Does Partisanship or Arguing Activate Political Motivated Reasoning?*

Andrew T. Little† Salvatore Nunnari‡

December 3, 2025

Abstract

Debates about when and whether (partisan) directional motives influence information processing are hard to resolve because rational and motivated learning often look similar. We develop an experimental design to distinguish between these possibilities which focuses on the order in which information is presented. A core tenet of Bayesian updating is that order should not impact final beliefs, but if some information changes the motivation to process other information, order effects may emerge. In our first study, we randomize the partisanship of real endorsements for ballot propositions, as well as whether participants learn about these endorsements before observing other information about the propositions. We find no evidence of motivated information processing across several tests. In a second study, we randomize whether participants themselves argue for or against a proposition, and whether they know this position before observing other information. This produces a strong order effect: being randomized to argue for versus against a position affects beliefs more when it is learned before information about the proposition is provided. We also find suggestive evidence that this order effect is driven by selective attention to information. Overall, our results suggest that motivated reasoning about politics is less prevalent than commonly believed, but may arise primarily when people are in an argumentative mindset.

*We are grateful to Don Green, Gabe Lenz, Minahil Malik, Carlo Prato, Michael Thaler, Joel van der Weele, and audiences at NYU, Bocconi University, Vanderbilt University, and the 2025 Benelux PECO in Rotterdam for helpful comments and suggestions. This study was approved by Bocconi Ethics Committee (Application RA000927). The pre-registrations are available at https://doi.org/10.17605/OSF.IO/XQ46N and https://doi.org/10.17605/OSF.IO/Q9RW2. We acknowledge financial support from the European Research Council (POPULIZATION Grant No. 852526).

†UC Berkeley, Travers Department of Political Science; andrew.little@berkeley.edu.

‡Bocconi University, Department of Economics; CEPR; CESifo; salvatore.nunnari@unibocconi.it.

1 Introduction

Discussions of politics frequently devolve into competing claims that those with different beliefs must be irrational. For example, consider how different groups may respond to a news story about an immigrant committing a crime. Those who support more immigration will likely accuse the media of overly focusing on examples which support a narrative that immigrants are dangerous rather than looking at overall crime rates. On the flip-side, those who oppose immigration will accuse their opponents of being motivated to explain away the costs of a more open policy. More generally, both sides of political debates often end up equally confident that they are interpreting information in a rational fashion while the other side's beliefs are driven by biases like motivated reasoning. Who is most susceptible to motivated reasoning, and when? Are people particularly irrational when arguing about politics?

A key barrier to answering questions like these is that reliably detecting motivated reasoning is hard. A specific challenge is that those who believe something is very likely because they are motivated to believe it respond to new information in a similar way as those who think it is very likely because of real prior information (Koehler, 1993; Tappin, Pennycook and Rand, 2020; Druckman and McGrath, 2019; Little, 2025). More broadly, we typically cannot assess whether someone is updating in response to new information in a rational way without knowing how they interpret that information. Formally, even if we know a prior and posterior belief, we cannot evaluate whether the update occurred via Bayes' rule without knowing the "true" likelihood function of the information. Existing work which recognizes this problem typically gets around it by using artificial signals which do imply a correct way to update beliefs (e.g., Hill, 2017; Thaler, 2024), though this comes at a cost of realism as it is not the kind of (political) information people encounter in the real world.

The first main contribution of this paper is to develop an experimental research paradigm which gets around these challenges, providing clean tests for motivated information processing with realistic political information. Our tests are based on detecting deviations from a general property of rational learning: the order in which information is learned should not matter (see, e.g., DeGroot, 2005, section 8.12). We also look at when certain kinds of pieces of information are more likely to be attended to and remembered.[1]

Motivated reasoning may lead to both selective attention and memory and "order effects." In particular, attention to or memory of a piece of information may depend on whether it is "what one

[1] Such selective attention and memory need not lead to biased beliefs if people understand the selection problem in what gets remembered, but substantial evidence indicates this is unlikely (e.g., Enke, 2020; Brundage, Little and You, 2024).

wants to hear." However, whether one piece of information is good or bad news may depend on other information. For example, suppose (as in our first study) people learn which party tends to support a policy and additional information which indicates that policy is good or bad. If people want to think their party supports good policies, they may interpret or pay attention to this other information differently if they first observe a partisan endorsement which "activates" a directional motive. More generally, by manipulating *when* people learn about the information which plausibly generates directional motives, we can compare their final beliefs holding fixed the information they ultimately possess.

Our second main contribution arises from applying this general design to two preregistered experiments, which share many features in common but differ in whether participants are only *evaluating* a policy (Study 1) or *arguing* about it as well (Study 2). In Study 1, the potential source of a directional motive is learning information about which party tends to support a proposition. In Study 2, we provide no partisan information and the potential source of a directional motive is being randomized to argue for or against the proposition. We find no evidence for motivated reasoning in Study 1, while we do in Study 2. Thus, our main substantive conclusion is that motivated reasoning is more prevalent when people are in an argumentative mindset, and this may be more important than partisanship by itself.

For both studies, we identify recent U.S. ballot propositions that had bipartisan support and opposition. In Study 1, we randomize the partisan composition of examples of supporters and opponents of the proposition, as well as whether participants observe some reasons (or "other information") about why others are in favor of or against the proposition. We elicit a prior belief about whether the participant would vote for the bill before learning both, and a posterior after learning an endorsement and reasons. To compare with previous studies, we also ask how persuasive they find the reasons.

Finally, and most important for our design, we randomize the order in which participants see the endorsement and the other information. For rational learners, this order should not matter for their ultimate evaluation. If participants believe their own party tends to support the bill before seeing the reasons, they may dismiss reasons against it, and if they believe the other party supports the bill they may dismiss reasons for it. In other words, if partisan motivated reasoning affects how other information is evaluated, we should see the effect of receiving a copartisan endorsement as more positive when it is revealed before the reasons rather than after.

We also conduct related tests on memory. The key prediction here is that information which is congruent with the partisan endorsement (e.g., a pro reason and a copartisan endorsement) will be remembered better if the partisan endorsement comes before the reasons.

Our results replicate several patterns from the past literature which are *not* clean tests of motivated reasoning. In particular, participants (1) expect that copartisans will endorse propositions they like but not those they dislike, (2) rate reasons consistent with their prior belief and congruent with the partisan treatment (e.g., a positive reason after learning a copartisan endorsement) as more persuasive, and (3) are heavily influenced by copartisan endorsements, which exert more than twice the influence on vote intention as receiving a positive reason. However, all of these results can be explained by rational inference among those who sincerely believe that they and their own party tend to support better policies.

In contrast, we find precisely estimated null results on all of the hypotheses which provide a cleaner test for motivated reasoning. In particular, we find no evidence that copartisan endorsements matter more when presented first. Whether looking at participants' beliefs about their own support for the proposition or an incentivized question about how many copartisans would support it our estimates tend to be negative and can rule out substantial positive effects. We also find nulls across subgroups by partisanship and prior beliefs. Similarly, we find no evidence that "party congruent" information is more likely to be recalled correctly or less likely to produce a "don't remember" response.

In the second study, we draw on experiments in economics which randomize the position people argue for (e.g., Babcock and Loewenstein, 1997; Schwardmann, Tripodi and Van der Weele, 2022) as well as whether they know this before encountering other information (e.g., Babcock et al., 1995; Gneezy et al., 2020). To be as close to Study 1 as possible, we use a subset of the same ballot propositions and the same key outcome measures, but rather than randomizing partisan endorsements, we randomize whether participants themselves argue for or against the proposition. The relevant order manipulation here is that some participants learn their position right after we give a basic description of the proposition and elicit prior beliefs, followed by observing four pieces of information that could be used in arguments, and then make their argument. Others do not learn their position until after they have observed the information. If arguing activates motivated reasoning, then not only should arguing for or against a position affect posterior beliefs, but this effect should be magnified when one knows the position when processing the interim information.

Unlike in our first study, but consistent with Babcock et al. (1995) and Gneezy et al. (2020), here we find strong evidence of an order effect. In our preregistered preferred specification, participants who are randomized to argue for the proposition but do not know this until after the information are about 7% more likely to say they would vote for it than those assigned to argue against. This difference increases to nearly 12% among those who do know the position at the outset. We find smaller but still significant differences in an incentivized question about whether

others would vote for the proposition.

One innovation relative to existing studies that randomize arguing position is that our design also allows for new tests of the mechanism of motivated information processing. Our preregistered test for the mechanism of motivated memory is that "congruent" information – i.e., positive information for those arguing for, and negative for those arguing against – receives relatively more attention and hence is more likely to be remembered when the position is known at the outset. The point estimates are in this direction, but are small and not significant at conventional levels.

We then conduct two post hoc tests which provide suggestive evidence that knowing the argument position does affect processing of information, even if this does not substantially affect answers to our memory questions. First, we ask for "interim" beliefs after each piece of information is revealed, and our key treatment groups see the same blend of positive and negative information. As more information is revealed, average beliefs move up for people who know they will be arguing for the proposition, relative to those who do not know the position or know they will argue against. Point estimates indicate this "biased interim learning" effect explains about half of the overall order effect. Second, we ask whether pieces of information are *used* in the argument. Unsurprisingly, congruent pieces of information are more likely to show up in the arguments participants make. More subtly, this effect is larger when the position (and hence congruence) is known when the information is revealed.

Overall, we interpret this as strong evidence of an order effect among those who need to argue for or against a policy. This effect appears to be at least partly driven by motivated attention to (and perhaps memory of) information that is congruent with one's position. As there is no party information given, this motivated information processing arises purely from the fact that people are arguing, rather than from any partisan motives.

Further, in both studies, we find little heterogeneity in patterns indicating motivated information processing across party lines, by strength of attachment to party, nor among those with different prior beliefs about the proposition or who make longer or shorter arguments. While there may be differences in predilection to motivated reasoning on variables we did not measure, we tentatively conclude that situational factors (arguing vs merely evaluating) are more important than personal traits.

Even more speculatively, our results suggest that the emphasis in the motivated reasoning literature on whether it undermines ordinary citizens' role in the functioning of democracy may be misplaced. Rather, those who are most invested in politics and spend time and effort arguing about it – elites, activists, and politicians themselves – may be most likely to process information in a motivated fashion. If we want to know whether motivated information processing harms democracy,

our emphasis should perhaps be on the beliefs of politicians rather than voters.

2 Related Literature

The question of whether the desire to hold certain beliefs affects how people process information has been heavily debated for decades; see, e.g., Kunda (1990) for an influential early review, Ditto et al. (2025) for a more recent review, and Bullock and Lenz (2019) for a review in the context of partisan bias. For example, our first experiment draws on how information about partisan endorsement affects evaluation of arguments and ensuing beliefs. It is widely known that individuals tend to support policies endorsed by copartisans (e.g., Lenz, 2009) and evaluate (partisan) congruent arguments and sources more favorably (e.g., Taber and Lodge, 2006; Kahan, 2015), though there is considerable debate about whether this is driven by the fact that partisan cues are informative or lead to biased evaluation.

A central challenge, which has been recognized since at least Koehler (1993), is that people holding different prior beliefs *should* respond to information in different ways, in a manner which may look irrational or driven by motivated reasoning (see also Tappin, Pennycook and Rand 2020; Druckman and McGrath 2019 for discussions of this "observational equivalence" problem, and Little 2025 for a formal treatment which we build on here.) Next we provide a brief overview of some related approaches to this problem, followed by discussion of other related literature.

Right-Answer Designs. One prominent approach is to give people information which does entail a correct Bayesian update, allowing for a clear test of deviations from Bayes' rule. For example, when eliciting prior beliefs and then giving a signal which is true with a given probability (known to the participants), this implies a correct benchmark. A general theme from studies that do this is that participants respond to information in the direction indicated by Bayes' rule but not perfectly (e.g., Grether 1980; Hill 2017; see Benjamin 2019 for an overview).

