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.]

May, 2025

Abstract

Increasing political polarization in the United States has led individuals to process information in ways that reinforce their prior beliefs, often dismissing sources associated with opposing views or the other party. While partisan-motivated reasoning provides a well-established explanation for this phenomenon, we investigate and document a largely overlooked determinant: the role of surprise in belief updating. Using two experimental studies, we exposed participants to identical political information about the likely outcome of the 2024 presidential election, while varying its source, and measured belief updating via the Log-Likelihood Ratio. Results show that belief updating was primarily driven by the level of surprise, not source credibility. Counter-stereotypical signals—when a source provided unexpected information—elicited stronger updates, overriding partisan bias. Our findings suggest that surprise disrupts ideological echo chambers, challenging entrenched beliefs more effectively than source alignment alone.

Keywords: Partisanship, information provision, media, surprise, polarization, belief updating.
Word count: 9253.

*We would like to thank Dani Sandu and Honorata Mazepus 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; Economics Department, University of Amsterdam; 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, it is not surprising that people tend to engage primarily with information sources that align with their views while distrusting or ignoring media associated with the opposing side. Shaped by both psychological and social factors, these tendencies are reinforced by media, as news outlets compete for attention by emphasizing narratives that appeal to their audience's ideological preferences.[1] As a result, individuals not only choose media sources that align with their views but also develop greater skepticism toward alternative sources, often ending up in ideological information bubbles.

While this phenomenon aligns with a broad literature on partisan-motivated reasoning, which suggests that individuals selectively process information in ways that reinforce their prior beliefs and ideological affiliations (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; Peterson and Iyengar, 2021; Guay and Johnston, 2022), it raises a critical question: Is there a mechanism that can facilitate individuals' incorporation of new information, overriding psychological (e.g., partisan) biases in a highly polarized environment?

Existing research has primarily focused on factors such as source credibility, partisan alignment, and ideological congruence in shaping belief updating. However, 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: 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.

On the psychological and communication side, there is growing evidence that sur-

[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. Socially, political identity functions as a form of group affiliation, and consuming media from the "other side" can signal a lack of loyalty, sometimes leading to social discomfort or even exclusion.

prise plays a meaningful role in how individuals process and respond to information. For instance, inoculation theory posits that exposing individuals to weakened, counter-attitudinal information can increase their resistance to persuasion and misinformation. Crucially, this exposure works by surprising individuals with arguments they had not previously considered, prompting greater cognitive engagement and resilience (e.g., Compton et al., 2021; Roozenbeek and Van der Linden, 2024).[2] 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 unexpected information.

In this paper, we investigate how the level of surprise in political information (relative to prior beliefs) shapes belief updating across different media sources. Specifically, we examine whether high-surprise information can disrupt partisan echo chambers and challenge entrenched beliefs, or whether it is more likely to be dismissed as unreliable. By exploring this underexamined mechanism, we aim to shed light on the potential for surprising political information to act as 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 Presidential Elections; then were exposed to a study-specific signal (*pro-Trump* signal in study A and *anti-Trump* signal in study B) with exogenously varied sources (*The New York Times* or *Fox News*[3]); and subsequently reported their posterior beliefs. This allowed us to measure belief updating using tools from information theory and assess the role of surprise in shaping belief updating across partisan and media contexts. Both studies were conducted on the same day, November 1, 2024, four days before the US Presidential Elections.

Our main finding is that belief updating is primarily driven by the element of surprise in the provided information rather than other, more traditional factors. Specifically, the impact of the source of the signal are moderated by the level of surprise, rather than by source credibility alone. Contrary to classical source credibility or partisanship (motivated) reasoning effects, belief updating was not primarily driven by whether a signal came from *The New York Times* or *Fox News*. Instead, the surprise shaped responses to the

[2]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 or caricature opposition views, making them seem illogical, extreme, or out of touch (e.g., Guriev and Treisman, 2022; Bronnikov, 2025).

[3]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.

source. In Study A, we observe the strongest belief update among Republicans who read the pro-Trump signal coming from *The New York Times*. In Study B, the strongest belief update was observed among Democrats when they read the anti-Trump signal coming from *Fox News*. These results highlight that the credibility of the source mattered less than the extent to which the source was surprising in delivering a counter-stereotypical signal.

With regard to partisan identity we find mixed results. While positive partisanship (strong attachment to one's party) significantly correlated with belief updating in Study A, the effect disappeared in Study B. This suggests that belief updating is not just a function of surprise but also how partisans rationalize surprising information based on their ideological identity and the current political situation. We also observe other asymmetries. Specifically, we find that Republicans seem to be more resistant to anti-Trump information than Democrats are to pro-Trump signals, which could have been due to asymmetries in election expectations at the time.

Our results contribute to several strands of research. 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; Bronnikov, 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 in political science and economics, 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 paper builds on this emerging perspective by introducing an information-theoretic framework to belief updating, and testing the role of surprise in a domain traditionally dominated by psychological or sociological explanations. This perspective challenges the standard assumption that partisans will reflexively discount out-group sources and instead posits that the unexpectedness of a signal—regardless of source—can shape belief updating. More specifically, we show that the effect of the source is moderated by the level of surprise rather than determined solely by source credibility or partisan alignment. In both of our studies, participants updated their beliefs more strongly when the signal was surprising relative to their priors, even if the signal came from an out-group source. Our results suggest that partisanship does not categorically block belief updating; rather, surprise can act as a catalyst for belief updating, contingent on the interaction between prior beliefs and the perceived incongruity of the signal and its source.

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). Our results suggest a revision of how we understand selective exposure by showing that people can update their beliefs when exposed to surprising counter-stereotypical signals, even from out-group sources. 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). These patterns are thought to reinforce ideological echo chambers, making belief updating across partisan lines increasingly rare. However, our findings suggest a more nuanced picture that reframes the "echo chamber" idea (e.g., Sunstein, 2001; Levy and Razin, 2019; Cinelli et al., 2021): under the right conditions, ideological media can challenge rather than reinforce prior beliefs. Namely, we show that participants were more responsive to political signals when those signals were incongruent with the perceived ideological stance of the source. This suggests that individuals are not only selectively exposed to ideologically congruent sources, but also can interpret and integrate information from ideologically incongruent sources in meaningful ways when it violates expectations. In this way, surprise serves as a key moderator of how and when cross-cutting information is processed, offering a revision to the standard selective exposure framework. Rather than functioning as impermeable silos, ideological media ecosystems may occasionally act as conduits for belief updating—if and when the content they carry is sufficiently surprising.

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). We demonstrate that belief updating is not entirely frozen in polarized environments. 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.[4] The effect of surprise takes place across party lines, suggesting that polarization does not eliminate the capacity for belief updating, but instead shapes the conditions under which it occurs. Our results indicate that surprise can temporarily disrupt the otherwise stable dynamics of polarized information processing by increasing the perceived informativeness of a signal, even if it comes from a distrusted source. Thus, rather than viewing polarization as a complete barrier to persuasion or belief updating, we highlight how surprising, expectation-violating content—particularly when embedded in ideologically incongruent sources—can momentarily open cracks in partisan echo

[4]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).

chambers.

