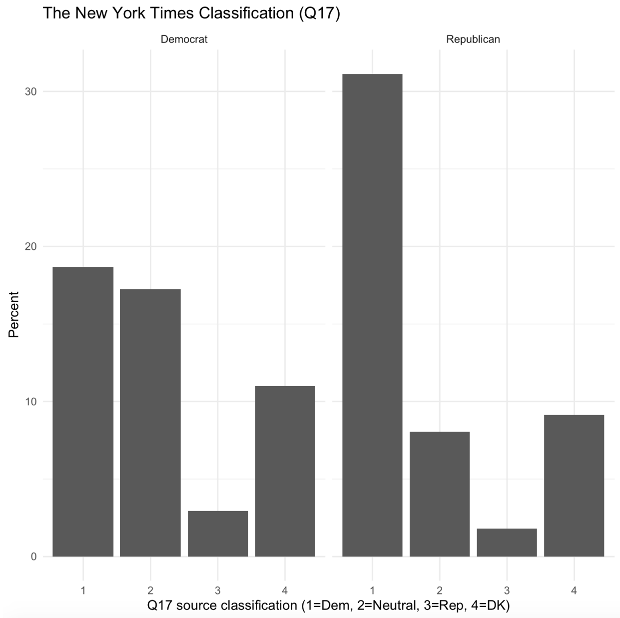


Figure 16: *The New York Times* Classification: How Democrats (left panel) and Republicans (right panel) classified the source.

Note: The Source Classification question asks participants to classify the source from which the signal was provided as (1) Democrat, (2) Neutral, (3) Republican source, or (4) I do not know. This graph captures the differences by party: self-identified Democrats are on the left side, and self-identified Republicans are on the right side of the histogram.

We first examine how participants classify the ideological orientation of the information source (Q17), conditional on the assigned outlet (see Figures 16 and 17). When the signal was attributed to *The New York Times*, Democrats and Republicans differed significantly in their perceptions of the source (Wilcoxon rank-sum test: $W = 11290$, $p = 0.001$). A similar but substantially stronger partisan divergence emerges when the signal was attributed to *Fox News* ($W = 12328$, $p < 10^{-6}$). Pearson chi-square tests confirm these patterns: source classification is strongly associated with party affiliation both for *The New York Times* ($\chi^2(3) = 18.70$, $p < 0.001$) and for *Fox News* ($\chi^2(3) = 35.65$, $p < 10^{-6}$).

Next, we test whether participants' classifications differ across treatments, that is, whether the same outlet is perceived differently depending on whether it delivers favorable or unfavorable information about Trump (see Figures 18 and 19). Pooling across parties, we find extremely strong treatment effects in both studies. In Study A ("Trump wins"), Q17 differs sharply between *The New York Times* and *Fox News* conditions ($W = 5315.5$, $p < 10^{-10}$). An equally strong difference appears in Study B ("Trump loses"; $W = 5001.5$, $p < 10^{-11}$). Chi-square tests corroborate these findings, revealing a very

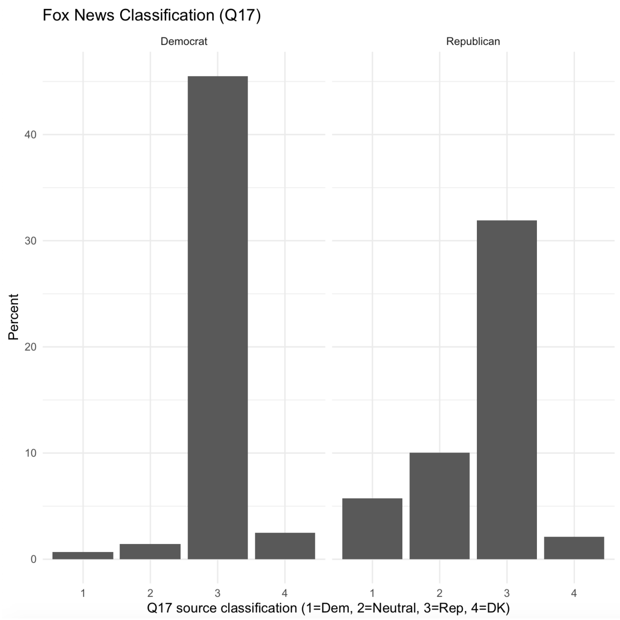


Figure 17: *Fox News* Classification: How Democrats (left panel) and Republicans (right panel) classified the source.