A related (often overlapping) approach is to exploit the symmetry in information to show that participants respond differently to good or bad news (Hill, 2017; Eil and Rao, 2011; Möbius et al., 2022). Thaler (2024) provides a particularly sharp example of this kind of design, by eliciting prior beliefs and then providing a noisy signal which indicates the truth is either above or below the median of the prior, which by construction is not informative about whether the signal is correct. Still, participants tend to "trust" signals which would move their beliefs in the desired direction.

Some other recent work that relies on giving questions with a correct answer suggests that motivated reasoning may arise more when people are confused. For example Hagenbach and

Saucet (2025) find people do a better job of applying skepticism in a disclosure game when the naive interpretation of the information is less favorable, and Exley and Kessler (2024) find people will make basic addition errors if doing so leads to more pleasant beliefs.[2]

Alas, most information we observe in the world does not take the form of a signal which is correct with a known probability, nor does it tell us the truth is likely above or below our prior.[3] Our general goal here is to develop a similarly sharp test for motivated reasoning which uses more natural pieces of information.

Order (and Memory). Closer to our approach methodologically is work which studies how the order in which information is presented may affect beliefs and recall. For example, Bransford and Johnson (1972) show that relatively vague sentences ("First you arrange things into different groups depending on their makeup") are more likely to be recalled when contextual information (they are sentences about doing laundry) is given before rather than after. Importantly, motivated reasoning is not the only reason why the order of information may matter.[4] The "serial-position" effect is another major finding in the study of memory. When multiple pieces of information are presented, those near the beginning and end are more likely to be remembered than those in the middle (e.g., Murdock Jr, 1962).[5]

While it is impossible to pinpoint that any order effect must be driven by motivated reasoning, our experimental designs reduce the potential for alternative explanations in a few ways. First, our information which may trigger directional motives always comes first or last, which past work tends to find to be the most impactful positions (Murdock Jr, 1962). Second, we provide a suite of auxiliary tests of the mechanism that other information receives more or less attention based on whether it is known to be congruent with directional motives. Finally, our key substantive results come from comparing a similar order manipulation across two settings that plausibly vary in the

[2]While not exactly relying on the existence of a correct answer, Lilley and Wheaton (2024) present a related test by comparing belief updating in response to information when presented as fact versus as a hypothetical. They argue that in the hypothetical case there is less incentive to apply motivated reasoning, and indeed find less updating in response to the real versus hypothetical information, particularly for "unfavorable" information.

[3]Our experiments hold the information people have fixed, which precludes a direct comparison with some past work which looks at motivated information seeking (Taber and Lodge, 2006; Grossman and Van Der Weele, 2017; Chen and Heese, Forthcoming). This mechanism could lead to biased beliefs even without motivated processing of the information people do encounter, a point we revisit in the conclusion.

[4]One example studying order and motivated reasoning using a much more abstract setting is Je and Youn (2024), who manipulate the order of learning whether a signal is informative changes beliefs in a "ball and urn" experiment, finding this does influence beliefs for those who will be paid more when a "preferred color" is chosen, though only for those with certain prior beliefs.

[5]Other theoretical work provides potential reasons for order effects driven by related ideas like confirmation bias (Rabin and Schrag, 1999), as well as less related mechanisms like improper updating in multivariate settings (Cheng and Hsiaw, 2022; Koçak, 2018).

degree to which directional motives are activated.

Arguing (and Order). One setting with relatively strong evidence for motivated reasoning is where people argue for or against positions, and their position is randomized. This tends to move beliefs in line with assigned position, often with large effects (e.g., Babcock and Loewenstein, 1997; Schwardmann and Van der Weele, 2019; Schwardmann, Tripodi and Van der Weele, 2022; Zhang and Rand, 2025). In particular, we build on work which manipulates the order in which people learn their incentives and receive other information (Babcock et al., 1995; Gneezy et al., 2020; Saccardo and Serra-Garcia, 2023). This work typically finds that revealing the arguing position or incentive before receiving other information magnifies the effect of the position.[6] In addition to allowing for a closer contrast to our study without arguing, the order manipulation here gives a more precise test that the effect of arguing is driven at least in part by motivated information processing, rather than an experimenter demand effect (see also Zhang and Rand, 2025, which provides an alternative approach to testing this connection by showing that randomizing incentives to pay attention to information has a similar effect as directly randomizing a persuasive position).

In addition to our application to a political setting, a key innovation relative to this work is to contrast the results in an experiment with arguing to one that is otherwise similar but where participants themselves only evaluate information. Arguing may be a particularly good domain to study motivated reasoning, and some even claim that arguing is central to understanding how humans reason more widely (Mercier and Sperber, 2011). Still, at a minimum there are situations where actively arguing plays a larger or smaller role in belief formation, so it is valuable to see whether people exhibit more or less motivated information processing based on this distinction.[7]

Memory. Finally, some of our mechanism tests draw on a growing literature, primarily in economics, which tests for motivated reasoning not by studying how beliefs are updated but what information is remembered. Suppression of unpleasant information is the key driver of motivated reasoning in the theoretical treatment in an influential series of papers by Bénabou and Tirole (2002, 2016). Several recent empirical papers find evidence that "pleasant" information is more likely to be recalled correctly, in the context of one's performance on an IQ test (Zimmermann, 2020; Chew, Huang and Zhao, 2020), investment success (Gödker, Jiao and Smeets, 2025), or

[6] One exception to this is Pace et al. (2025) where the ultimate choice is a consumption decision, and perhaps tellingly they do not find a similar order effect in this setting.

[7] Related ideas in popular books are that we may exhibit more cognitive biases when in a "soldier mindset" (Galef, 2021).

generosity in a behavioral game (Carlson et al., 2020).[8]

3 Theory

To fix ideas, we develop a theory in the context of our first study, though also note how it applies to the second.

3.1 Setup and Key Assumptions

Consider a participant forming a belief about whether a proposed ballot initiative is "good," in the sense that they would prefer to vote for it rather than against it if well-informed on the topic. Let $\omega = 1$ mean the proposal is good and $\omega = 0$ mean it is bad.

Consider the case where beliefs may be influenced by two pieces of information. One piece is an indicator for whether the policy is endorsed by in-party politicians, $s_p \in \{0, 1\}$. In our first study, we will always show either copartisans endorsing and out-partisans opposing ($s_p = 1$, hereafter copartisan endorsement), or copartisans opposing and out-partisans endorsing ($s_p = 0$, hereafter out-party endorsement). That is, in the theory we will not distinguish between, e.g., a positive effect of a copartisan endorsement vs opposition from the out-party, since in the experiment these will always perfectly covary, as they typically covary in real scenarios. In our second study, the analog of the "partisan information" is the position (for or against) which the participant will be asked to argue.

The other piece of information is an argument about merits (or demerits) of the proposal itself and it is written $s_o \in \{0, 1\}$, where $s_o = 0$ is a negative piece of information and $s_o = 1$ is a positive piece of information.

Let s refer to the information held after some information is revealed, which could be one of these signals or both. If the prior belief that the proposition is good is p, the standard Bayesian posterior belief is:

$$\mathcal{P}(\omega = 1|s) = \frac{p\mathcal{P}(s|\omega = 1)}{p\mathcal{P}(s|\omega = 1) + (1 - p)\mathcal{P}(s|\omega = 0)}.$$

To differentiate with interim beliefs after only some information is revealed, we often call this the *final posterior* belief. It will be convenient to study the logit transformation of this posterior belief,

[8] See Little, Platas and Raffler (2023) for a discussion of how some of these results may be driven by "differential guessing" in the presence of uncertainty, which also guides how we design some of our tests.

which can be written:

$$\Lambda(\mathcal{P}(\omega = 1|s)) = \Lambda(p) + \log(l(s))$$

where $\Lambda(x) = \log(x/(1-x))$ is the logit function and $l(s) = \mathcal{P}(s|\omega = 1)/\mathcal{P}(s|\omega = 0)$ is the likelihood ratio of information s, i.e., how relatively likely information s is when the policy is good versus not. When $l(s) > 1$ the signals are more likely when $\omega = 1$; hence, $\log(l(s)) > 0$ and observing this signal increases the belief that $\omega = 1$; conversely, if $l(s) < 1$, observing the signal decreases the belief that $\omega = 1$.

Note that in the case where s contains both pieces of information ($s = (s_o, s_p)$) this belief is not a function of the order in which these signals were observed (or if they are observed at the same time). This is the formal expression of the standard feature of Bayesian updating that the order with which information is revealed does not affect the ultimate belief (DeGroot, 2005, section 8.12).

Importantly, this fact is true even if the partisan information affects how the other information is interpreted for a Bayesian. Intuitively, when this is true, the *interim* updates upon observing the other information may depend on knowledge of the partisan information. However, as long as both pieces of information are ultimately learned, participants will end up at the belief implied by both pieces of information.[9]

To keep formulas simple in the main text, we assume that the signals are conditionally independent ($\mathcal{P}(s_o, s_p|\omega) = \mathcal{P}(s_o|\omega)\mathcal{P}(s_p|\omega)$) and so we can write the final posterior belief upon observing both as:[10]

$$\Lambda(\mathcal{P}(\omega = 1|s_o, s_p)) = \Lambda(p) + \log(l(s_o)) + \log(l(s_p)). \tag{1}$$

It is natural to assume that $l(s_j = 1) \geq 1$ and $l(s_j = 0) \leq 1$, $j \in \{o, p\}$, meaning the participant updates (weakly) positively upon observing copartisan endorsements and pro arguments, and negatively for out-party endorsements and con arguments. In general, we assume these inequalities are strict, though note that in the context of the second study, where s_p is just information about the participant's position in a debate, it may be more natural to assume $l(s_p) = 1$, i.e., this does not

[9]As a simple example, suppose s_o is fully informative about ω when $s_p = 1$ (in which case $s_o = \omega$), and is not informative when $s_p = 0$. Formally, let $Pr(s_o = 1|s_p) = Pr(s_p = 1|s_o) = Pr(\omega|s_p) = 1/2$, $Pr(\omega = 1|s_o = 0, s_p) = 1/2$, and $Pr(\omega = 1|s_o = s_p = 1)$. In this case, a Bayesian who first observes s_p will not update in either direction after this information, as it simply tells how to interpret s_o which they don't yet know. Observing s_o will not change the belief about ω at all when observing $s_p = 0$, while she will known with certainty that $\omega = s_o$ when $s_p = 1$. If first observing s_o, a Bayesian will update partially in the direction of s_o upon this signal (in particular, to $1/4$ if $s_o = 0$ and to $3/4$ if $s_o = 1$), since they do not yet know if s_o is informative. When observing $s_p = 1$ she will then fully update in the direction of s_o, while observing $s_p = 0$ will move the belief back to the prior of $1/2$. In either case, the final belief will be 0 when $s_p = 1, s_o = 0$; 1, when $s_p = s_o = 1$; and $1/2$ when $s_p = 0$.

[10]See Appendix A for a discussion of the case without conditional independence, which again does not change the Bayesian analysis and does not change the qualitative conclusions of our analysis with motivated reasoning under a mild assumption.

actually provide true information about whether the policy is good or bad.

Now suppose there is some probability that a given signal will either be ignored (that is, not committed to memory) at the time it is seen, or not remembered (that is, not retrieved from memory) at the time when participants are asked to report their belief about ω (i.e., how likely they are to vote for a ballot proposition). To account for a variety of possible mechanisms, we define a_j as an indicator for whether the participant "accepts" piece of information j.

$$\Lambda(\mathcal{P}(\omega = 1|s_o, s_p)) = \Lambda(p) + a_o \log(l(s_o)) + a_p \log(l(s_p)).$$

If we interpret acceptance as a precondition for successful encoding in memory, then when a signal is not accepted ($a_j = 0$) it has no effect on the posterior, and it also cannot be recalled later.