2 Theoretical Framework and Hypotheses

2.1 Belief updating

We will measure belief updating within the standard framework of 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). We start with the binary state space, where each state corresponds to a winner of the upcoming presidential elections, 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, which we elicit in the experiment.

After elicitation of prior beliefs, participants read an opinion of one of the two political experts. 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 (Harris will win). In both studies, we vary the source of the signal as quoted by either *Fox News* or *The New York Times* (treatments).

Within information theory this can be formalized as participants receiving a signal. Signal S is a piece of relevant information shown to the participant. The two main characteristics of a signal are the *opinion* of the expert (whether Trump will win or lose) and the *identity* of the source that reproduces this opinion (*Fox News* or *The New York Times*). In this sense, S can take one of the four potential values:

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

Figure 1: Summary of Signals

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 subjects 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 subjects incorporate information into their beliefs, as further discussed later in this section. For a more elaborate discussion of this interpretation of the Bayes formula, we refer to Rabin (2013) and Lois et al. (2023).

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 for Harris, and subsequently taking logarithms. The LLR is inferred from the log-prior and log-posterior odds, 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 in their belief upon receiving the signal. Positive LLR implies that the participant has interpreted the received signal S as evidence in favor of Trump winning, whereas negative LLR implies that the participant has interpreted S as evidence of Harris winning. And of course, LLR being equal to 0 implies that the participant has not taken the signal into account, and has not updated the prior belief. Furthermore, the absolute value of LLR reflects how strongly the evidence is perceived by the participant. That is, the further away from zero LLR is (in one or the other direction), the stronger the effect of the evidence is on the participant, and therefore the more the participant has updated.

Note that using LLR, rather than the absolute difference between posterior and prior beliefs, allows us to control for the role of prior beliefs. For instance, the LLR that we obtain when beliefs are updated from 80% to 90% is much larger than the LLR we obtain when beliefs are updated from 50% to 60% (see more details in Appendix A). Although

in both cases the absolute belief updating is the same in percentage points, in the former case the participant has incorporated much more information than in the latter case. In other words, the participant has interpreted the signal as much stronger evidence in the former than in the latter case. This is consistent with information theory, in the sense that LLR reflects how *surprising* the participants find the evidence in relation to their prior beliefs.

Note 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 science experiments where it is unclear how each participant interprets the likelihood of a signal.

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

For the purpose of this study, we construct five hypotheses—three associative and two causal—all of which were pre-registered before the experiment.[5]

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—the underlying mechanism of information incorporation—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-aligned narratives rather than objective truth (e.g., Kahan, 2013; Petersen et al., 2013; Bolsen et al., 2014; Donovan et al., 2020; Guay and Johnston, 2022).

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

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.*

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 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).

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. 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.

The causal effect of the source on belief updating is critical for understanding the role of media or sources from which the signal is coming. 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 mediating belief change. Recent work in behavioral economics supports this perspective, showing that unexpected information can drive belief updating even when it originates from ideologically incongruent sources (e.g., Bronnikov and Drouvelis, 2025). 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 expectations, 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, 2025) suggest that belief updating is driven by the degree to which new information is surprising relative to prior expectations. From this perspective, even signals from out-group or distrusted 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 based on a *favorable* signal (expert opinion that Trump wins). 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 $N_A = 268$, limited to US nationals and excluding participants from Study 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 elections. 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 B).

Following this, participants were shown a signal that presented the opinion of Nate Silver, a polling analyst and public figure, who predicted that Donald Trump would win the election (upcoming at the time of the experiment). Participants were told that this information comes from one of the two different sources: *The New York Times* or *Fox News*. Table 1 shows the precise wording (see Figures 4 and 5 in Appendix B for the operationalization in 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*. While the content of the signal remained identical, the source varied across conditions.

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 were rewarded monetarily 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 elections.[6]

Additional measures and controls. We gathered additional variables to gain a more complete picture of the participants' traits and how they might affect belief updating. These additional variables include indices for positive and negative partisanship, measures of cognitive reflection ability, and sociodemographic details. All of these measures are commonly used in previous, related research (a full description of each is provided in Appendix F).

[6]Following a novel 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 subjects 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.*

Note: The signal was accompanied by a brief introduction of Nate Silver: *Nate Silver is a well-known American statistician, writer and founder of FiveThirtyEight (famous website that focuses on opinion polls)*.

3.2 Main Results

Our analysis strategy is aligned with the pre-registration.[7] Looking at first three correlational hypotheses we find only one significant correlation with LLR that we describe in the result.

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 party 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). There is 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, standalone effect on belief updating after accounting for other factors such as party affiliation and prior beliefs. Under signals predicting Trump's high chances of winning, Democrats exhibit no significant belief updating regardless of the source. This limited response reflects the lack of surprise in the signal content, as it contradicts their partisan priors (that Trump is unlikely to win) and provides little perceived informational novelty, regardless of the source. For Republicans, belief updating is stronger, particularly under signals from The New York Times, compared to Fox News. This suggests moderate surprise in receiving a signal that aligns with their par-*

[7] See the pre-registration document on the AsPredicted platform (#196961).

tisan priors (that Trump is likely to win) but comes from a source perceived as ideologically opposed. The greater informational content of The New York Times signal likely arises from the counter-stereotypical nature of the source. Conversely, Fox News, as a trusted in-group source, delivers a signal that Republicans expect, resulting in less surprise and a weaker belief update.

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 suggests that the signal contained a moderate level of surprise for Democrats. The source, *The New York Times*, aligns with their general trust in the source but conflicts with their partisan priors that Trump is unlikely to win. The positive LLR reflects that Democrats found the signal somewhat informative, 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—-it was unexpected, but not enough to substantially shift their prior beliefs.

When the signal was attributed to *Fox News*, Democrats exhibited an even smaller belief update, with an LLR of 0.018 (SE=0.031, 95% CI: [–0.043, 0.079]). Here, the negligible belief update suggests that the signal from Fox News carried minimal surprise for Democrats. 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 significantly from Democrats' prior expectations (both about the source and the signal content). As a result, the LLR indicates that the signal had almost no effect on their posterior beliefs, as it provided little to no new or 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 reflects high surprise in information theory 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 unexpected nature of the signal's alignment with their prior beliefs (Trump's high chances) and its delivery by a generally out-group source significantly increased the signal's information content. As a result, Republicans viewed the signal as highly credible and incorporated it more strongly into their posterior beliefs. The higher LLR highlights that the signal reduced their uncertainty by providing a surprisingly aligned piece of evidence.

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 contained lower surprise for Republicans. Since Fox News is a source they expect to align with their priors (that Trump has a high chance of winning), the signal did not significantly deviate from their expectations. Consequently, the information content of the signal was limited, as it served more as a reinforcement of existing beliefs rather than a surprising new 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 updated their beliefs under the pro-Trump signal (Republicans were more receptive to the signal, while Democrats resisted it). When the source is Fox News, the result is not statistically significant (see full details in Appendix C). The comparison over sources does not show any significant results (see Appendix D for details).