Note: The Source Classification question asks participants to classify the source from which the signal was provided as (1) Democrat, (2) Neutral, (3) Republican source, or (4) I do not know. This graph captures the differences by party: self-identified Democrats are on the left side, and self-identified Republicans are on the right side of the histogram.

strong association between source attribution and classification in both Study A ($\chi^2(3) = 162.33$, $p < 2 \times 10^{-16}$) and Study B ($\chi^2(3) = 148.67$, $p < 2 \times 10^{-16}$).

Importantly, these treatment effects persist within each party. In Study A, Democrats classify *The New York Times* and *Fox News* significantly differently ($W = 1126$, $p < 10^{-7}$), as do Republicans ($W = 1490$, $p < 0.001$). The same pattern holds in Study B: Democrats ($W = 1322$, $p < 10^{-5}$) and Republicans ($W = 1121$, $p < 10^{-7}$) both exhibit highly significant differences in their classifications across sources.

Taken together, these results demonstrate that participants possess clear beliefs about the ideological orientation of media outlets. Both Democrats and Republicans systematically distinguish between *The New York Times* and Fox News, and they do so consistently across experimental contexts.

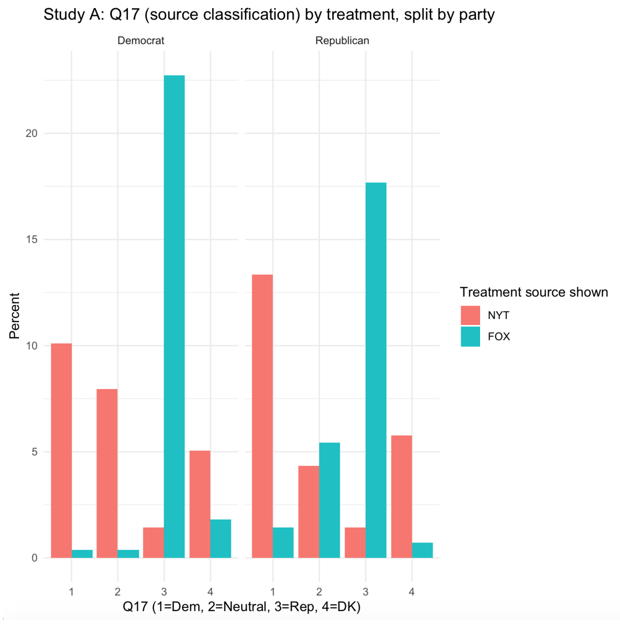


Figure 18: Source classification by treatment and party in Study A ("Trump Wins").

Note: The figure displays the percentage of respondents in each party who classified the information source as (1) Democrat, (2) Neutral, (3) Republican, or (4) Don't Know (Q17), separately by treatment source (New York Times vs. Fox News). Panels correspond to respondents' party affiliation (Democrat vs. Republican). Study A presents a favorable signal about Trump's electoral chances.

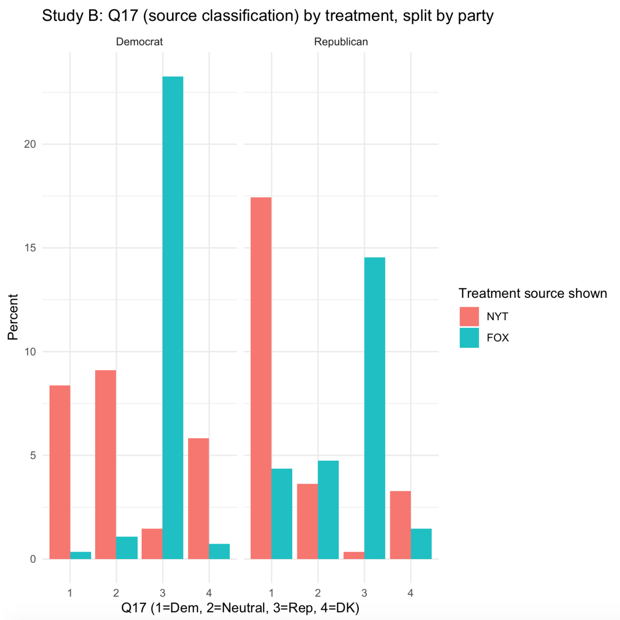


Figure 19: Source classification by treatment and party in Study B ("Trump Loses").

Note: The figure displays the percentage of respondents in each party who classified the information source as (1) Democrat, (2) Neutral, (3) Republican, or (4) Don't Know (Q17), separately by treatment source (New York Times vs. Fox News). Panels correspond to respondents' party affiliation (Democrat vs. Republican). Study B presents an unfavorable signal about Trump's electoral chances.

H Additional Measures And Controls

In addition to the main experimental design, we collected supplementary data to better understand participants' characteristics and how these traits may influence their belief updating. This additional data encompasses indices of positive and negative partisanship, assessments of cognitive reflection ability, and various socio-demographic factors.