Next, if such acceptance is a random variable as well, we can write the expected logit posterior belief averaging over the realization of what gets remembered as:

$$E[\Lambda(\mathcal{P}(\omega = 1|s_o, s_p))] = \Lambda(p) + \mathcal{P}(a_o = 1|\cdot) \log(l(s_o)) + \mathcal{P}(a_p = 1|\cdot) \log(l(s_p)).$$

The core idea of the theory is that even if the *likelihood functions* do not depend on the order in which information is presented—which again must be true in standard treatments of Bayesian learning—it might affect the attention paid to information.

In particular, following Little (2025), suppose that people are more likely to accept information when they know it is positive information about the in-party or in-party positions. Put another way, they may be apt to ignore negative information about their preferred party. However, if people do not know the party context when the other piece of information arrives, they cannot apply this "partisan filter." In our second study, the same idea applies, except the filter leads to a greater acceptance rate for pieces of information that can support the argument one is assigned to make. More relevant for our memory tests, information being accepted initially is a necessary condition for it to be remembered later, and so the order in which information is revealed may affect what is remembered.

More precisely, let $f \in \{o, p\}$ indicate which information is revealed first. Define "good news" as a positive piece of information about an in-party policy or a negative piece of information about an out-party policy. Likewise, define "bad news" as a negative piece of information about an in-party policy or a positive piece of information about an out-party policy. We make the following assumption:

Assumption 1. Good news is more likely to be remembered when party (or argument position) is

known:

$$\mathcal{P}(a_o = 1 | f = p, s_p = s_o) > \mathcal{P}(a_o = 1 | f = o, s_p = s_o)$$

and bad news is less likely to be remembered when party (or argument position) is known:

$$\mathcal{P}(a_o = 1 | f = p, s_p \neq s_o) < \mathcal{P}(a_o = 1 | f = o, s_p \neq s_o)$$

While we cannot directly test this assumption, which applies to when beliefs about the policy are elicited, we do indirectly test it by looking at recall later in the survey.

An important implicit aspect of this assumption is that if people learn the partisan valence or argument relevance of a piece of information later, they do not (or cannot) go back and fully "re-evaluate" whether the other piece of information was good or bad news. There are two reasons we believe this assumption is reasonable in the context of our experiment and more generally. First, as discussed above, for information to be remembered, people must pay attention to and encode it in memory when first seeing it. If the piece of information is not committed to memory in the first place, it cannot be remembered even if later information renders it more favorable. Second, even if incongruent arguments are partially remembered, it may be cognitively costly to go back and re-evaluate every piece of information one has about a random variable upon learning partisan information which could potentially change how the past information is viewed.

In order to keep the focus on this channel of selective memory driven by party information activating an information filter, we assume that while party information itself may be forgotten, the probability of remembering s_p does not depend on the order in which information is revealed or the content of either signal.

Assumption 2. Recall of s_p does not depend on f or s_j:

$$\mathcal{P}(a_p = 1 | s_o, s_p, f) = q > 0, \forall s_o, s_p, f$$

In Appendix A we discuss how letting this probability vary affects the results.

3.2 Predictions

The effect on support for the proposition of learning that the proposition is endorsed by the in-party vs the out-party as a function of the other piece of information and what information is

revealed first can be written as

$$\begin{aligned}\Delta_c(f, s_o) &= E[\Lambda(\mathcal{P}(\omega = 1|s_o, s_p = 1, f))] - E[\Lambda(\mathcal{P}(\omega = 1|s_o, s_p = 0, f))] \\ &= \underbrace{q(\log(l(s_p = 1)) - \log(l(s_p = 0)))}_{\text{direct party effect}} \\ &+ \underbrace{(\mathcal{P}(a_o|f, s_p = 1, s_o) - \mathcal{P}(a_o|f, s_p = 0, s_o)) \log(l(s_o))}_{\text{differential acceptance}}\end{aligned}$$

The first term captures the fact that the participant will be more likely to support the in-party-endorsed proposition regardless of what information is revealed first. (Note that in the context of our second study, if $l(s_p) = 1$ for both signals, this term drops out.) The second term captures how endorsement by the in-party affects the incorporation of the other information. The key question is how this changes when information about endorsers is revealed first:

$$\begin{aligned}\beta_{c1}(s_o) &\equiv \Delta_c(f = p, s_o) - \Delta_c(f = o, s_o) \\ &= \log(l(s_o))((\mathcal{P}(a_o|f = p, s_p = 1, s_o) - \mathcal{P}(a_o|f = p, s_p = 0, s_o)) \\ &- (\mathcal{P}(a_o|f = o, s_p = 1, s_o) - \mathcal{P}(a_o|f = o, s_p = 0, s_o)).\end{aligned} \tag{2}$$

Empirically, we will primarily be interested in the average difference in the effect of learning the endorsement first versus second, accounting for cases where the other piece of information is negative or positive:

$$\beta_{c1} = \sum_{s_o \in \{0,1\}} \mathcal{P}(s_o)\beta_{c1}(s_o) \tag{3}$$

These differences are all positive:[11]

Proposition 1. *When Assumption 1-2 hold, the effect of a copartisan endorsement (or arguing for a position) on the final posterior belief is more positive when revealed first:*

i for both negative and positive other facts ($\beta_{c1}(s_o) > 0$, $s_o \in \{0, 1\}$), and

ii averaging over negative and positive other facts ($\beta_{c1} > 0$).

See the Appendix A for a proof. Intuitively, when the information is good, it exerts a larger impact when party is revealed first because it is more likely to be remembered. Conversely, the

[11]The order effects for pro and con arguments are more nuanced; see the appendix for an analysis.

effect of bad information about policies supported by in-party politicians is attenuated when the participant knows the endorsement. Our primary specification in Study 1 will look at the average effect of learning the copartisan endorsement first versus second, averaging over positive and negative information, which is positive because both of the differences being averaged are positive. Similarly, in Study 2 we primarily look at the difference in the average effect of the argument position based on whether it is revealed before or after other information. In both cases, we take an order effect as evidence of motivated information processing, and a lack of an order effect as evidence against motivated information processing.[12]

4 Study 1

We now test these theoretical predictions in two survey experiments on real statewide ballot propositions. The core evaluation task in both studies is deciding how likely one would be to vote on ballot propositions which had bipartisan endorsement. In Study 1, we aim to see if learning information about what party tended to support the proposition before seeing a reason ("information") for or against the proposition affects the ultimate evaluation.[13] We identify propositions, as well as endorsements and reasons using the website Ballotpedia.org.

For example, one of our propositions is Colorado Proposition 131 in 2024, which would have introduced Ranked Choice Voting in the elections for several statewide offices. Our aim is to see how information about partisan endorsements and arguments for or against the proposition affect evaluations. For the endorsements, in the Proposition 131 case, in one treatment participants observe a screen with the following:

> Here are some examples of politicians who supported and opposed this ballot measure, and their party affiliation.
>
> Supporters:
> U.S. Sen. John Hickenlooper (Dem)
> Gov. Jared Polis (Dem)

[12] As discussed in (Little, 2025), a form of motivated reasoning which solely influences prior beliefs but not whether subsequent information is accepted or remembered would not produce such an order effect. However, this form of motivated reasoning is nearly impossible to detect by any design since it produces identical updating in response to any sequence of information. Further, motivated reasoning which only changes priors is less problematic in the long term, as people with different motives will eventually converge in beliefs with more information (Blackwell and Dubins, 1962).

[13] We described this other information to participants as an "argument," but to avoid confusion with arguments made by participants themselves in Study 2 we refer to it here as information or reasons.

Opponents
U.S. Representative Lauren Boebert (Rep)
Former State Representative Dave Williams (Rep)

An example of positive information/reasons for this proposition is:

> "Prop 131 is about giving voters more voice, choice and power in our elections. It's about giving us the power to vote our true preferences. ... It's about making candidates represent all of us. It's about making our leaders produce better results on the issues that we care about."

The randomization and outcome measures are described below.

4.1 Research Design

We recruited 1500 participants on Prolific, restricted to U.S. residents and nationals who have a high-quality track record on the platform and who identified as either Democrats or Republicans. Quota sampling ensured roughly equal representation from each party. Participants received a base payment of $3 with an opportunity to earn up to $2 as a bonus based on answers to factual questions. The median completion time was 25 minutes. The study was preregistered at OSF.[14]

The survey consisted of four parts:

1. **Party ID and Attention Check.** Participants first indicated their party preference and answered four questions measuring "social partisanship" from Huddy, Mason and Aarøe (2015), along with an attention check.[15]

2. **Proposition Evaluation.** Participants were shown six recent propositions on U.S. state ballots that received bipartisan support and opposition. The details of these propositions appear in the Appendix, and the order of presentation was randomized.

   For each proposition, participants first saw a short factual description and answered a comprehension question on the same screen (required for continuation). They then reported their *prior* belief—the probability they would vote in favor of the proposition if on their ballot. For simplicity and to differentiate from other belief elicitations, we refer to the beliefs that one would vote for the measure as their *support* for the measure. They also report an (incentivized) belief about the share of copartisans who would support the proposition.

[14] The preregistration is available at `https://doi.org/10.17605/OSF.IO/XQ46N`.
[15] Around 95% of participants passed the check.

Next, participants viewed information about partisan endorsements and reasons either for or against the proposition. The **order of these two pieces of information was randomized**, which plays a key role in our analysis.

For the endorsement, we showed two supporters and two opponents. We refer to the treatment where participants saw two Democratic supporters and two Republican opponents as a "Democratic endorsement," and similarly define a "Republican endorsement." When the endorsers match the participants' own party identification, we call it a "copartisan" or "in-party" endorsement, and call the opposite case an "out-party" endorsement. Because the selected propositions had endorsements and opposition across party lines, we could randomize whether endorsers or opponents were Democrats without deception. Each endorser's occupation was also displayed and later used for memory questions. As a manipulation check, participants then estimated the share of endorsers who were Democrats.

For the other information, we selected a short (three–four sentence) excerpt in favor of or against the proposition from Ballotpedia and randomized whether participants received a "positive" or "negative" information/reasons. After reading the reasons, participants rated their persuasiveness on a 0–100 scale.

Finally, participants again reported how likely they would be to vote for the measure—the *posterior* support. They also report a posterior belief about the share of copartisans who would support the proposition. For ease of comparison with the theory and Study 2 where we measure support more times, we typically call this the *final* (posterior) support and belief about others' support.

3. **Filler.** Participants completed an unrelated filler task, including forecasts of future political events, feeling thermometers toward partisans, and hypothetical money-allocation tasks measuring moral universalism (Enke, Rodriguez-Padilla and Zimmermann, 2022; Enke, Rodríguez-Padilla and Zimmermann, 2023).

4. **Memory.** Participants then answered 10 memory questions about the earlier information. For each proposition, they received three types of questions: (a) which of several sentences had appeared in the argument; (b) whether one of the endorsers was a Democrat or Republican; and (c) a "neutral" question about an endorser's occupation.

4.2 Hypotheses

We preregistered two sets of hypotheses: those concerning **evaluations** (belief updating) and those concerning **memory**. In line with Proposition 1, the key hypothesis about evaluations is that in-party endorsements matter more when presented first. For the memory questions, we define two outcomes: *correct* (an indicator for a correct answer) and *nr* (an indicator for answering "don't remember"). In the main text we focus on questions about whether a phrase was used in the reason, as these lead to the most straightforward prediction and are similar to the recall question used in Study 2.[16] In line with Assumption 1, the key hypothesis here is that information–endorsement congruence should matter more for recall when the endorsement is seen first.