Lastly, we state 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.*

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

	(1) LLR	(2) LLR	(3) LLR
Prior		–0.005	–0.006*
		(0.003)	(0.003)
Source = Fox News	–0.059	–0.047	–0.034
	(0.064)	(0.063)	(0.080)
Party = Republicans	0.075	0.183**	–0.143
	(0.056)	(0.068)	(0.235)
Positive Partisanship	0.060	0.063*	0.056
	(0.032)	(0.032)	(0.040)
Negative Partisanship	–0.023	–0.025	–0.057
	(0.025)	(0.025)	(0.030)
Fox News × Republicans			–0.025
			(0.107)
Republicans × Positive Partisanship			0.007
			(0.057)
Republicans × Negative Partisanship			0.075
			(0.049)
Source Classification	0.002	0.005	0.005
	(0.038)	(0.038)	(0.038)
Age	–0.002	–0.002	–0.002
	(0.002)	(0.002)	(0.002)
Education	0.019	0.014	0.015
	(0.017)	(0.016)	(0.016)
Employment	–0.012	–0.011	–0.011
	(0.015)	(0.015)	(0.015)
Gender	–0.047	–0.060	–0.069
	(0.054)	(0.056)	(0.058)
Ethnicity	0.030	0.038	0.037
	(0.026)	(0.027)	(0.027)
Place of Birth	–0.069	–0.069	–0.075
	(0.039)	(0.037)	(0.039)
Student	0.080	0.084	0.082
	(0.072)	(0.071)	(0.072)
Constant	–0.083	0.168	0.368
	(0.185)	(0.223)	(0.227)
R^2	0.067	0.097	0.106
N	268	268	268

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 G.

To explore interaction effects, we estimate three OLS regression models (reported in Table 3). In the first model, we focus solely on the direct effects on the LLR, omitting prior beliefs. The second model introduces the prior belief variable. Finally, the third, most comprehensive model includes both the prior belief and interaction effects along with the direct effects on the LLR. In all specifications in Table 3, controls (e.g., age, education, gender, etc.) are included. Specifications without controls (that can be found in Table 17 in Appendix H) demonstrate that their exclusion does not alter the key patterns.

In the baseline model (Column 1 in Table 3), the analysis excludes both the prior belief measure and any interaction terms. As a result, the regression captures only the direct effects of the signal's source (e.g., whether the signal came from Fox News), partisan identity (whether a participant identifies as Republican), partisanship indices (positive and negative), and several demographic controls. For instance, the coefficient on "Party = Republicans" is positive (0.075), suggesting that, in this direct specification, Republican identifiers tend to update their beliefs differently compared to the baseline group. However, without the prior belief measure, this model cannot account for the key element of surprise—that is, the discrepancy between what participants originally believed and the information they received. The low R^2 (0.067) indicates that omitting the prior leaves a substantial amount of variance unexplained, implying that the surprise component is a critical driver of belief updating.

In the second model (Column 2 in Table 3), the prior belief variable is introduced while still excluding interaction effects. The coefficient on Prior is -0.005 (with a standard error of 0.003), which—although modest—suggests that higher prior beliefs (i.e., beliefs more aligned with the new signal) are associated with smaller updates in LLR. In other words, when participants' initial beliefs are closer to the signal (and thus less surprising), the magnitude of their belief updating is reduced. Moreover, after controlling for the prior, the effect of "Party = Republicans" increases to 0.183 and becomes statistically significant, indicating that partisan identity has a stronger role in belief updating when the initial level of surprise is taken into account. The improvement in R^2 (from 0.067 to 0.097) confirms that including prior beliefs—thereby capturing the surprise element—enhances the model's explanatory power.

The full specification (Column 3 in Table 3) incorporates both the prior belief variable and interaction terms that allow the effects of the signal source to vary by partisan identity and partisanship. Here, the coefficient on Prior remains negative (-0.006, significant at the 5% level), reinforcing the notion that a higher prior (i.e., less surprising information) leads to a smaller update. In this model, the direct effect of receiving the

signal from Fox News becomes less negative (-0.034) compared to the baseline, while the main effect of party identification (i.e., "Party = Republicans") flips to -0.143—likely reflecting that the interaction terms are now capturing much of the nuanced variation. Although the interaction terms (e.g., "Fox News × Republicans," "Republicans × Positive Partisanship," and "Republicans × Negative Partisanship") are not robustly significant, they hint at the possibility that the impact of the source and the strength of partisan feelings might moderate how surprising the signal is perceived to be. The overall R^2 rises to 0.106, suggesting that even though the additional interaction terms add modest incremental explanatory power, the central driver of belief updating remains the degree of surprise as measured by the prior.[8]

4 Study B: "Trump Loses"

This study examines how individuals update their beliefs about Trump's likelihood of winning when presented with information *unfavorable* to the presidential candidate, with participants 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 $N_B = 260$ was restricted to individuals with US nationality and those who had not participated in Study A.

4.1 Experimental Design

The design of the experiment in Study B is identical to experiment in Study A except for the signals the participants received. In Study B, we provided them with a signal representing the opinion of a researcher and public figure, Allan Lichtman. His opinion, that Trump would lose the elections, was reported by both *The New York Times* and *Fox News*, see Table 4.

4.2 Main Results

We proceed first with correlational hypotheses as in study A.

Result B1. *We find no significant correlation between LLR and party affiliation; LLR and*

[8]A noteworthy aspect of the analysis is the source classification, where subjects were asked to classify the source of the signal (i.e., either *The New York Times* or *Fox News*) as (1) Democrat, (2) Neutral, (3) Republican source, or (4) I do not know. Comparing these classifications across party lines, we find significant differences in how Democrats and Republicans perceive *The New York Times* and *Fox News* (see Appendix E for details). Building on this observation, we incorporate the source classification variable into our regression analysis. However, we do not find significant effects of source classification on the LLR.

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.*

Note: The signal was accompanied by a brief introduction of Nate Silver: *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.*

the source; LLR and negative or positive partisanship.

We find no significant correlations between belief updating and standard variables, thus providing no support for hypotheses H(1) to H(3).

Next, we analyze hypotheses H(4a) and H(4b) showcasing the primary result of study B.

Result B2. *On average, the source of the signal did not have a direct, standalone effect on belief updating after accounting for other factors such as party affiliation and prior beliefs. Updating under signals from both The New York Times and Fox News significantly, Democrats experience greater belief updating (stronger LLRs) from Fox News due to the unexpected nature of the source delivering a pro-Democrat signal, which increases surprise and informational content. Significantly updating under the signal from The New York Times only, Republicans, on the other hand, exhibit more belief updating from the source that does not align with their expectations for partisan bias, while Fox News elicits minimal surprise and thus limited belief updating.*

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).