The positive partisanship index captures the extent of participants' emotional attachment to and identification with their own political party. It is constructed from responses to seven questions, such as, "When I talk about my political party, I say 'we' instead of 'them", rated on a scale from 1 (strongly disagree) to 7 (strongly agree).[14] The index reflects the average score across these items and provides a measure of the participant's sense of belonging and commitment to their party.

The negative partisanship index quantifies participants' negative attitudes toward the opposing political party. It is based on responses to six questions, including, "I think people of the opposing party are bad people" and "I enjoy it when the opposing party does poorly in the polls".[15] These items, also rated on a 1-7 scale, are averaged to create a measure of the participant's emotional hostility and aversion to the out-group.

The Cognitive Reflection Test (CRT) is designed to assess an individual's ability to override intuitive responses and engage in reflective, analytical thinking. Initially introduced by Frederick (2005), the CRT typically presents participants with a series of short questions formulated in a way that often elicits a quickly developed, intuitive answer, which is usually incorrect. The challenge lies in the participant's ability to question their initial, instinctive responses and instead employ more deliberate reasoning to arrive at the correct answer. The CRT reveals how individuals handle problems requiring reflective reasoning, providing insight into the broader cognitive mechanisms of decision-making.[16] To account for potential demographic influences, we collected information on

[14]These statements are as follows: (i) My political party understands my concerns; (ii) My political party represents people like me; (iii) The members of my political party think like me; (iv) When I talk about my political party, I say 'we' instead of 'them'; (v) I care about what other people think about my party; (vi) It bothers me when my party does poorly in the polls; (vii) When I meet somebody who supports my party, I feel connected.

[15]These statements are as follows: (i) I enjoy it when the opposing party does poorly in the polls; (ii) I care about what other people think about the opposing party; (iii) When I meet somebody who supports the opposing party, I feel disconnected; (iv) When I hear somebody criticize the opposing party, it makes me feel good; (v) I think people of the opposing party are bad people; (vi) I dislike the opposing party more than I like my party.

[16]In our study, the following four questions were used: If you are running a race and you pass the person in second place, what place are you in? [Four answers were provided: (i) First, (ii) Second, (iii) Third, (iv) Not enough information]; A farmer had 15 sheep and all but 8 died. How many are left? [Four answers were provided: (i) 15, (ii) 8, (iii) 7, (iv) Not enough information]; Emily's father has three daughters. The first two are named April and May. What is the third daughter's name? [Four answers were provided: (i) June, (ii) July, (iii) Emily, (iv) Not enough information]; How many cubic feet of dirt are there in a hole that is 3' deep × 3' wide × 3' long? [Four answers were provided: (i) 27, (ii) 9, (iii) 0,

participants' age, highest level of education, and geographic location (state) during the study. These standard socio-demographic variables provide context for understanding variation in responses and belief updating across different population groups.

(iv) Not enough information].

I Variables Description

Here we provide the description of the variables used in the regressions.

Education ranges from Less than high school (1) to Professional degree (8).

Gender is Female (0), Male (1), or Unavailable (2).

Ethnicity includes White (0), Black (1), Asian (2), Mixed (3), Other (4), and Unavailable (5).

Birthplace is coded as US (0), Europe (1), Asia (2), Africa (3), Latin America (4), or Unavailable (5).

Language is English (0), Other (1), or Unavailable (2).

Student Status is No (0), Yes (1), or Unknown (2).

Employment Status includes Full-time (0), Part-time (1), Starting new job (2), Not paid/Homemaker/Retired/Disabled (3), Other (4), Unknown (5), and Unemployed (6).

J Regressions' Results with Recoded LLR

As a robustness check, we re-estimate the regressions reported in Tables 3 and 6 using a recoded version of the LLR that retains observations with extreme beliefs. Specifically, to avoid dropping respondents who reported priors or posteriors equal to 0 or 100—which would otherwise yield undefined log-odds—we recode these values to 0.1 and 99.9, respectively. We then recompute the log-likelihood ratio following eq. 2 using these adjusted beliefs. This procedure preserves the full sample, allowing us to assess whether our results are sensitive to the exclusion of extreme prior or posterior responses.