4.3 Results

Intermediate Outcomes. Before getting to our key tests, we provide some analysis of intermediate outcomes to test whether our treatments work as intended and whether we can produce patterns similar to past studies on whether arguments are rated as persuasive.

Figure 1 shows the answer to the question about what share of endorsers are Democrats among several groups. In all of the figures to come, gray dots represent group averages, and black dots differences among groups with 95% confidence intervals.[17] The left facet provides a manipulation check: those who were shown a pair of Democratic endorsers expect about 70% of all endorsers to be Democrats, while those shown Republican examples expect this to be around 30% lower. The right facet tests the idea that participants will tend to expect their own party politicians to agree with them. The left two gray dots plot the average of Democrats who assigned a prior belief that they would vote for the proposition at or above 50%, and Democrats with a prior belief below this. Those who are positive about the proposition themselves expect the share of Democratic endorsers to be about 20% higher. The right two dots show the same comparison for Republicans; there is no significant difference here, perhaps because the question was phrased in terms of the share of Democrats.

Next we look at the evaluation of the quality of reasons given for the proposition. The left facet shows that participants who see a congruent reason (positive for those with a prior greater than or equal to 50, negative for those with a lower prior) rate the quality as about 10 points higher (on a 0-100 scale). The right panel compares three groups. First, those who see the reason before the partisan endorsement. For those who see the endorsement first, we look at the average

[16] We discuss other memory questions in Appendix C.

[17] These figures treat observations as independent; see Table A1 in Appendix D.1 for a regression-based analysis with fixed effects and clustering.

Figure 1: Study 1, Manipulation Check and Beliefs about Democratic Support

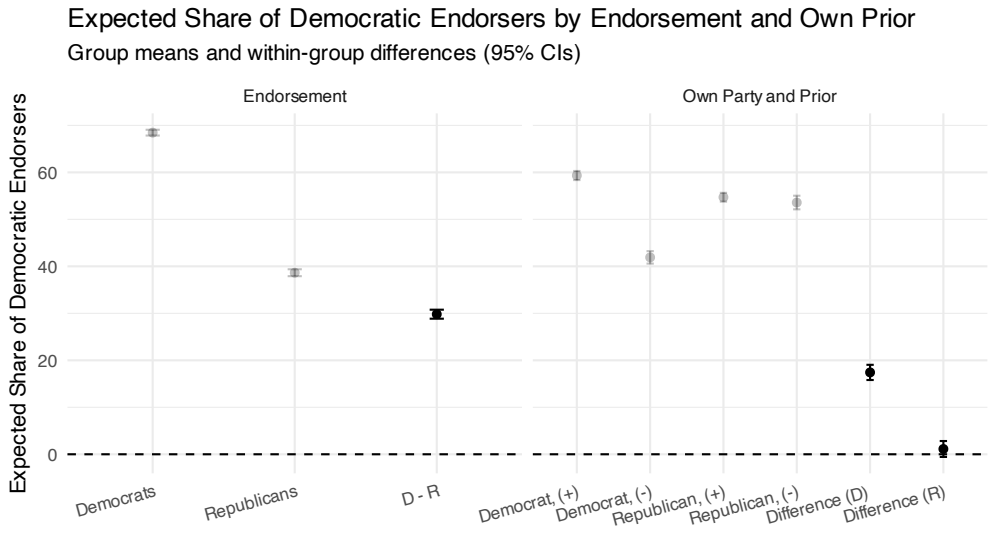
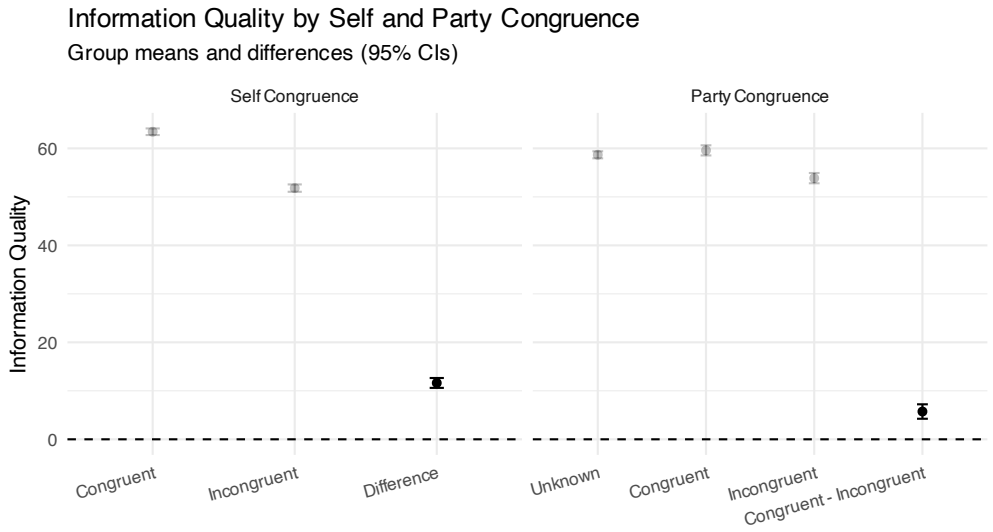


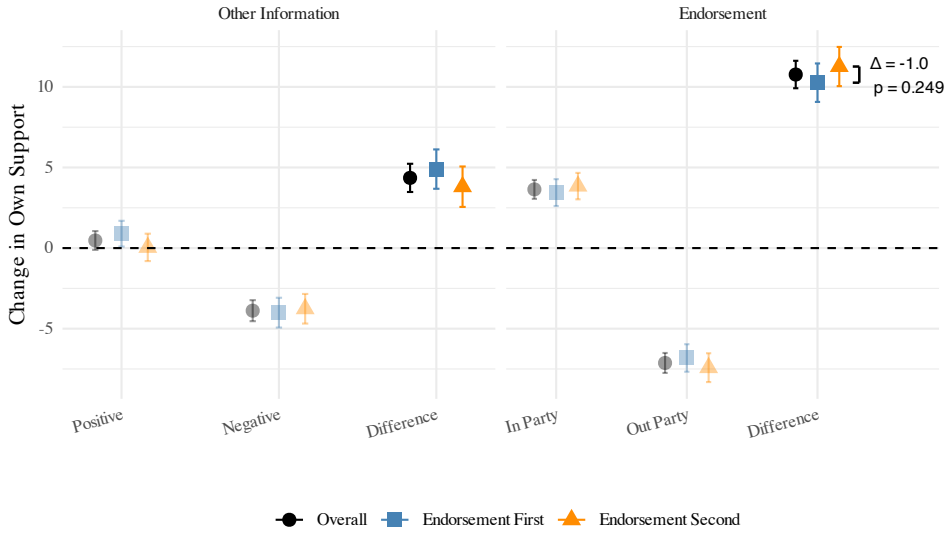
Figure 2: Study 1, Predictors of Information Quality Ratings



among those where the information and endorsement are congruent (i.e., in-party endorsement and positive information, or out-party endorsement and negative information) or incongruent. Reasons known to be congruent with the party endorsement are rated slightly higher compared to unknown, and those known to be incongruent are rated moderately lower. The contrast between known congruent and known incongruent is statistically significant.[18]

[18]See Table A2 for a regression version of this analysis with a similar conclusion.

Figure 3: Study 1, Predictors of Change in Support



Overall, these analyses show that participants tend to favorably rate reasons in line with their prior beliefs, and to some extent when they think the argument is likely to come from a copartisan. However, this could be driven by a sincere belief that one's own views and those by copartisans tend to be well-supported, and hence arguments not in line with them are not as sound. At a minimum, these do not comprise clean tests of motivated information processing

Evaluations. Our next analysis contains our cleaner test. We first present a visual version of the key comparisons, and then move to our preregistered regressions.

Figure 3 shows the average change in support from the prior to the final posterior among treatment groups (translucent points) and key differences among groups (solid points). Gray and black circles are averages not conditioning on the order in which information is presented, blue squares represent those who saw the endorsements first, and orange triangles those who saw the endorsement second.

The left facet looks at the effect of positive vs negative reasons. Starting with the black, positive information leads to a slight increase on average, while negative information leads to a modest decrease, with an overall difference of around 5%. There is no clear order effect for these treatments (as the theory predicts for both Bayesians and motivated reasoners.) The right facet shows the effect of in-party vs out-party endorsement is nearly twice this size on average. More importantly, our key prediction is that the difference between an in-party and out-party endorsement should be larger when the endorsement comes first (blue square above orange triangle). However, the point

Table 1: Study 1, Predictors of Final Support

	(1)	(2)	(3)	(4)	(5)
Prior (self)		0.75^{***}	0.75^{***}	0.75^{***}	0.70^{***}
		(0.01)	(0.01)	(0.01)	(0.01)
Positive	5.50^{***}	4.74^{***}	4.27^{***}	4.24^{***}	4.04^{***}
	(0.62)	(0.44)	(0.62)	(0.62)	(0.65)
Copartisan	11.86^{***}	11.47^{***}	11.83^{***}	11.87^{***}	12.49^{***}
	(0.71)	(0.53)	(0.70)	(0.70)	(0.74)
End first			0.00	-0.05	0.18
			(0.80)	(0.80)	(0.80)
Positive × End first			0.92	1.01	0.68
			(0.86)	(0.86)	(0.92)
Copartisan × End first			-0.72	-0.78	-0.78
			(0.86)	(0.85)	(0.92)
Fixed effects	No	No	No	Prop	Prop+Part
N	8622	8622	8622	8622	8622

Notes: Estimates from linear regressions of the final posterior (self) belief of voting for the proposition, in percentage points. Standard errors (clustered by participant) in parentheses. $*p < .05$, $**p < .01$.

estimate reveals the opposite, though the difference is not statistically significant.

Table 1 contains our preregistered version of this analysis, with regressions predicting the final support, expressed as a percentage.[19] The first column contains a basic specification with our two key treatments. Being assigned to a positive reason vs a negative reason increases the final support by around 5.5%, while receiving a copartisan endorsement vs a non-copartisan endorsement increases the support by around 11.9%. Both are consistent with expectations, and highly statistically significant. Column (2) adds the prior support, which leads to marginally different coefficients (and also lower standard errors).

Column (3) contains our first main test for order effects by adding an interaction term between receiving the endorsement first and the two main treatments. The key coefficient is the interaction between copartisan and endorsement first; analogous to the difference between the blue square and orange triangle in the right facet of figure 3. The motivated reasoning hypothesis is that this should be positive, but it is negative and not statistically distinguishable from zero. Column (4)

[19] Our preregistration did not specify how we would cluster our standard errors. We opt for a standard approach and cluster by participant in our main analysis, and then present results with alternative clustering approaches in the Appendix; in this case Table A3. Unless otherwise noted (as is true for this analysis), claims of statistical significance are unaffected by clustering choice.

adds proposition fixed effects, and column (5) is our preregistered specification with proposition and participant fixed effects. The key coefficient remains negative and not significant.[20] This null result is reasonably precise: the main effect of a copartisan endorsement is over 10%, and the upper bound on the 95% confidence interval on the interaction term is less than 1%.

Appendix Tables A5 and A6 contain versions of this analysis using logit-transformed beliefs to be closer to theory, and the (incentivized) belief about others supporting the proposition as the key dependent variable. We find similar null results on our key interaction coefficient. Appendix Table A7 presents variants of the regression in column (5) for various subsets of the data by partisanship. We find some heterogeneity in the reactions to the main treatments, but the key interaction term is negative and not significant for all groups. Appendix Table A4 shows little heterogeneity by ballot proposition.