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 both credible and informative. This

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** (0.065)	**–0.259** (0.085)
Fox News treatment	**–0.423** (0.096)	–0.093 (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.

signal aligns with their prior beliefs (that Trump is unlikely to win), and its delivery by a trusted source adds to its perceived validity. The moderate level of surprise stems from the fact that while the signal is consistent with their expectations about Trump's chances, it still provides new information that helps reduce uncertainty. 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 suggests that Democrats found the signal about Trump's low chances of winning from *Fox News* to be highly surprising. 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) likely made it highly informative and credible, leading to a significant belief update. The greater magnitude of the LLR compared to *The New York Times* suggests that the surprise effect increased the information content of the signal, amplifying its impact on 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 likely carried moderate 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 narratives. This balance of prior expectations and source credibility meant the signal provided informational content sufficient

to reduce uncertainty and drive 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 found the signal from *Fox News* to have low surprise and thus low information content. While the signal contradicted Republicans' prior beliefs about Trump's chances, it came from *Fox News*—a source they generally trust and perceive as aligned with their worldview. Because of this alignment, the signal may have been interpreted as less surprising or as an outlier, reducing its perceived informativeness and resulting in minimal belief updating. From an information theory perspective, the low surprise factor of the signal limited its ability to reduce uncertainty or provide significantly new information.

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 updated their beliefs under the anti-Trump signal (Republicans were more surprised by the signal, whereas Democrats were likely unsurprised, see full details in Appendix C). 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 D for details).

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

Result B3. *Belief updating is notably impacted by the interaction between the signal source and party affiliation, with Republicans showing a stronger reaction when the information comes from Fox News. However, the influence of party affiliation on partisanship measures (whether positive or negative) does not significantly affect belief updating.*

To investigate this, we run OLS regressions similar to those in Study A. The results are reported in Table 6. Models without controls (presented in Table 18 in Appendix H) show that their omission does not affect the main patterns.

In this baseline model (Column 1 in Table 6), we exclude the prior belief measure as well as any interaction effects. As a result, the model solely captures the direct associations between the signal's source (e.g., *Fox News*), party identification (Republicans), partisanship measures, and other controls such as age, education, and demographics. For example, the coefficient on party identification ("Party = Republicans") is positive (0.194), suggesting that Republican identifiers tend to update their beliefs in one direction relative to the omitted category. However, neither the effect of receiving the signal from *Fox News* (-0.051) nor that of party identity is statistically significant in isolation. Importantly, because this specification does not control for the prior beliefs, it cannot

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

	(1) LLR	(2) LLR	(3) LLR
Prior		–0.009**	–0.009**
		(0.003)	(0.003)
Source = Fox News	–0.051	–0.054	–0.185
	(0.102)	(0.100)	(0.126)
Party = Republicans	0.194	0.336**	0.178
	(0.100)	(0.105)	(0.368)
Positive Partisanship	–0.044	–0.024	–0.062
	(0.034)	(0.032)	(0.044)
Negative Partisanship	0.084*	0.069	0.108*
	(0.039)	(0.036)	(0.051)
Fox News × Republicans			0.348*
			(0.153)
Republicans × Positive Partisanship			0.096
			(0.061)
Republicans × Negative Partisanship			–0.116
			(0.075)
Source Classification	0.052	0.041	0.027
	(0.053)	(0.052)	(0.051)
Age	0.007*	0.005	0.006*
	(0.003)	(0.003)	(0.003)
Education	0.043	0.039	0.032
	(0.027)	(0.027)	(0.027)
Employment	–0.011	–0.004	–0.008
	(0.025)	(0.025)	(0.025)
Gender	–0.051	–0.044	–0.036
	(0.074)	(0.072)	(0.071)
Ethnicity	–0.051	–0.060	–0.062
	(0.046)	(0.045)	(0.045)
Place of Birth	0.108*	0.117**	0.109*
	(0.042)	(0.043)	(0.042)
Student	0.036	0.017	0.033
	(0.054)	(0.054)	(0.055)
Constant	–1.037**	–0.523	–0.415
	(0.341)	(0.337)	(0.393)
R^2	0.081	0.118	0.144
N	260	260	260

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 G

capture the surprise element inherent in the belief-updating process. As a result, its explanatory power is relatively low ($R^2 = 0.081$), indicating that omitting the prior likely leaves out a crucial determinant of the magnitude of the update.

In the second model (Column 2 in Table 6), the prior belief variable is included. Here, the coefficient on Prior is -0.009 and is statistically significant at the 1% level. This negative coefficient implies that a higher prior (i.e., beliefs that are closer to the signal) is associated with a smaller LLR update. In other words, when the new signal is less surprising because it aligns more closely with what participants initially believed, they update their beliefs by a smaller amount. Additionally, the effect of "Party = Republicans" increases to 0.336 and becomes statistically significant, suggesting that partisan identity matters for belief updating when controlling for one's prior. Overall, the inclusion of the prior enhances the model's fit (R^2 increases to 0.118) and directly supports the idea from information theory: the greater the discrepancy (or surprise) between what was expected and what is learned, the larger the belief update.

The full specification (Column 3 in Table 6) incorporates both the prior belief measure and interaction terms that capture how the effect of the signal source interacts with partisan identity and partisanship. The coefficient on Prior remains -0.009 (and significant), reinforcing that the extent of surprise—the gap between prior and new information—drives the LLR update. Notably, the direct effect of receiving the signal from *Fox News* becomes more negative (-0.185), though it is not significant by itself. However, the interaction term "Fox News × Republicans" is positive (0.348, significant at the 5% level), indicating that for Republican identifiers, the impact of a signal delivered by *Fox News* is amplified relative to other sources. This suggests that when partisan alignment and source credibility coincide, the perceived surprise (or lack thereof) of the information is modulated by these contextual factors. The additional interaction terms with positive and negative partisanship, while not statistically significant, point to further nuances in how emotional and evaluative dimensions of partisanship might condition the belief update. With an R^2 of 0.144, this specification explains more of the variance in belief updating, capturing both the direct surprise effect and the way this effect is moderated by source and partisan factors.[9]

5 Conclusion

This paper explores whether surprising political information—measured as deviation from prior beliefs—can prompt belief updating and potentially disrupt partisan echo

[9]As in Study A, driven by the same reason—significantly different perceptions of both sources by both Democrats and Republicans—here we also include the source classification variable in regressions. However, our analysis does not reveal significant effects of source classification on the LLR.

chambers in polarized environments.

Our findings underscore the central role of surprise in shaping belief updating across partisan lines. In our experiment, rather than reacting to signals based solely on whether they came from an ideologically aligned or opposing media source, participants responded most strongly when the signal itself defied their expectations. Republicans, for instance, were more likely to revise their beliefs when a pro-Trump signal came from an unexpected outlet like *The New York Times*, rather than a predictable source such as *Fox News*. Conversely, Democrats showed the greatest belief updating when anti-Trump information was delivered by *Fox News*, not because the signal was inherently more credible, but because it was more surprising. These patterns suggest that the power of political information lies not just in who says it, but in how much the signal violates prior assumptions—making surprise a key lever for belief updating in a polarized media environment.

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. 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 a testable mechanism—surprise—that can potentially disrupt it.

From a policy perspective, our findings open up new possibilities: rather than attempting to reduce polarization by merely correcting misinformation—which often proves ineffective or even backfires—strategies that leverage expectation violations may prove more fruitful. Designing interventions that embed counter-attitudinal signals in unexpected or counter-stereotypical sources, for instance, may 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 serve as an entry point into otherwise impermeable ideological bubbles, its power unlikely to be inexhaustible. Put differently, the effectiveness of surprise likely depends on its 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; over time, and with repeated exposure, even signals that initially violate expectations may lose their disruptive power, becoming assimilated into the background of what individuals come to anticipate.