J.1 Study A

Table 15: Direct and Interaction Effects on the LLR in Study A (Recoded)

	Dependent variable:		
	LLR		
	(1)	(2)	(3)
Prior		−0.001	−0.002
		(0.003)	(0.003)
Source = Fox News	−0.068	−0.065	−0.048
	(0.068)	(0.067)	(0.068)
Party = Republicans	0.120*	0.149**	−0.212
	(0.062)	(0.075)	(0.262)
Positive Partisanship	0.107**	0.109**	0.059
	(0.044)	(0.043)	(0.041)
Negative Partisanship	−0.060	−0.061	−0.052*
	(0.045)	(0.044)	(0.031)
Fox News × Republicans			−0.047
			(0.135)
Republicans × Positive Partisanship			0.104
			(0.086)
Republicans × Negative Partisanship			−0.024
			(0.098)
Controls	✓	✓	✓
Observations	277	277	277
R^2	0.091	0.092	0.100

Note: This table presents regression results examining direct and interaction effects on the recoded LLR. Standard errors are shown in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 15 reports OLS regressions using the recoded LLR, where extreme prior and posterior beliefs (0 or 100) are adjusted to avoid undefined log-likelihood ratios and retained in the sample. This increases the number of observations from 268 to 277 and allows participants with absolute beliefs to contribute to the analysis.

Across specifications, prior beliefs no longer exhibit a statistically significant association with belief updating. In Columns (2) and (3), the coefficient on Prior is small and close to zero, indicating that once extreme responses are retained via recoding, the direct effect of prior beliefs on the magnitude of updating becomes substantially attenuated. This contrasts with the baseline LLR results, where higher priors were associated with significantly smaller updates, consistent with a surprise-based mechanism.

Partisan identity continues to matter. In Columns (1) and (2), Republicans exhibit significantly larger belief updates than Democrats (0.120 and 0.149, respectively), although this main effect disappears once interaction terms are introduced in Column (3). Positive partisanship is positively and significantly associated with belief updating in the first two specifications, suggesting that individuals with stronger favorable partisan affect respond more strongly to the signal. Negative partisanship enters with a negative sign and becomes weakly significant in the full model, indicating that stronger out-party animus is associated with smaller updates once partisan interactions are accounted for.

The source of the signal itself does not exert a statistically significant direct effect: receiving information from *Fox News* rather than *The New York Times* is associated with slightly smaller updates, but these differences are not distinguishable from zero in any specification. Likewise, none of the interaction terms in Column (3)—including *Fox News* × Republicans or interactions between party identity and partisan affect—reach conventional significance levels. This suggests limited evidence that source effects or partisan intensity systematically moderate belief updating once extreme beliefs are retained.

Relative to the original analysis, which excluded respondents with priors or posteriors at 0 or 100, the recoded specification yields two important changes.

First, the previously significant negative effect of prior beliefs disappears. This indicates that the relationship between prior beliefs and updating identified in the baseline analysis—consistent with an information-theoretic notion of surprise—is driven largely by respondents with interior beliefs; individuals holding absolute priors behave differently, and their inclusion weakens the overall relationship between priors and updating.

Second, coefficients on partisan variables—especially positive partisanship—become larger in magnitude, while source effects remain negligible. This shift suggests that participants with extreme beliefs amplify the role of partisan dispositions in belief updating, even as the informational content of the signal itself continues to play a limited role.

Taken together, the recoded results confirm that partisan identity and affective attachments remain central predictors of belief updating in Study A, while the source of information has little independent influence. However, allowing respondents with extreme beliefs to enter the analysis substantially weakens the role of prior beliefs and reduces evidence for a pattern consistent with a surprise-based updating mechanism. This pattern implies that individuals holding absolute convictions respond to political information in a qualitatively different way: their updates are less shaped by surprise and more closely tied to partisan orientation. Overall, the findings underscore that belief updating appears more strongly associated with partisan variables than with source identity.

J.2 Study B

Table 16: Direct and Interaction Effects on the LLR in Study B (Recoded)

	Dependent variable:		
	LLR		
	(1)	(2)	(3)
Prior		-0.007^{*}	-0.008^{*}
		(0.004)	(0.004)
Source = Fox News	0.028	0.027	-0.062
	(0.115)	(0.115)	(0.153)
Party = Republicans	0.235^{*}	0.382^{***}	0.360
	(0.121)	(0.119)	(0.494)
Positive Partisanship	-0.089	-0.070	-0.136
	(0.057)	(0.056)	(0.110)
Negative Partisanship	0.072^{*}	0.063	0.145^{**}
	(0.042)	(0.040)	(0.060)
Fox News × Republicans			0.259
			(0.218)
Republicans × Positive Partisanship			0.145
			(0.116)
Republicans × Negative Partisanship			-0.191^{**}
			(0.088)
Controls	✓	✓	✓
Observations	275	275	275
R^2	0.065	0.083	0.102

Note: This table presents regression results examining direct and interaction effects on the recoded LLR. Standard errors are shown in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 16 reports OLS regressions using the recoded LLR, which retains respondents with extreme prior or posterior beliefs by adjusting values at 0 and 100. This increases the sample from 260 to 275 observations and allows individuals with absolute expectations about Trump's electoral prospects to enter the analysis.