Memory. Now we turn to the memory questions. In the main text, we focus on memory questions which ask participants if they saw a phrase in the argument. The first three columns of Table 2 use an indicator for being correct as the dependent variable. We multiply this indicator by 100 so coefficients can be interpreted in percentage points as in our other regressions. The fourth to sixth use an indicator for answering "don't remember" (again multiplied by 100). The independent variables are an indicator for the reason being congruent with the prior ("Self Congruent"), congruent with the party endorsement (e.g., a for reason when the endorsement was copartisan), and an interaction between the latter and whether the endorsement came first and hence congruence was known. Columns (1) and (4) only uses two congruence measures, columns (2) and (5) add the interaction terms, and column (3) and (6) add participant and question fixed effects.

The key prediction is that the interaction between party congruent and the endorsement coming first should be positive for being correct and negative for answering "don't remember." In both cases the coefficient is in this direction but not statistically significant without fixed effects, and the sign flips (and is still not significant) when fixed effects are added. The coefficient on self-congruence is positive for getting the answer correct without fixed effects, indicating people may be more likely to remember phrases from arguments which are congruent with their prior belief. However, this coefficient is not significant with fixed effects, and the coefficient on answering "don't remember" (inconsistent with people paying more attention to congruent arguments). Overall, we find little evidence of motivated memory of the reasons. Appendix Table A8 shows the main specifications with alternative ways of computing standard errors, and Appendix Tables A9

[20]The interaction between receiving a pro reason and receiving the endorsement first is positive but not statistically significant. Our pre-registration also proposed a comparison between these coefficients, which is not statistically significant.

Table 2: Study 1, Predictors of Recall, Phrases from Reasons

	(1) Correct	(2) Correct	(3) Correct	(4) DR	(5) DR	(6) DR
Party Congruent x End First		2.13	−0.49		−1.61	1.10
		(3.12)	(2.90)		(2.57)	(2.32)
Party Congruent	−1.70	−2.75	−3.62	0.68	1.48	−0.47
	(1.53)	(2.22)	(2.09)	(1.25)	(1.79)	(1.61)
Endorsement First		−1.26	1.84		1.27	−1.08
		(2.20)	(2.06)		(1.75)	(1.62)
Self Congruent	4.99**	4.99**	1.20	1.64	1.64	2.35
	(1.52)	(1.52)	(1.52)	(1.25)	(1.25)	(1.23)
Fixed effects	No FE	No FE	Q + Part	No FE	No FE	Q + Part
N	4309	4309	4309	4309	4309	4309

Notes: Estimates from linear regressions predicting answering a memory question correctly (columns 1-2) or "don't remember" (columns 3-4). Standard errors (clustered by participant) in parentheses.
*$p < .05$, **$p < .01$.

and A10 show similar analysis of the memory questions about the endorsers, which also provides no evidence of motivated information processing.

4.4 Discussion

The results in this study replicate many patterns often ascribed to motivated information processing. We find that learning a partisan endorsement has a large effect on voting intention, much larger than the effect of seeing a pro reason. We find that participants rate reasons consistent with their prior far more persuasive than those which go against it, and find reasons consistent with what copartisans say more persuasive than reasons consistent with what out-party members say. We also find some evidence that participants answer recall questions correctly about reasons consistent with their prior more often than those inconsistent with their prior.

However, as discussed above, all of these findings can also be explained by participants simply finding partisan endorsements informative, and finding reasons that go against their beliefs less credible, perhaps because they have confidence in their own evaluation.

In contrast, all of the predictions that our theory indicates are a test for motivated information processing – or, at least, non-Bayesian information processing – turn up null results, often quite sharp. On balance we view this as strong evidence against motivated information processing in this context.

One potential reason for the null results is that these are relatively neutral propositions without

any prior partisan valence. However, it is precisely this neutrality which allows us to randomize partisan endorsements and makes belief change plausible. Further, if we were to ask about a more hot-button issue like abortion, those who already feel strongly about the issue likely also have strong prior beliefs about which party supports different policies,.

5 Study 2

Our second study builds directly on the framework of Study 1 but shifts from *evaluating* policies to *arguing* about them. In doing so, it examines whether motivated reasoning becomes more pronounced when participants are placed in an explicitly argumentative role. The theoretical idea is that when individuals must defend a position, they may attend more closely to information that supports that position and discount or forget information that undermines it.

The design also provides another test for order effects. Rational (Bayesian) learning implies that the order in which information and incentives are presented should not matter: the same set of information should yield the same posterior beliefs regardless of sequence. However, if being tasked with defending a position changes how subsequent information is processed or remembered, then we should observe systematic differences depending on whether participants learn their argumentative position before or after seeing the relevant information.

In contrast to Study 1—which randomized the order of partisan endorsements and third-party arguments—Study 2 randomizes whether participants know *their own side* before observing policy information. This allows us to distinguish between purely informational effects of persuasion and those arising from motivated attention or memory.

5.1 Research Design

We recruited 1500 participants on Prolific, restricted to U.S. residents and nationals who have a high-quality track record on the platform and identified as either Democrats or Republicans. Slower recruiting of Republican participants and the possibility that people report a different ID in the survey led to a sample that was around 60% Democrat and around 40% Republican. As in Study 1, participants received a base payment of $3 and up to $2 in bonuses for correct factual responses in the memory section. The study was preregistered at OSF.[21] As pre-registered, participants who failed two initial attention checks are excluded from the analyses below.[22]

[21] The pre-registration is available at `https://doi.org/10.17605/OSF.IO/Q9RW2`.

[22] In the interest of space we do not present results without excluding these two participants, which are essentially identical to the ones we do present.

Each participant evaluated three recent statewide ballot propositions that had both bipartisan support and opposition. The study employed a 2×2 factorial randomization at the participant–proposition level:

1. whether the participant was assigned to **argue in favor of** ($for = 1$) or **against** ($for = 0$) the proposition; and

2. whether this assigned position was **revealed before** ($known = 1$) or **after** ($known = 0$) the participant read additional information about the proposition.

A secondary randomization determined whether participants saw a set of mostly positive or mostly negative informational items, drawn at random from predefined pools of short factual or opinionated statements.[23]

The survey unfolded as follows:

1. **Introduction and Priors.** Same as Study 1

2. **Information and Position Assignment.** Participants were informed they would later need to write an argument about the proposition. The sequence of information and position revelation was randomized:

   - In the **Known** condition, participants learned whether they would argue for or against before viewing any additional information.
   - In the **Unknown** condition, they saw the information first and only later learned their assigned side.

   In both cases, participants viewed four short pieces of information one at a time, with interim belief elicitations after each item.

3. **Argument and Posteriors.** After viewing all information and learning their assigned position (if not already known), participants wrote their argument. This task was incentivized: participants knew that one of their three arguments was going to be matched with another participant's argument on the same proposition and that they would receive a bonus if their argument was considered more persuasive by the research team. They then reported their final posterior support and beliefs about others' support.

[23] In particular, half of participants are assigned to a pool with four positive and one negative piece of information, while the other half draw from a pool with the opposite composition. Since they observe 4/5 of these, most will observe 1/4 or 3/4 positive pieces of information, while a smaller number observe 0/4 or 4/4 positive.

4. **Memory Section.** Same as Study 1

As in Study 1, this design holds constant the total information set but varies the *timing* of when participants' directional motivation is activated. If participants process information rationally, learning one's assigned side before or after should not affect posterior beliefs or recall accuracy. A larger effect of known positions on congruent information or memories would indicate motivated information processing.

5.2 Hypotheses

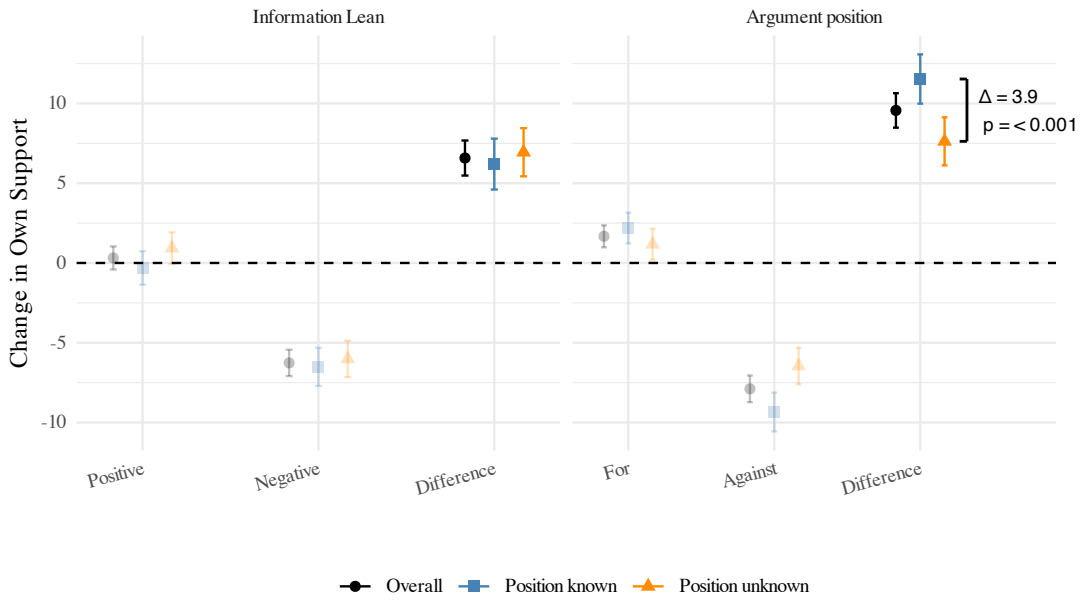
We preregistered two sets of hypotheses: those concerning **evaluations** (belief updating) and those concerning **memory**. In line with the evidence from previous research discussed above, a first hypothesis about evaluations is that being assigned to argue *for* a proposition will increase support for it, even if the role is revealed only after viewing the information. In line with Proposition 1, our key hypothesis about evaluations is that the effect of arguing for a position will be larger when the role is known *before* seeing the information. This second hypothesis is evidence of motivated reasoning. Regarding memory, a first hypotheses is that information congruent with one's own assigned side (positive if arguing for, negative if arguing against) will be more likely to be recalled correctly. Our key hypothesis though is that the memory advantage for congruent information will be stronger when the argumentative position is known in advance, reflecting motivated encoding or attention.

5.3 Results

Evaluation. Figure 4 presents a visual version of the key results. The translucent points represent group averages in change from prior to final support, and the solid points are differences between treatment groups. Black points do not differentiate between people who learn their arguing position before or after the other information, while blue squares subset to people who learn before and orange triangles to people who learn after. The left facet compares people who saw 0/4 or 1/4 positive pieces of information to people who saw 3/4 or 4/4 (see footnote 23), and shows that beliefs move in the direction of the information lean. This effect does not significantly vary by whether the argument position is known.

The right facet contains our key test. Starting with the black pooled points, those who are randomized to argue for increase their support somewhat on average, while those randomized to argue against decrease their support, with an overall difference of around 10%. Consistent with our main hypothesis about motivated reasoning, comparing the blue square and orange triangle

Figure 4: Study 2, Predictors of Change in Support



indicates this difference is about 4% larger among people who learn the position at the outset than those who learn it after the information is presented.

Table 3 shows these results in our preregistered regression form. Each column reports coefficients from a regression where the dependent variable is the final support after learning all of the information, one's position, and making the argument (in percentage points). In all regressions throughout we cluster standard errors by participant.[24]

Our preregistered preferred specification is in column (6), though we start with more basic versions. Column (1) only includes whether one is randomized to argue for the proposition ("for"), and the share of information that is positive for the proposition. Receiving more positive information has a substantial effect on the final support: going from no positive to all positive information increases the average final posterior by 12%. The effect of being randomized to argue for the proposition is somewhat smaller at 8%, which is still substantial and significant ($p < .01$).[25]

Column (2) completes our key 2×2 randomization by adding an interaction term between

[24]Our preregistration neglected to specify how we would cluster standard error, and clustering by participant is a standard choice so we opt for that in the main text. Tables A14 and A21 in the Appendix show that alternative ways of computing standard errors do not change our key results.