With this in mind, the "budget" for surprise is likely to be finite—especially in polar-

ized political contexts, where skepticism toward opposing sources is already high. Attempts to sustain belief updating by escalating the magnitude of surprise (e.g., through sensitive or emotionally charged content) may also carry risks. While more extreme signals might generate temporary attention or disruption, they may also trigger backlash and further polarize audiences. Thus, although surprise can be a powerful catalyst for belief updating, it is best understood as an occasional rather than continuous mechanism—capable of opening brief windows for reflection, but unlikely to serve as a unique long-term solution. Effective use of surprise must therefore be strategic, calibrated, and cognizant of its (likely diminishing) returns and potential side effects.

References

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.

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, "(Politically) Motivated Information Acquisition: A Minimal Formal Framework," *Working paper*, 2025.

_ **and Michalis Drouvelis**, "The Role of Emotions in Asymmetric Belief Updating," *Working paper*, 2025.

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.

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.

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.

Cinelli, Matteo, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini, "The echo chamber effect on social media," *Proceedings of the National Academy of Sciences*, 2021, *118* (9), e2023301118.

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.

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.

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, 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.

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.

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.

Levy, Gilat and Ronny Razin, "Echo chambers and their effects on economic and political outcomes," *Annual Review of Economics*, 2019, *11* (1), 303–328.

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 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.

_ , _ , 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 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.

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.

Sunstein, Cass R., *Republic.com*, Princeton University Press, 2001.

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.

A Details on the Log-likelihood Ratio

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

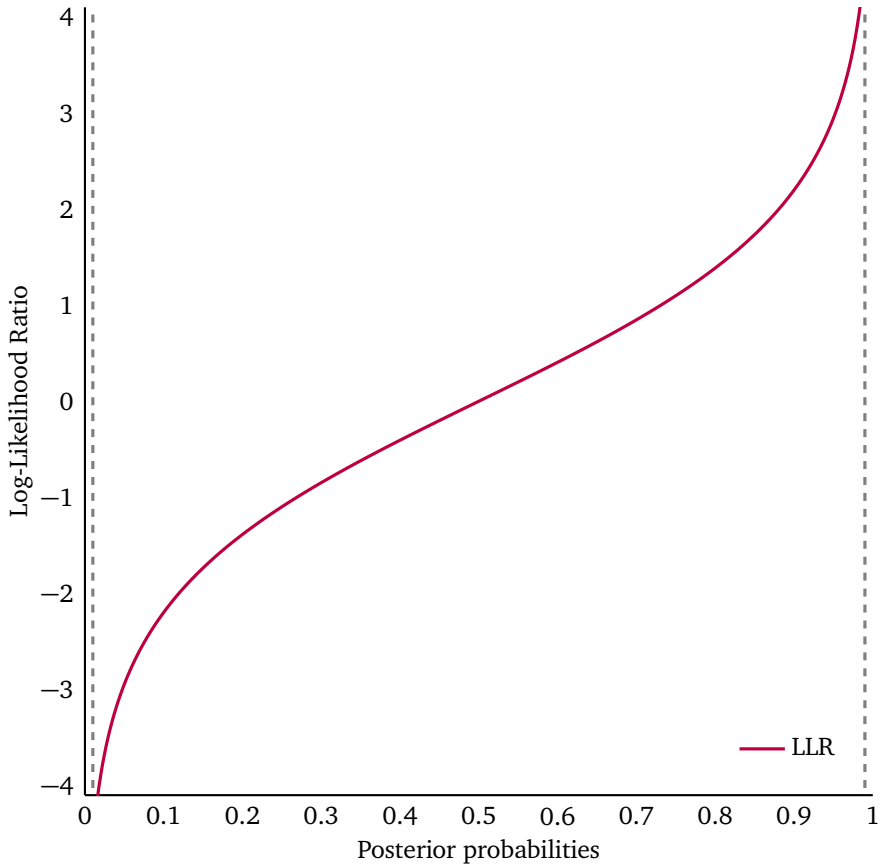

Figure 2: The Log-Likelihood Ratio for a binary Sample Space.

Notes: Fixing priors, the Log-Likelihood Ratio is given by log(posterior/(1 − posterior)). The graph is undefined at $\theta = 0$ and $\theta = 1$ 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.

In Section 2, we provide the following example, emphasizing the non-linear nature of the LLR function: when beliefs are updated from 80% to 90% is much larger than the LLR we obtain 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 probabilities: it is defined as the log of the odds (i.e., log(posterior/(1−posterior))). 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 from 50% to 60%

$$\text{LLR}_{50\to60} = \log\left(\frac{0.60}{1-0.60}\right) - \log\left(\frac{0.50}{1-0.50}\right) \tag{3}$$

$$= \log(1.5) - \log(1) = \log(1.5) \approx 0.405 \tag{4}$$

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

$$\text{LLR}_{80\to90} = \log\left(\frac{0.90}{1-0.90}\right) - \log\left(\frac{0.80}{1-0.80}\right) \tag{5}$$

$$= \log(9) - \log(4) \approx 0.811 \tag{6}$$

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 Elements of Instructions: Priors/Posteriors and signals

B.1 How Priors/Posteriors Were Introduced

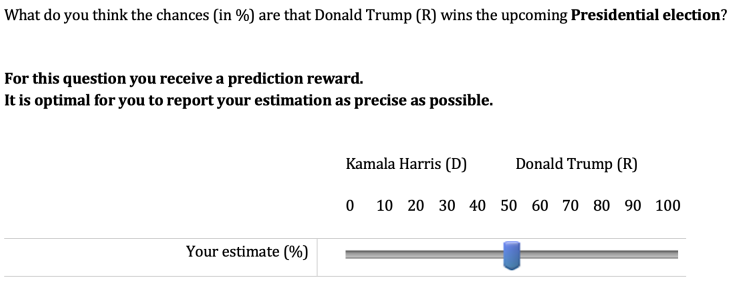

Figure 3: Beliefs Elicitation Screenshot.

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

B.2 Study A: How signal (Signal) Was Introduced

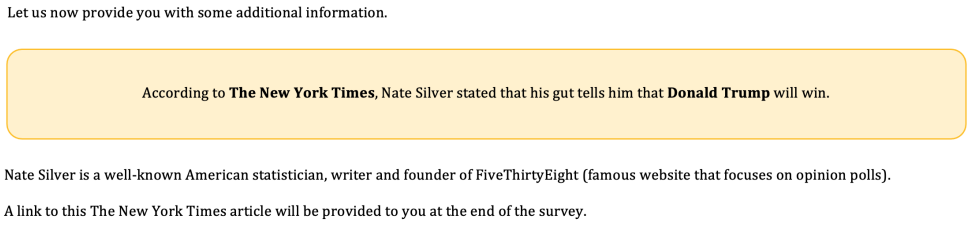

Figure 4: signal communicating Trump's High chances of winning reported by the New York Times.