As in Study A, prior beliefs remain negatively associated with belief updating once recoded values are included, although the magnitude is smaller than in the original specification. In Columns (2) and (3), the coefficient on Prior is approximately −0.007 to −0.008 and statistically significant at the 5% level, indicating that participants whose initial beliefs were closer to the unfavorable signal update less. This finding continues to support a pattern consistent with a surprise-based mechanism, though its strength is attenuated relative to the baseline results.

Partisan identity plays a prominent role. Republicans exhibit significantly larger updates than Democrats in Columns (1) and (2) (0.235 and 0.382, respectively), although this main effect becomes statistically insignificant in Column (3) once interaction terms are introduced. Positive partisanship enters with a negative sign across all specifications but does not reach conventional significance. By contrast, negative partisanship is positively associated with belief updating and becomes strongly significant in the full model (0.145 in Column (3)), suggesting that stronger out-party animus is linked to larger responses to unfavorable information.

Turning to source effects, receiving the signal from *Fox News* does not exert a statistically significant direct impact on belief updating in any specification. In the full model, the *Fox News* × Republicans interaction is positive but not significant, indicating limited evidence that Republicans respond more strongly to unfavorable news when it originates from *Fox News* once extreme beliefs are retained.

Notably, the interaction between Republican identity and negative partisanship is negative and statistically significant in Column (3). This suggests that among Republicans, higher levels of negative partisanship dampen belief updating in response to unfavorable information, pointing to a potential defensive response among those with stronger affective hostility toward the opposing party.

Relative to the baseline Study B results, which excluded respondents with priors or posteriors equal to 0 or 100, the recoded specification produces several meaningful shifts.

First, while prior beliefs remain statistically significant, their effect is weaker in magnitude, indicating that individuals with extreme beliefs partially dilute the overall surprise effect. Second, the role of partisan identity becomes more pronounced in the simpler specifications, with larger Republican coefficients in Columns (1) and (2). Third, the moderating role of negative partisanship changes: whereas the original analysis highlighted a positive *Fox News* × Republicans interaction, the recoded model instead emphasizes a significant Republicans × Negative Partisanship interaction, suggesting that affective polarization rather than source alignment becomes the dominant conditioning factor once extreme respondents are included.

At the same time, direct source effects remain negligible in both versions of the analysis, reinforcing the conclusion that the identity of the media outlet plays a limited independent role in shaping belief updates.

Overall, the recoded results for Study B corroborate the central role of partisan identity and affective polarization in shaping responses to unfavorable political information. Although prior beliefs continue to matter, their influence weakens once participants with absolute expectations are retained. Instead, negative partisanship emerges as a key moderator, particularly among Republicans, indicating that emotional opposition to the out-party constrains belief updating in the face of adverse news. As in Study A, these findings suggest that belief updating is driven less by source credibility and more by partisan psychology, especially when individuals hold strong prior convictions.

K Robustness to Alternative Model Specifications

To assess whether our results are sensitive to modeling belief updating as a change score, we conduct an additional robustness analysis using an ANCOVA specification that conditions directly on baseline beliefs, following recommendations in Nyhan (2021). Rather than estimating belief updating via the log-likelihood ratio, this approach regresses posterior beliefs in log-odds form on prior beliefs in log-odds form, treatment indicators, partisanship measures, their interactions, and the same set of demographic controls used in the main analyses. Formally, we estimate:

$$\log\left(\frac{\text{Posterior}_i}{100-\text{Posterior}_i}\right) = \alpha + \beta_1 \log\left(\frac{Prior_i}{100-Prior_i}\right) + \beta_2 Source_i + \tag{11}$$

$$\beta_3 Party_i + \beta_4 (Source_i \times Party_i) + \mathbf{X}_i^{\top}\gamma + \varepsilon_i, \tag{12}$$

with additional specifications incorporating interactions between party affiliation and positive and negative partisanship. This ANCOVA framework provides a direct test of treatment effects while explicitly conditioning on baseline beliefs and provides a closely related robustness check to the LLR formulation used in the main results. Tables 17 and 18 report these estimates for Studies A and B, respectively.

K.1 Study A

As a robustness check, we re-estimate the Study A models using an ANCOVA specification in log-odds space, regressing posterior beliefs on prior beliefs and the same set of treatment indicators, partisanship measures, interactions, and controls reported in Table 3. These results are presented in Table 17.