[25]In addition to the somewhat different specification, one reason why the coefficient here is larger on "Share Pro" than "For" while in figure 4 the difference between a positive and negative information lean is smaller is that most participants received either $1/4$ or $3/4$ pieces of positive information (with a minority receiving $0/4$ or $1/4$ (see footnote 23) and the mostly positive vs mostly negative contrast is mostly capturing that smaller difference.

Table 3: Study 2, Predictors of Final Support

	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	44.56^{**}	46.49^{**}	-3.12^{**}			
	(1.01)	(1.22)	(0.85)			
Prior			0.84^{**}	0.85^{**}	0.83^{**}	0.84^{**}
			(0.01)	(0.01)	(0.01)	(0.01)
For	8.06^{**}	5.09^{**}	7.42^{**}	7.39^{**}	7.17^{**}	7.12^{**}
	(0.95)	(1.30)	(0.75)	(0.75)	(0.91)	(0.90)
Known		-3.74^{**}	-2.81^{**}	-2.73^{**}	-3.53^{**}	-3.42^{**}
		(1.33)	(0.85)	(0.85)	(0.96)	(0.95)
For × Known		6.00^{**}	3.84^{**}	3.85^{**}	4.69^{**}	4.70^{**}
		(1.86)	(1.07)	(1.07)	(1.30)	(1.29)
Share pro	12.39^{**}	12.23^{**}	11.57^{**}	11.61^{**}	10.28^{**}	10.34^{**}
	(1.45)	(1.45)	(0.83)	(0.83)	(0.95)	(0.95)
R^2	0.03	0.03	0.69	0.69	0.80	0.81
Fixed effects	No	No	No	Prop	Resp	Resp+Prop
N	4506	4506	4506	4506	4506	4506

Notes: Estimates from linear regressions of final posterior probability of supporting the proposition (self), in percentage points. Standard errors (clustered by participant) in parentheses.
$^{*}p < .05$, $^{**}p < .01$.

"for" and whether the argument is revealed before getting the information ("known"). Adding this term decreases the coefficient on "for" and the interaction is positive, indicating that the effect of being randomized to argue for the proposition is substantially larger for those who learn the position earlier. In columns (3)-(6) we add a control for the prior. This improves the precision of our key coefficients and leads to some change in the point estimates.[26]

Columns (4)-(6) introduce fixed effects for the proposition and/or respondent. While these changes have some effect on the magnitudes of the key coefficients, they are always positive and significant (at $p < .01$). In our (preregistered) preferred specification in column (6), which includes the prior and both fixed effects, the estimated effect of being randomized to argue in favor when it comes after the information is 7.1%, and this increases by around 4.7% when the position is revealed before the information.

Overall this analysis provides strong evidence of an order effect when introducing information and a position to argue, consistent with motivated information processing, and with past studies

[26]This is likely because, as shown in Appendix Table A11, there is some imbalance in the prior support across treatment groups, though this is not driven by differential attrition by treatment.

on order effects in arguing studies (Babcock et al., 1995; Gneezy et al., 2020; Saccardo and Serra-Garcia, 2023). Appendix Table A15 presents a similar set of regressions, but where the dependent variable (and relevant prior) are the belief about how many others would assign a greater than 50% chance to voting for the proposition. The coefficients on "for" and "for x known" are generally about half of the magnitude as for the self-evaluation, but still always positive and significant. Appendix Table A16 shows analogous regressions using logit-transformed probabilities; again the key coefficient is positive and significant. Appendix Table A17 presents regressions testing for whether the effect of the share of positive information is different depending on what one knows about the argument position. If motivated information processing dampens the effect if inconvenient information this will mean the effect of more positive information is smaller when the position is known. We generally find point estimates in this direction but they are not statistically significant.

Appendix Tables A18, A19, and A13 look for heterogeneity by partisanship, prior belief/argument length, and ballot proposition, respectively. The key coefficient is positive and ranges from around 2 to 6 across these subgroups, and most are statistically significant.

Memory. Table 4 presents our analysis of whether there is motivated memory of information. Each observation is the answer to a memory question; in Panel A the dependent variable is an indicator for being correct, and in Panel B it is an indicator for answering "don't remember." We multiply the dependent variable by 100 so we can interpret coefficients as percentages as in other tables.

Our preregistered specification is column (4), though again we show simpler specifications leading up to this. Columns (1) and (2) use whether the phrase was from a "congruent" piece of information (i.e., positive for those arguing for, negative for those arguing against), with column (2) adding question and respondent fixed effects. The motivated memory prediction in Panel A is that this coefficient should be positive, though this is not our main prediction as we argue it could be driven by differential guessing (i.e., people who are unsure tend to guess they saw congruent information). However, the coefficient is slightly negative and not significant. In Panel B the prediction is that this coefficient should be negative, which it is, albeit not significant. In columns (3) and (4) we add the interaction with the position being known when the information is seen. Our key prediction is that this coefficient should be positive in panel A and negative in Panel B. Both of these are in the predicted direction with a magnitude of 1.5%, though not significant at $p < .05$.

Column (5) adds a control for whether the truth is that the phrase was in fact seen, which post hoc analysis found was a strong predictor of how people answer. This is likely because when

Table 4: Study 2, Predictors of Memory

	(1)	(2)	(3)	(4)	(5)
Panel A: DV Correct					
(Intercept)	65.70^{**}		66.24^{**}		
	(0.63)		(0.82)		
Congruent	-0.73	-0.50	-1.52	-1.32	-0.97
	(0.68)	(0.66)	(0.92)	(0.89)	(0.83)
Known			-1.09	-1.28	-0.74
			(1.06)	(1.03)	(0.98)
Congruent × Known			1.59	1.66	0.78
			(1.33)	(1.29)	(1.18)
Seen					27.55^{**}
					(0.85)
R^2	0.00	0.24	0.00	0.24	0.31
Panel B: DV Don't Remember					
(Intercept)	20.64^{**}		20.36^{**}		
	(0.61)		(0.76)		
Congruent	-0.50	-0.56	0.18	0.15	-0.06
	(0.54)	(0.54)	(0.73)	(0.73)	(0.70)
Known			0.57	0.81	0.48
			(0.95)	(0.86)	(0.84)
Congruent × Known			-1.36	-1.42	-0.89
			(1.07)	(1.07)	(1.01)
Seen					-16.73^{**}
					(0.70)
R^2	0.00	0.29	0.00	0.29	0.33
N	18024	18024	18024	18024	18024
Fixed effects	No	Q+Resp	No	Q+Resp	Q+Resp

Notes: Estimates from linear regressions predicting a correct answer (panel A) and answering "Don't Remember" (panel B) to memory questions, in percentage points. Standard errors (clustered by participant) in parentheses.
$^{*}p < .05$, $^{**}p < .01$.

participants did not see a phrase they may struggle to know whether to respond that they did not see it or do not remember. Adding this control leads to a smaller in magnitude but more precisely estimated key coefficient in both panels.

Post Hoc Analyses. On the whole, the results on motivated memory are not as strong or clear as the order effect in the evaluations. Motivated memory may play some role in the order effect observed in our main analysis, but other mechanisms are likely at play. To explore other possible drivers for the order effect, we present two additional non-preregistered analyses. (In both cases we did preregister doing exploratory analysis along these lines, but did not preregister any specific tests.)

First, we present a visualization on when the different treatment subgroup diverge in their beliefs. Figure 5 plots the average change from the previous support at each elicitation, broken down by our key 2×2 randomization. The thick lines are for those who learn their position before the information, and thin lines for those learning the position after. Blue lines are for those who argued for, and red for those who argued against.

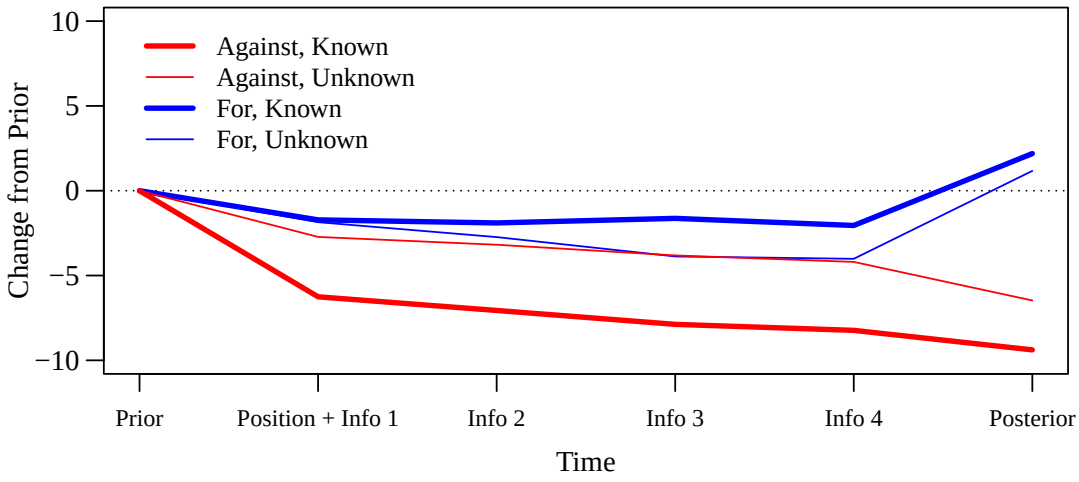
After the prior (normalized to be zero), the second elicitation happens after the "before" group learns their position, and all see one piece of information. In general, all tend to move their beliefs down, perhaps because they do not consider as many arguments against before they start to see information. This dip is largest for those who also learn they will be arguing against.

The next three elicitations ("Info 2" to "Info 4") are after the second to fourth pieces of information. In this phase we see some additional separation between those who know their position and those who do not, mostly driven by the "for" group becoming relatively more positive. Since all that is happening here is participants reading more information, we view this as suggestive evidence of selective incorporation or weighting of information which is congruent with one's arguing position.

The final elicitation follows the "after" group learning their position and everyone making their argument. Here we see a relatively parallel shift where all move in the direction of the argument they make.

By the posterior, we can visually see the key result: the gap between those arguing pro and con is larger for those who knew the position at the outset. This seems to be driven by two forces. First, as noted above, there is more separation between the groups as participants read more information, and those who know their position can selectively pay attention to and weight the information which matches their argument (particularly in the "for" group). Second, for those who learn their position early, there tends to be (relative) movement towards this position twice: when they learn

Figure 5: Dynamics of Support



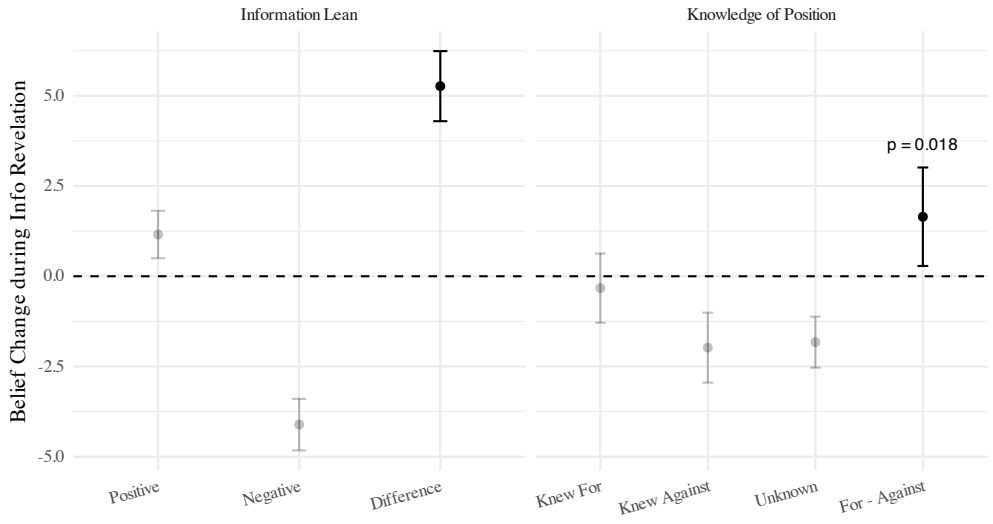
it, and when they actually make the argument. For those learning the position later, they learn the position and make the argument at the same time, but this does not seem to matter much more than just making the argument. That is, the "unbundling" of learning the position and making the argument may lead to a larger aggregate effect.