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

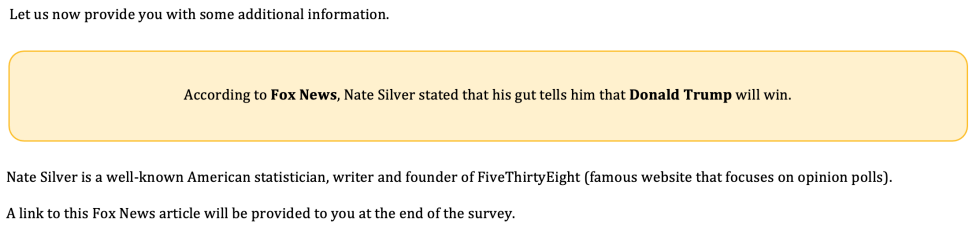

Figure 5: signal communicating Trump's High chances of winning reported by *Fox News*.

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

B.3 Study B: How signal (Signal) Was Introduced

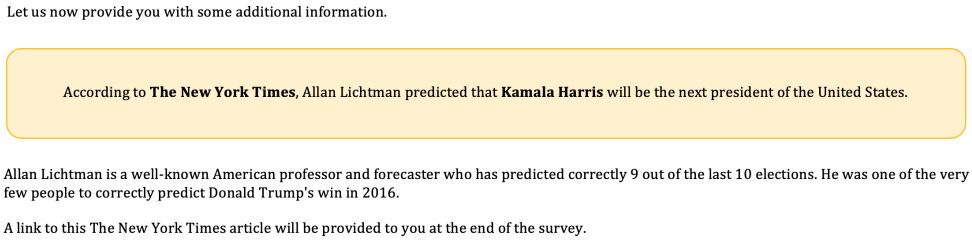

Figure 6: signal communicating Trump's Low chances of winning reported by the New York Times.

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

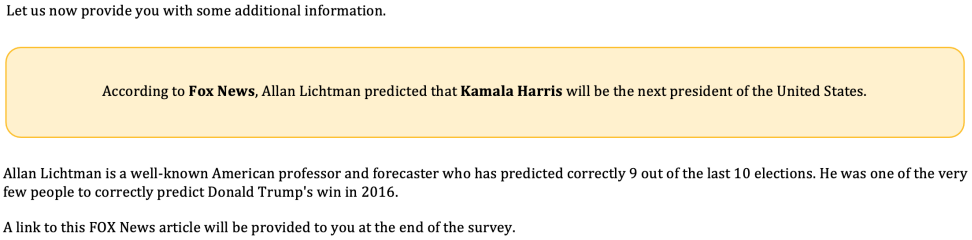

Figure 7: signal communicating Trump's Low chances of winning reported by *Fox News*.

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

C Comparison Over Party

C.1 Study A

We use the two-sample Wilcoxon rank-sum (Mann-Whitney) test compare the distributions of the log-likelihood ratio between Democrats and Republicans under two distinct conditions in Study A. These conditions are based on the source of the information: either *The New York Times* or *Fox News*.

We start with testing the first condition, when the source is *The New York Times* (see Table 7 in Appendix C). The p-value (0.031) is statistically significant at the $\alpha = 0.05$ threshold, allowing us to reject the null hypothesis. Republicans have a higher rank sum (4812.5) than Democrats (4098.5), indicating that Republicans updated their beliefs more positively (or less negatively) compared to Democrats in response to the signal from *The New York Times*. This result reflects partisan differences, namely: (i) Democrats, for whom the signal conflicts with their prior beliefs, likely downplayed or resisted the signal even though it came from a trusted source; and (ii) Republicans, for whom the signal aligns with their priors, were more likely to incorporate the information into their belief updating. Overall, there is a statistically significant difference in how Democrats and Republicans updated their beliefs when the pro-Trump signal was attributed to *The*

New York Times. Republicans were more receptive to the signal, while Democrats resisted it.

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 $>$ $\|z\|$): 0.0307			
Exact p-value: 0.0305			

Next, we turn to the second condition, that is when the source is *Fox News* (see Table 8 in Appendix C). The p-value (0.055) is marginally insignificant at the $\alpha = 0.05$ threshold but suggests a possible trend of differences between the two groups. Republicans have a higher rank sum (4912) than Democrats (4268), similar to Condition 1. This indicates that Republicans again updated their beliefs more positively (or less negatively) than Democrats when the pro-Trump signal was attributed to *Fox News*. This suggests that (i) Democrats may have discounted the signal even further when it came from *Fox News*, a source they likely associate with pro-Republican bias, and (ii) Republicans, on the other hand, may have been more receptive to the signal because it aligns both with their priors and with the perceived ideological stance of *Fox News*. Overall, while the result is not statistically significant, the trend suggests that Republicans were more likely to update their beliefs in response to the signal from *Fox News*, while Democrats were more resistant.

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			

C.2 Study B

We use the two-sample Wilcoxon rank-sum (Mann-Whitney) test compare the distributions of the log-likelihood ratio between Democrats and Republicans under two distinct conditions in Study B. These conditions are based on the source of the information: either *The New York Times* or *Fox News*.

To begin, we test the first condition, when the source is *The New York Times* (see Table 9 in Appendix C). The p-value (0.069) suggests that the result is marginally insignificant at the conventional $\alpha = 0.05$ threshold. However, it is close enough to suggest that there might be some meaningful differences between the two groups. The higher rank sum for Republicans indicates that Republicans tended to have higher LLR values than Democrats when the signal predicting Trump's loss came from *The New York Times*. This suggests that Republicans updated their beliefs more positively (less negatively) in response to the signal from this source, relative to Democrats. Overall, while the result is not statistically significant at the $\alpha = 0.05$ level, the trend indicates that Republicans were less swayed by the *The New York Times*' signal (or perhaps even skeptical of it), compared to Democrats, who likely exhibited stronger reductions in belief in Trump's chances of winning.

Next, we turn to the second condition, that is when the source is *Fox News* (see Table 10). The p-value (0.022) indicates that the result is statistically significant at the $\alpha = 0.05$ threshold. This allows us to reject the null hypothesis and conclude that there is a meaningful difference between Democrats and Republicans in their LLR values. Similar to the first condition, Republicans had a higher rank sum than Democrats. This implies that Republicans exhibited less negative belief updating (or possibly even positive updates) compared to Democrats when the signal came from *Fox News*. The significant difference suggests that Republicans found the signal from *Fox News* more credible and consistent with their priors. In contrast, Democrats likely dismissed or downplayed

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			

the signal when it came from a source they perceive as ideologically opposed to their own views. Overall, the statistically significant result underscores that Republicans and Democrats reacted differently to the signal from *Fox News*. Republicans were more receptive to the signal, whereas Democrats were likely resistant.

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			

D Comparison Over Source

D.1 Study A

The results presented here analyze the differences in belief updating between the two sources of the signal, *The New York Times* and *Fox News*, in Study A. This comparison is conducted separately for Democrats and Republicans (using the Wilcoxon rank-sum (Mann-Whitney) test. The focus is on how individuals from each party respond to the pro-Republican signal (from Silver, predicting Trump's victory) based on its attributed source.