Consistent with the LLR results in Table 3, prior beliefs strongly predict posterior beliefs across all specifications (Columns A1–A3: $\beta = 0.984$, 0.934, and 0.926, all $p < 0.01$), indicating substantial persistence in log-odds beliefs. This mirrors the negative coefficients on Prior in the LLR models (Columns 2–3: $\beta = -0.005$ and -0.006), confirming that posterior beliefs remain strongly anchored in initial beliefs, consistent with the information-theoretic interpretation that updating depends on the informational distance between the signal and prior beliefs.

As in the main analysis, the *Fox News* source indicator remains statistically insignificant throughout (Columns A1–A3: $\beta = -0.073$, -0.048, and -0.042), providing no evidence of a direct persuasive effect of source identity once priors are accounted for. Republican identification is positive and statistically significant in the additive specification (Column A2: $\beta = 0.159$, $p < 0.01$), closely paralleling the LLR results (Table 3, Column 2: $\beta = 0.182$, $p < 0.001$), but becomes statistically insignificant once inter-

Table 17: Robustness (ANCOVA log-odds): Study A

	Dependent variable:		
	$\log\left(\frac{\text{Posterior}_i}{100-\text{Posterior}_i}\right)$		
	(A1)	(A2)	(A3)
logit(Prior)	0.984^{***}	0.934^{***}	0.926^{***}
	(0.038)	(0.042)	(0.040)
Source = Fox News	−0.073	−0.048	−0.042
	(0.056)	(0.053)	(0.067)
Party = Republicans		0.159^{***}	−0.143
		(0.061)	(0.232)
Positive Partisanship		0.085^{**}	0.057
		(0.033)	(0.040)
Negative Partisanship		−0.047	-0.052^{*}
		(0.029)	(0.031)
Fox News × Republicans			−0.019
			(0.110)
Republicans × Positive Partisanship			0.058
			(0.062)
Republicans × Negative Partisanship			0.010
			(0.061)
Controls	✓	✓	✓
Observations	277	277	277
R^2	0.852	0.861	0.862
Note:			*p<0.1; **p<0.05; ***p<0.01

action terms are introduced (Column A3: $\beta = -0.143$), again matching the pattern observed in the LLR models.

Positive partisanship exhibits a modest positive association in the additive ANCOVA model (Column A2: $\beta = 0.085$, $p < 0.05$), comparable to its effect in the LLR regressions (Table 3, Column 2: $\beta = 0.062$, $p < 0.01$), but this effect attenuates and loses statistical significance in the full interaction specification. Negative partisanship becomes marginally negative in the interacted model (Column A3: $\beta = -0.052$, $p < 0.10$), closely corresponding to the LLR result (Table 3, Column 3: $\beta = -0.057$, $p < 0.10$). Importantly, none of the interaction terms—including *Fox News* × Republicans—reach conventional levels of statistical significance in either framework (ANCOVA Column A3: $\beta = -0.019$; LLR Column 3: $\beta = -0.027$).

Taken together, the ANCOVA results in Table 17 closely replicate the substantive conclusions of the LLR analysis in Table 3. Conditioning directly on prior beliefs yields the same qualitative pattern: posterior beliefs are overwhelmingly anchored in initial beliefs, with modest contributions from positive partisanship and little evidence that responsiveness to the signal varies systematically by source-partisan alignment. This confirms that the main findings are not artifacts of change-score modeling and are robust to alternative specifications that explicitly control for baseline beliefs.

K.2 Study B

Table 18: Robustness (ANCOVA log-odds): Study B

	Dependent variable:		
	$\log\left(\frac{\text{Posterior}_i}{100-\text{Posterior}_i}\right)$		
	(B1)	(B2)	(B3)
logit(Prior)	0.926***	0.886***	0.880***
	(0.039)	(0.042)	(0.043)
Source = Fox News	0.010	0.009	−0.092
	(0.094)	(0.092)	(0.124)
Party = Republicans		0.342***	0.204
		(0.104)	(0.389)
Positive Partisanship		−0.072*	−0.135*
		(0.041)	(0.074)
Negative Partisanship		0.070*	0.135**
		(0.038)	(0.055)
Fox News × Republicans			0.270
			(0.173)
Republicans × Positive Partisanship			0.136
			(0.083)
Republicans × Negative Partisanship			−0.155**
			(0.079)
Controls	✓	✓	✓
Observations	275	275	275
R^2	0.760	0.772	0.777
Note:			*p<0.1; **p<0.05; ***p<0.01

As a robustness check, we re-estimate the Study B models using an ANCOVA specification in log-odds space, regressing posterior beliefs on prior beliefs and the same set of treatment indicators, partisanship measures, interactions, and controls reported in Table 6. These results are presented in Table 18.