To test for the former effect – which is in line with motivated information processing – systematically, we conduct a post hoc (not preregistered) analysis of how support change from the "Info 1" elicitation to the "Info 4" elicitation. Figure 6 presents a simple visual version of the analysis.[27] Here we look at the average change in support from after the first piece of information to after the fourth. The left facet shows that, unsurprisingly, those who see more positive information tend to update in this direction. The right facet breaks the average change down by knowledge of arguing position. All groups decline on average, though this decline is largest among those who know they will argue against the position. In relative terms, those who know they will argue for the position have a more positive change in average beliefs than those who know they will argue against. This is significant at $p < .05$ with a simple difference of means test, as is the related coefficient from a regression with fixed effects and clustering presented in Appendix Table A20.

We take this as evidence that knowing the position leads to motivated incorporation of information that is in line with the argument participants know they will make. The point estimates on knew for vs knew against change indicate that among those who know their position during this phase, those in the pro group move about 2% towards believing they would vote for the proposition. Since by construction there can be no for vs against movement among the group who does

[27] See Appendix Table A20 for a similar analysis in regression form.

Figure 6: Study 2, Predictors of Support Dynamics



not know their position, we can interpret this effect as explaining about half of the $\approx 4\%$ difference in being randomized to argue for when it is known at the outset versus not.

Next, we present analysis of whether the information presented was used in the arguments.[28] Again we present a simple visual version of this analysis in the main text, with regression-based analysis in Appendix Table A23. The left facet contains a sanity check: when the phrase was actually shown to the participant, they are more likely to use it in the argument.[29] The overall contrast in the right panel contains another sanity check, showing that information congruent with the participant position is also much more likely to be used.

More subtly, the effect of the information being congruent is larger if this congruence is known at the time the information is seen; i.e., for participants who learn their position at the outset. We take this as additional suggestive evidence that people pay more attention to information which can be helpful in crafting arguments when they know it will be helpful.

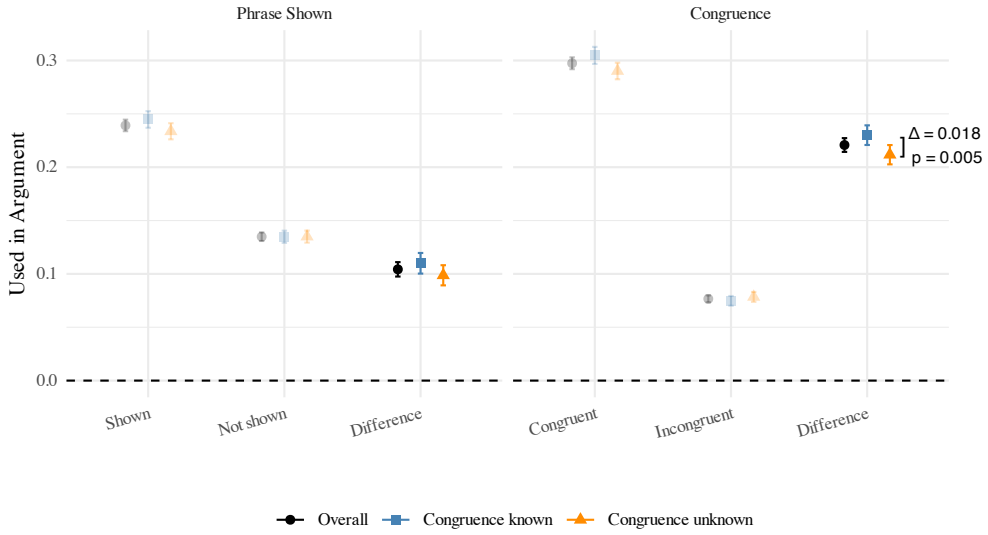
In sum, as with the analysis of belief change as information is introduced, we interpret this as suggestive evidence of some motivated information processing. Participants seem to give more attention or weight to information which they know will be useful for the argument they make. However, the ultimate effect of this additional attention on memory is modest.[30]

[28]To do this, we use the ChatGPT API to run queries of the form "Was [information] used in the text of [argument]." See Appendix C for more details about this coding.

[29]The fact that pieces of information which are not shown to participants sometimes get coded as used could mean that participants happen to use language consistent with some of the unshown information, or measurement error.

[30]Appendix Table A24 shows that phrases from pieces of information which are used in arguments are moderately

Figure 7: Study 2, Predictors of Argument Content



6 Discussion

The contrast between our two studies suggests that partisanship does less to activate motivated information processing than argument. However, we certainly do not mean to imply that partisanship cannot play a role in the relationship between arguing and motivated reasoning. In our setting we isolate the effect of arguing by randomly assigning positions, but in reality people will tend to sort into arguing for the positions they want, often those in line with their partisan attachments. While we do not have direct evidence to this effect, we suspect that if people were randomized to argue with whatever position they like versus not, this may activate patterns consistent with partisan motivated reasoning. Our results also suggest that one way to decrease polarization relative to this natural setting would be to encourage people to not only listen to arguments that go against their beliefs, but to actually argue against their priors.

Two other limitations of our analysis are that we restrict to online samples and focus on ballot propositions that are not partisan-coded at the outset. The former issue could be addressed in future work. The fact that we rely on relatively neutral policies is important for our design but harder to address given our approach. If we picked policies that already have a strong partisan affiliation (e.g., abortion, immigration, taxation), it would be harder to find endorsements from both parties, and participants would likely respond to these cues less. Still, it may be possible to come up with variants of our information-order-based design that could work on a wider range of issues.

more likely to be recalled correctly in the memory stage, and less likely to elicit a "don't remember" response.

As a concluding thought, it is noteworthy that most of the prominent work claiming belief biases undermine democratic performance focuses on ordinary citizens (e.g., Taber and Lodge, 2006; Bullock and Lenz, 2019; Achen and Bartels, 2017) rather than political professionals (politicians, staffers, journalists, etc.). And most theories that allow for someone to have biased beliefs typically point to voters as well (e.g., Ashworth and De Mesquita, 2014; Little, Schnakenberg and Turner, 2022; Callander and Carbajal, 2022). This may be for justifiable theoretical reasons – e.g., ordinary citizens have smaller incentives to form accurate beliefs about politics – or just due to the relative ease of getting samples of the general population. However, there may be reasons to think it is the professionals who are most susceptible to some belief biases, including motivated information processing. To use the common terminology from Kunda (1990), even if elites have more accuracy motives, they also plausibly have more directional motives, as they have more reason to care about the political world.

The results here suggest another reason to think political elites who spend much arguing for their positions may be particularly apt to succumb to motivated information processing. Future empirical work could test this conjecture more formally, and future theoretical work should pay more attention to the possibility that voters are the relatively rational ones when it comes to politics.

References

Achen, Christopher H. and Larry M. Bartels. 2017. *Democracy for Realists: Why Elections do Not Produce Responsive Government*. Princeton, NJ: Princeton University Press.

Ashworth, Scott and Ethan Bueno De Mesquita. 2014. "Is Voter Competence Good for Voters?: Information, Rationality, and Democratic Performance." *American Political Science Review* 108(3):565–587.

Babcock, Linda and George Loewenstein. 1997. "Explaining Bargaining Impasse: The Role of Self-Serving Biases." *Journal of Economic Perspectives* 11(1):109–126.

Babcock, Linda, George Loewenstein, Samuel Issacharoff and Colin Camerer. 1995. "Biased Judgments of Fairness in Bargaining." *American Economic Review* 85(5):1337–1343.

Bénabou, Roland and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *Quarterly Journal of Economics* 117(3):871–915.

Bénabou, Roland and Jean Tirole. 2016. "Mindful Economics: The Production, Consumption, and Value of Beliefs." *Journal of Economic Perspectives* 30(3):141–64.

Benjamin, Daniel J. 2019. "Errors in Probabilistic Reasoning and Judgment Biases." *Handbook of Behavioral Economics: Applications and Foundations 1* 2:69–186.

Blackwell, David and Lester Dubins. 1962. "Merging of Opinions with Increasing Information." *Annals of Mathematical Statistics* 33(3):882–886.

Bransford, John D and Marcia K Johnson. 1972. "Contextual Prerequisites for Understanding: Some Investigations of Comprehension and Recall." *Journal of Verbal Learning and Verbal Behavior* 11(6):717–726.

Brundage, Matt, Andrew T Little and Soosun You. 2024. "Selection Neglect and Political Beliefs." *Annual Review of Political Science* 27(1):63–85.

Bullock, John G and Gabriel Lenz. 2019. "Partisan Bias in Surveys." *Annual Review of Political Science* 22(1):325–342.

Callander, Steven and Juan Carlos Carbajal. 2022. "Cause and Effect in Political Polarization: A Dynamic Analysis." *Journal of Political Economy* 130(4):825–880.

Carlson, Ryan W, Michel André Maréchal, Bastiaan Oud, Ernst Fehr and Molly J Crockett. 2020. "Motivated Misremembering of Selfish Decisions." *Nature Communications* 11(1):1–11.

Chen, Si and Carl Heese. Forthcoming. "Fishing for Good News: Motivated Information Acquisition." *Journal of Political Economy: Microeconomics* .

Cheng, Haw and Alice Hsiaw. 2022. "Distrust in Experts and the Origins of Disagreement." *Journal of Economic Theory* 200:105401.

Chew, Soo Hong, Wei Huang and Xiaojian Zhao. 2020. "Motivated False Memory." *Journal of Political Economy* 128(10):3913–3939.

DeGroot, Morris H. 2005. *Optimal statistical decisions*. John Wiley & Sons.

Ditto, Peter H, Jared B Celniker, Shiri Spitz Siddiqi, Mertcan Güngör and Daniel P Relihan. 2025. "Partisan Bias in Political Judgment." *Annual Review of Psychology* 76.

Druckman, James N and Mary C McGrath. 2019. "The Evidence for Motivated Reasoning in Climate Change Preference Formation." *Nature Climate Change* 9(2):111–119.

Eil, David and Justin M Rao. 2011. "The Good News-Bad News Effect: Asymmetric Processing of Objective Information About Yourself." *American Economic Journal: Microeconomics* 3(2):114–38.

Enke, Benjamin. 2020. "What You See Is All There Is." *Quarterly Journal of Economics* 135(3):1363–1398.

Enke, Benjamin, Ricardo Rodriguez-Padilla and Florian Zimmermann. 2022. "Moral Universalism: Measurement and Economic Relevance." *Management Science* 68(5):3590–3603.

Enke, Benjamin, Ricardo Rodríguez-Padilla and Florian Zimmermann. 2023. "Moral Universalism and the Structure of Ideology." *Review of Economic Studies* 90(4):1934–1962.

Exley, Christine L and Judd B Kessler. 2024. "Motivated Errors." *American Economic Review* 114(4):961–987.

Galef, Julia. 2021. *The Scout Mindset: Why Some People See Things Clearly and Others Don't*. Penguin.

Gneezy, Uri, Silvia Saccardo, Marta Serra-Garcia and Roel van Veldhuizen. 2020. "Bribing the Self." *Games and Economic Behavior* 120:311–324.