First, we test the condition when the receivers are Democrats (see Table 11 in Appendix D). The p-value (0.254) is much higher than the conventional significance threshold ($\alpha = 0.05$), so we fail to reject the null hypothesis. This means there is no evidence of a significant difference in belief updating for Democrats based on the source of the pro-Republican signal. The rank sums are slightly higher for *The New York Times* (4940.5) than for *Fox News* (4512.5), but the difference is not statistically significant. This suggests that Democrats showed a similar level of resistance to the signal regardless of its source. Overall, for Democrats, the source of the pro-Republican signal predicting Trump's victory had no significant impact on their belief updating. This result suggests that Democrats' responses were primarily shaped by the *lack of informational surprise*, as the signal content sharply conflicted with their prior beliefs and thus was easily dismissed—regardless of whether it came from an ideologically aligned or incongruent source.

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 $>$ $\lvert z \rvert$): 0.2535			
Exact p-value: 0.2551			

Next, we test the condition when the receivers are Republicans (see Table 12 in Appendix D). The p-value (0.4387) is much higher than the conventional significance threshold ($\alpha = 0.05$), so we fail to reject the null hypothesis. This means there is no evidence of a significant difference in belief updating for Republicans based on the source

of the pro-Republican signal. The rank sums are slightly higher for *The New York Times* (4455) than for *Fox News* (4191), but the difference is not statistically significant. This suggests that Republicans also showed similar levels of belief updating regardless of the source. Overall, for Republicans, the source of the pro-Republican signal predicting Trump's victory had no significant impact on their belief updating. This result suggests that Republicans' responses were shaped by the *low surprise* of the signal, which aligned with their prior beliefs and expectations. Because both the content and the attributed sources were broadly congruent with Republicans' worldview, the information likely lacked sufficient novelty to trigger differential updating across source conditions.

Thus, in Study A, when comparing belief updating across sources (*The New York Times* vs. *Fox News*), the results indicate no statistically significant differences for either Democrats or Republicans. This highlights that the content of the signal (predicting Trump's victory) played a more central role than the attributed source in shaping participants' responses. These findings suggest that variation in belief updating was primarily driven by the degree of *surprise*—specifically, how expected or unexpected the signal was relative to participants' prior beliefs—rather than by source credibility or ideological alignment alone. For both groups, the signal lacked sufficient informational novelty to generate differential responses across sources.

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			

D.2 Study B

The results presented here analyze the differences in belief updating between the two sources of the signal, *The New York Times*and *Fox News*, in Study B. This comparison is conducted separately for Democrats and Republicans using the Wilcoxon rank-sum (Mann-Whitney) test. The focus is on how individuals from each party respond to the pro-Democrat signal (from Lichtman, predicting Trump's loss) based on its attributed source.

First, we test the condition when the receivers are Democrats (see Table 13 in Appendix D). The p-value (0.469) is much higher than the conventional significance threshold ($\alpha = 0.05$), so we fail to reject the null hypothesis. This means that there is no evidence of a significant difference in belief updating for Democrats based on the source of the pro-Democrat signal. The rank sums for both sources are very similar (4412.5 vs. 4365.5), suggesting that Democrats updated their beliefs in a comparable way regardless of whether the signal came from *The New York Times* or *Fox News*. Overall, for Democrats, the source of the signal had no statistically significant impact on their belief updating. This indicates that Democrats treated the signal about Trump's predicted loss similarly whether it was attributed to *The New York Times* (a trusted in-group source) or *Fox News* (a typically out-group source). This result suggests that the signal aligned closely with Democrats' prior beliefs and thus failed to produce informational surprise, making the source attribution relatively inconsequential in shaping belief updating.

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			

Next, we test the condition when the receivers are Republicans (see Table 14 in Appendix D). The p-value (0.8848) is much higher than the conventional significance threshold ($\alpha = 0.05$), so we fail to reject the null hypothesis. This means that there is no evidence of a significant difference in belief updating for Republicans based on the source of the pro-Democrat signal. Similar to the Democrats, the rank sums for Republicans are almost identical (4157.5 vs. 4098.5), suggesting that Republicans also updated

their beliefs similarly, irrespective of whether the signal came from *The New York Times* or *Fox News*. Overall, for Republicans, the source of the signal also had no statistically significant impact on their belief updating. This suggests that Republicans' resistance to the pro-Democrat signal was consistent regardless of whether it came from *The New York Times* (an out-group source) or *Fox News* (an in-group source). This result may reflect a *lack of informational surprise*, as the signal sharply conflicted with Republicans' priors, making it easy to dismiss—regardless of the credibility or ideological alignment of the attributed source.

Thus, in Study B, when comparing belief updating across sources (*The New York Times* vs. *Fox News*), the results indicate no statistically significant differences for either Democrats or Republicans. This highlights that the content of the signal (predicting Trump's loss) played a more central role than the attributed source in shaping participants' responses. These findings contrast with typical source credibility effects observed in other studies, suggesting that the *level of informational surprise*—determined by how the signal aligns with or deviates from partisan priors—can override the influence of source identity. In this case, the signal either strongly confirmed prior beliefs (for Democrats) or sharply contradicted them (for Republicans), leaving little room for source-driven differences in belief updating.

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			

E Sources Classifications

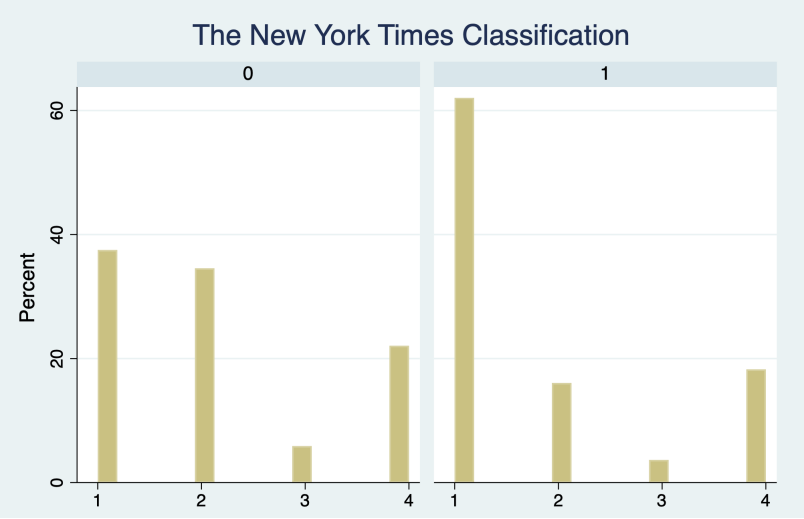

Figure 8: *The New York Times* Classification: How Democrats (right panel) and Republicans (left panel) classified

Note: The Source Classification question asks subjects 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.