Consistent with the LLR results in Table 6, prior beliefs strongly predict posterior beliefs across all specifications (Columns B1–B3: $\beta = 0.926$, 0.886, and 0.880, all $p < 0.01$). This mirrors the negative and highly significant coefficients on Prior in the LLR models (Columns 2–3: $\beta = -0.009$ and -0.010, both $p < 0.001$), confirming that posterior beliefs remain strongly anchored in baseline beliefs, consistent with the interpretation that updating depends on the informational distance between the signal and prior beliefs.

The *Fox News* source indicator remains statistically insignificant in the additive specifications (Columns B1–B2: $\beta = 0.010$ and 0.009), closely paralleling the null effects in the corresponding LLR models (Table 6, Columns 1–2: $\beta = -0.006$ and -0.019). In the full interaction specification (Column B3), the direct Fox coefficient becomes negative but statistically insignificant ($\beta = -0.092$), similar in direction to the LLR result (Column 3: $\beta = -0.165$), indicating that any source effect operates primarily through interaction channels.

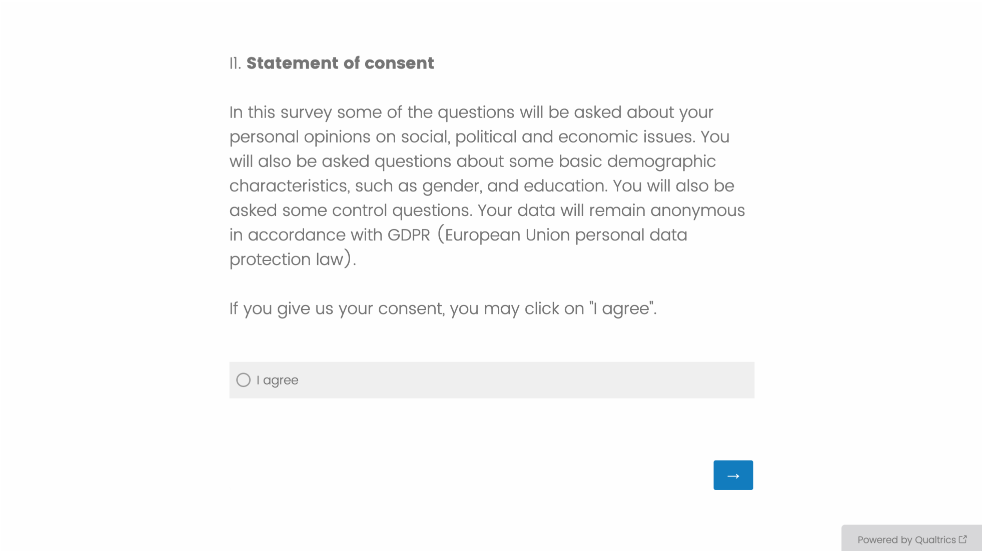
Republican identification remains positive and statistically significant in the additive ANCOVA model (Column B2: $\beta = 0.342$, $p < 0.01$), closely matching the LLR estimate (Table 6, Column 2: $\beta = 0.321$, $p < 0.001$). However, once interaction terms are introduced (Column B3), the main effect of Republican identification becomes statistically insignificant ($\beta = 0.204$), again replicating the pattern observed in the LLR framework (Column 3: $\beta = 0.159$).

Turning to affective partisanship, positive partisanship is negative and marginally significant in the ANCOVA specification (Column B2: $\beta = -0.072$, $p < 0.10$; Column B3: $\beta = -0.135$, $p < 0.10$), broadly consistent with its weak and statistically insignificant role in the LLR models. Negative partisanship remains positively associated with posterior beliefs (Column B2: $\beta = 0.070$, $p < 0.10$; Column B3: $\beta = 0.135$, $p < 0.05$), closely corresponding to the LLR estimates (Table 6, Columns 2–3: $\beta = 0.068$ and 0.108).

Importantly, the *Fox News* × Republican interaction, which is positive and statistically significant in the LLR specification (Table 6, Column 3: $\beta = 0.356$, $p < 0.01$), becomes positive but statistically insignificant in the ANCOVA model (Column B3: $\beta = 0.270$). At the same time, the interaction between Republican identification and negative partisanship becomes negative and statistically significant in the ANCOVA specification (Column B3: $\beta = -0.155$, $p < 0.05$), suggesting that once baseline beliefs are explicitly conditioned upon in log-odds space, some heterogeneity previously captured by the source-by-party interaction may instead be absorbed by affective polarization.