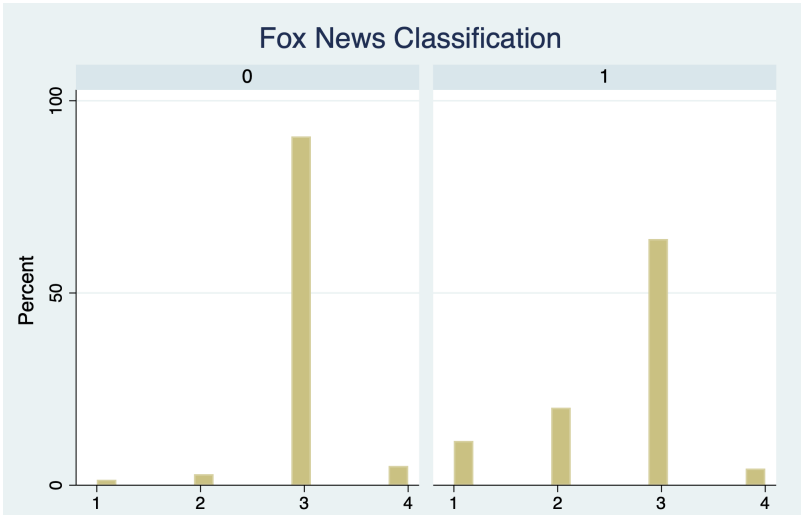

Figure 9: *Fox News* Classification: How Democrats (right panel) and Republicans (left panel) classified

Note: The Source Classification question asks subjects 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.

Classification of The New York Times. This analysis examines whether Democrats and Republicans systematically differ in how they classify the media source when the source is restricted to *The New York Times* (see Table 15). The rank sum for Democrats is significantly higher than expected, meaning they tend to classify *The New York Times* differently from Republicans in a systematic way. More specifically, Democrats are more likely to classify *The New York Times* as a Democratic or neutral source; Republicans are more likely to classify *The New York Times* as a Republican source or say "I do not know" less frequently. The fact that the distributions are not equal suggests partisan bias in media perception: (i) Democrats may see *The New York Times* as aligned with their views (Democrat-leaning or neutral); (ii) Republicans may disagree with this classification, potentially rating it as more neutral or even Republican (if they trust it) or may avoid classifying it clearly. The large z-score (3.280) and the low p-value (0.001) indicate a robust difference in perception between the two groups. The shift in rank sums shows that the effect is not random but part of a systematic pattern in how partisans interpret media sources.

Table 15: Two-Sample Wilcoxon Rank-Sum Test: Classification of *The New York Times*

PARTY	**Obs**	**Rank Sum**	**Expected**
Democrats (0)	136	20,606.5	18,632
Republicans (1)	137	16,794.5	18,769
Combined	273	37,401	37,401
Variance Adjustments			
Unadjusted variance		425,430.67	
Adjustment for ties		-62,985.37	
Adjusted variance		362,445.29	
Hypothesis Test			
Null Hypothesis		$Q17(PARTY = 0) = Q17(PARTY = 1)$	
z-statistic		3.280	
p-value		0.001	

Classification of Fox News. This analysis examines whether Democrats and Republicans systematically differ in how they classify the media source when the source is restricted to *Fox News* (see Table 16). The rank sum for Democrats is significantly higher than expected, meaning they tend to classify *Fox News* differently from Republicans in a systematic way. More specifically, Democrats are more likely to classify *Fox News* as a Republican-leaning source. Republicans are more likely to classify *Fox News* as neutral or even Democratic, or they may be less likely to choose "I do not know." The strong statis-

Table 16: Two-Sample Wilcoxon Rank-Sum Test: Classification of *Fox News*

PARTY	**Obs**	**Rank Sum**	**Expected**
Democrats (0)	140	22,197.5	19,600
Republicans (1)	139	16,862.5	19,460
Combined	279	39,060	39,060
Variance Adjustments			
Unadjusted variance		454,066.67	
Adjustment for ties		-211,551.67	
Adjusted variance		242,515.00	
Hypothesis Test			
Null Hypothesis		$Q17(PARTY = 0) = Q17(PARTY = 1)$	
z-statistic		5.275	
p-value		0.000	

tical difference suggests that Democrats and Republicans have fundamentally different perceptions of *Fox News*: (i) Democrats overwhelmingly classify it as Republican-leaning; (ii) Republicans may be more varied in their classification, possibly considering it more neutral or even Democratic in some cases (depending on their views on editorial shifts within *Fox News*). The very large z-score (5.275) and the extremely small p-value (0.000) indicate a strong, systematic difference in how Democrats and Republicans classify *Fox News*. This is even stronger than the difference observed for *The New York Times* in the previous test, suggesting that *Fox News* is perceived as even more partisan compared to *The New York Times*.

F 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).[10] 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".[11] 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 that influence judgment and decision-making.[12] To account for potential demographic influences, we

[10]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.

[11]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.

[12]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,

collected information on 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.

Finally, to account for potential demographic influences, we gathered data on participants' age, highest level of education, and geographic location (state) at the time of the study. These standard socio-demographic variables offer valuable context for interpreting differences in responses and belief updating across various population groups.

(iv) Not enough information].

G Variables Description

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

Source Classification categorizes news sources as Democrat (1), Neutral (2), Republican (3), and "I do not know" (4).

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).

H Regressions' Results without Controls

H.1 Study A

Table 17: Direct and Interaction Effects without Controls on the LLR in Study A

	(1) LLR	(2) LLR	(3) LLR
Prior		–0.005	–0.005*
		(0.003)	(0.003)
Source = Fox News	–0.065	–0.052	–0.044
	(0.056)	(0.052)	(0.074)
Party = Republicans	0.064	0.160*	–0.157
	(0.058)	(0.063)	(0.213)
Positive Partisanship	0.055*	0.057*	0.041
	(0.028)	(0.028)	(0.035)
Negative Partisanship	–0.021	–0.023	–0.044
	(0.025)	(0.025)	(0.032)
Fox News × Republicans			–0.018
			(0.107)
Republicans × Positive Partisanship			0.029
			(0.054)
Republicans × Negative Partisanship			0.046
			(0.050)
Constant	–0.072	0.151	0.337
	(0.115)	(0.155)	(0.176)
R^2	0.030	0.057	0.064
N	268	268	268

Standard errors in parentheses

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

H.2 Study B

Table 18: Direct and Interaction Effects without Controls on the LLR in Study B

	(1) LLR	(2) LLR	(3) LLR
Prior		–0.010**	–0.010**
		(0.003)	(0.003)
Source = Fox News	–0.017	–0.026	–0.179
	(0.081)	(0.078)	(0.112)
Party = Republicans	0.189*	0.348***	0.280
	(0.093)	(0.099)	(0.352)
Positive Partisanship	–0.016	0.006	–0.016
	(0.031)	(0.029)	(0.041)
Negative Partisanship	0.070	0.055	0.089
	(0.039)	(0.036)	(0.049)
Fox News × Republicans			0.355*
			(0.162)
Republicans × Positive Partisanship			0.073
			(0.061)
Republicans × Negative Partisanship			–0.109
			(0.073)
Constant	–0.562**	–0.136	–0.092
	(0.208)	(0.216)	(0.298)
R^2	0.028	0.074	0.101
N	260	260	260

Standard errors in parentheses

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