Overall, the ANCOVA results in Table 18 confirm the central role of prior beliefs in shaping posterior beliefs and broadly replicate the additive partisan patterns observed in Table 6. While the magnitude and statistical significance of the *Fox News* × Republican interaction attenuate under this alternative specification, the broader conclusion remains unchanged: posterior beliefs are strongly anchored in prior beliefs, and partisan heterogeneity plays a secondary but meaningful role in shaping responses to political signals.

L Some Additional Elements of Instructions


II. **Statement of consent**

In this survey some of the questions will be asked about your personal opinions on social, political and economic issues. You will also be asked questions about some basic demographic characteristics, such as gender, and education. You will also be asked some control questions. Your data will remain anonymous in accordance with GDPR (European Union personal data protection law).

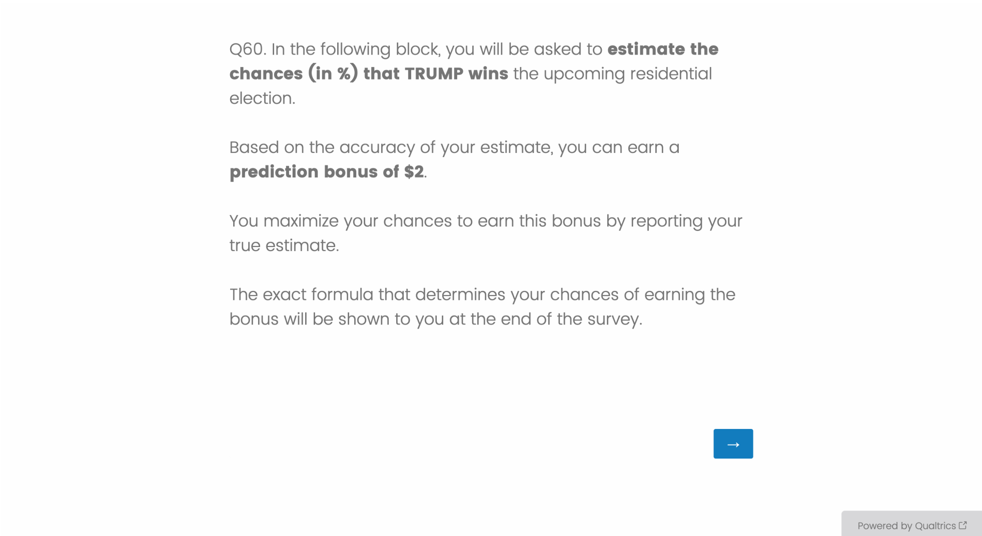
If you give us your consent, you may click on "I agree".

○ I agree

→

Powered by Qualtrics

Figure 20: Consent in instructions


Q60. In the following block, you will be asked to **estimate the chances (in %) that TRUMP wins** the upcoming residential election.

Based on the accuracy of your estimate, you can earn a **prediction bonus of $2**.

You maximize your chances to earn this bonus by reporting your true estimate.

The exact formula that determines your chances of earning the bonus will be shown to you at the end of the survey.

→

Powered by Qualtrics

Figure 21: Bonus payment introduction

Q61. In case you are interested, here is how your chances to earn the prediction bonus are calculated.

I1-1.
In case you are interested in seeing the formula that we will use to calculate your chances to receive the bonus, click on the link below.

Please note that the bonuses will only be calculated after the election outcome is known.

Tell me more

Figure 22: Short Explanation of the bonus payment construction

We apply the following procedure:

We take your last estimate of TRUMP winning the election. Once the election outcome is known, we calculate your prediction error. This is equal to how many percentage points your estimate was away from the election outcome. That is, if TRUMP wins, your prediction error is equal to 100% minus your estimate. On the other hand if HARRIS wins, your prediction error is equal to your estimate.

Then, we plug in the error in the following formula to obtain your chances to earn the bonus:

YOUR CHANCES=100%-(Prediction error)^2

Figure 23: More Detailed—in case the button "Tell me more" was clicked— explanation of the Bonus (first part).

Finally, we will randomly draw a number between 0 and 1 in excel. If this randomly generated number is smaller than YOUR CHANCES, you will receive the prediction bonus. Otherwise, if this randomly generated number is larger than YOUR CHANCES, you will not receive any bonus.

This procedure is called Binarized Scoring Rule and it is commonly used in economic experiments. It has the advantage that your chances to earn the prediction bonus are maximized if you report your true estimate.

Figure 24: More Detailed—in case the button "Tell me more" was clicked— explanation of the Bonus (second).