Gödker, Katrin, Peiran Jiao and Paul Smeets. 2025. "Investor Memory." *Review of Financial Studies* 38(6):1595–1640.

Grether, David M. 1980. "Bayes Rule as a Descriptive Model: The Representativeness Heuristic." *Quarterly Journal of Economics* 95(3):537–557.

Grossman, Zachary and Joel J Van Der Weele. 2017. "Self-Image and Willful Ignorance in Social Decisions." *Journal of the European Economic Association* 15(1):173–217.

Hagenbach, Jeanne and Charlotte Saucet. 2025. "Motivated Skepticism." *Review of Economic Studies* 92(3):1882–1919.

Hill, Seth J. 2017. "Learning Together Slowly: Bayesian Learning About Political Facts." *Journal of Politics* 79(4):1403–1418.

Huddy, Leonie, Lilliana Mason and Lene Aarøe. 2015. "Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity." *American Political Science Review* 109(1):1–17.

Je, Hyundam and Sora Youn. 2024. "Timing of Informativeness on Motivated Reasoning." Working Paper.

Kahan, Dan M. 2015. "The Politically Motivated Reasoning Paradigm." *Emerging Trends in Social & Behavioral Sciences, Forthcoming* .

Koçak, Korhan. 2018. "Sequential Updating: A Behavioral Model of Belief Change." Working Paper.

Koehler, Jonathan J. 1993. "The Influence of Prior Beliefs on Scientific Judgments of Evidence Quality." *Organizational Behavior and Human Decision Processes* 56(1):28–55.

Kunda, Ziva. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* 108(3):480.

Lenz, Gabriel S. 2009. "Learning and Opinion Change, Not Priming: Reconsidering the Priming Hypothesis." *American Journal of Political Science* 53(4):821–837.

Lilley, Matthew and Brian Wheaton. 2024. "Are Preconceptions Postconceptions? Evidence on Motivated Political Reasoning1." Working Paper.

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

Little, Andrew T, Keith E Schnakenberg and Ian R Turner. 2022. "Motivated Reasoning and Democratic Accountability." *American Political Science Review* 116(2):751–767.

Little, Andrew T., Melina Platas and Pia Raffler. 2023. "Limits on Learning: Selective Incorporation and Retention of Political Information." Working Paper.

Mercier, Hugo and Dan Sperber. 2011. "Why do Humans Reason? Arguments for an Argumentative Theory." *Behavioral and Brain Sciences* 34(2):57–74.

Möbius, Markus M, Muriel Niederle, Paul Niehaus and Tanya S Rosenblat. 2022. "Managing Self-Confidence: Theory and Experimental Evidence." *Management Science* 68(11):7793–7817.

Murdock Jr, Bennet B. 1962. "The Serial Position Effect of Free Recall." *Journal of Experimental Psychology* 64(5):482.

Pace, Davide D, Taisuke Imai, Peter Schwardmann and Joël J van der Weele. 2025. "Uncertainty About Carbon Impact and the Willingness to Avoid CO2 Emissions." *Ecological Economics* 227:108401.

Rabin, Matthew and Joel L Schrag. 1999. "First Impressions Matter: A Model of Confirmatory Bias." *Quarterly Journal of Economics* 114(1):37–82.

Saccardo, Silvia and Marta Serra-Garcia. 2023. "Enabling or Limiting Cognitive Flexibility? Evidence of Demand for Moral Commitment." *American Economic Review* 113(2):396–429.

Schwardmann, Peter, Egon Tripodi and Joël J Van der Weele. 2022. "Self-Persuasion: Evidence From Field Experiments at Two International Debating Competitions." *American Economic Review* 112(4):1118—-46.

Schwardmann, Peter and Joel Van der Weele. 2019. "Deception and Self-Deception." *Nature Human Behaviour* 3(10):1055–1061.

Taber, Charles S and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50(3):755–769.

Tappin, Ben M, Gordon Pennycook and David G Rand. 2020. "Thinking Clearly About Causal Inferences of Politically Motivated Reasoning: Why Paradigmatic Study Designs Often Undermine Causal Inference." *Current Opinion in Behavioral Sciences* 34:81–87.

Thaler, Michael. 2024. "The Fake News Effect: Experimentally Identifying Motivated Reasoning Using Trust in News." *American Economic Journal: Microeconomics* 16(2):1–38.

Zhang, Yunhao and David G Rand. 2025. "Self-Persuasion Does not Imply Self-Deception." *Cognition* 263:106215.

Zimmermann, Florian. 2020. "The Dynamics of Motivated Beliefs." *American Economic Review* 110(2):337–61.

A Additional Theoretical Analysis and Proofs

Proof of Proposition 1. We can rewrite equation 2 as:

$$\Delta_c(f = p, s_o) - \Delta_c(f = o, s_o) = \log(l(s_o))(d_r(1, s_o) - d_r(0, s_o)) \tag{4}$$

where

$$d_r(s_p, s_o) = \mathcal{P}(a_o|f = p, s_p, s_o) - \mathcal{P}(a_o|f = o, s_p, s_o) \tag{5}$$

is the relative likelihood of the other fact being remembered when party is revealed first vs second, as a function of the signals. When $s_o = 1$ (positive fact), $\log(l(s_o)) > 0$ and by assumption 1 $d_r(1, 1) > 0$ and $d_r(0, 1) < 0$, hence the expression is positive. When $s_o = 0$ (negative fact), $\log(l(s_o)) < 0$ and by assumption 1 $d_r(1, 0) < 0$ and $d_r(0, 0) > 0$, and so again the expression is positive. For part ii, the average effect is:

$$\sum_{s_o \in \{0,1\}} \mathcal{P}(s_o)[\Delta_c(f = p, s_o) - \Delta_c(f = o, s_o)] > 0$$

∎

Order Effects for Other Information. Now we compare the effect of observing pro argument vs a con argument if observed first versus second. When the endorsement is copartisan, observing the party first will increase the effect. However, when the endorsement is out-partisan, the pro argument will have a less positive effect when the party endorsement comes first. Depending on which effect dominates, the average effect of the argument may be larger or smaller when revealed first. Formally, let the effect of seeing a pro vs con argument as a function of the order and endorsement information be:

$$\begin{aligned}\Delta_o(f, s_p) &= E[\Lambda(\mathcal{P}(\omega = 1|s_o = 1, s_p, f))] - E[\Lambda(\mathcal{P}(\omega = 1|s_o = 0, s_p, f))] \\ &= \mathcal{P}(a_o = 1|f, s_o = 1, s_p) \log(l(s_o = 1)) - \mathcal{P}(a_o = 1|f, s_o = 0, s_p) \log(l(s_o = 0))\end{aligned}$$

The difference in seeing the pro vs con argument when party comes first vs second is then:

$$\begin{aligned}\beta_{o1}(s_p) &= \Delta_o(f = p, s_p) - \Delta_o(f = o, s_p) \\ &= d_r(s_p, 1) \log(l(s_o = 1)) - d_r(s_p, 0) \log(l(s_o = 0))\end{aligned}$$

where $d_r(s_p, s_o)$ is as defined in equation 5. (Recall this is the difference in remembering if party is revealed first vs second.) This has the following properties:

Proposition 2. *The difference in the effect of receiving a pro vs con argument when partisan information is revealed first versus second can be positive or negative, for both copartisan and out-party endorsements, and averaging across both.*

Proof For $s_p = 1$, $d_r(1,1) > 0$ and $d_r(1,0) < 0$. So the first term of $\beta_{o1}(1)$ is positive ($d_r(1,1)\log(l(s_o = 1)) > 0$), but the second term is negative ($-d_r(1,0)\log(l(s_o = 0)) < 0$).

Similarly, for $s_p = 0$, $d_r(0,1) < 0$ and $d_r(0,0) > 0$, and so $\beta_{o1}(0)$ can also be positive or negative.

Further, a weigthed average of these two terms can be positive or negative. ∎

To see the intuition, consider a copartisan endorsement. If this is known before seeing the argument, it will increase the effect of a pro argument, but also mute the effect of a con argument. So, the overall difference can be positive or negative depending on which effect dominates.

Correlated Information In the main analysis we assume that the effect of the party and other information is independent. This may not be true if, for example, participants interpret the other information in a different fashion when they believe the policy is supported by their own party. To allow for this possibility, we can write the logit-transformed posterior when remembering all information as:

$$\Lambda(\mathcal{P}(\omega = 1|s_o, s_p)) = \Lambda(p) + \log(l(s_o)) + \log(l(s_p)) + m(s_p, s_o) \tag{6}$$

where $l(s_j)$ is the marginal likelihood ratio of piece of information s_j not accounting for s_{-j}. The m term then captures any change to the posterior belief with both pieces of information relative to "additive" case. If $m(s_p, s_o) = 0$ for all signals, this is the independent case. If, for example, $m(1,1) > 0$, this would mean that the effect of receiving a signal of 1 on both dimensions is larger than the sum of the individual effects of the pieces of information.

If we maintain the assumption that party information is always remembered, then the change

in the effect of a copartisan endorsement as a function of when it is learned becomes:

$$\begin{aligned}\Delta_c(f, s_o) &= E[\Lambda(\mathcal{P}(\omega = 1|s_o, s_p = 1, f))] - E[\Lambda(\mathcal{P}(\omega = 1|s_o, s_p = 0, f))] \\ &= \underbrace{q(\log(l(s_p = 1)) - \log(l(s_p = 0)))}_{\text{direct party effect}} \\ &+ \underbrace{(\mathcal{P}(a_o|f, s_p = 1, s_o) - \mathcal{P}(a_o|f, s_p = 0, s_o))\log(l(s_o))}_{\text{differential acceptance (individual)}} \\ &+ \underbrace{(\mathcal{P}(a_o|f, s_p = 1, s_o)m(s_p = 1, s_o) - \mathcal{P}(a_o|f, s_p = 0, s_o)m(s_p = 0, s_o))}_{\text{differential acceptance (mutual)}}\end{aligned}$$

The difference in this effect when the endorsement is first versus second is then:

$$\begin{aligned}\beta_{c1}(s_o) &\equiv \Delta_c(f = p, s_o) - \Delta_c(f = o, s_o) \\ &= d_r(1, s_o)(\log(l(s_o)) + m(s_p = 1, s_o)) - d_r(0, s_o)(\log(l(s_o)) + m(s_p = 0, s_o)). \quad (7)\end{aligned}$$

A sufficient condition for this to be positive by the same argument as in the proof of Proposition 2 is:

$$\text{sign}(\log(l(s_o))) = \text{sign}(\log(l(s_o)) + m(s_p = 1, s_o)) = \text{sign}(\log(l(s_o)) + m(s_p = 0, s_o)) \quad (8)$$

In words, this states that the effect of a "positive" piece of information ($s_o = 1$) is positive regardless of the revelation of the party signal, and similarly for negative information. E.g., it may be the case that positive pieces of information have a smaller or larger impact when the party signal is in-party or out-party, but the conditional effect of the information can't reverse.

Memory of Party Information. We see less reason to expect that party endorsements would be selectively ignored or forgotten based on whether paired with a pro or con argument. (And, empirically we find that party endorsements exert a substantially larger effect than arguments.) If such recall does depend on whether the argument is known there are more terms to contend with. However, for the same reasons discussed in proposition 2, this would not necessarily change the overall impact of learning the party first in a predictable direction. For example, if party is more likely to be remembered when it is "good news", this will decrease recall if party is revealed first with a pro argument, but increase recall if party is revealed first with a con argument.

Separating Acceptance and Recall. Suppose the probability a piece of information is remembered can be split into two stages: accepting it into memory in the first place $a \in \{0, 1\}$, and then recalling it from memory $r \in \{0, 1\}$. A piece of information is remembered if and only if $a\, r = 1$. Let $n_a \in \{g, b, u\}$ correspond to whether the information is known to be good news (g), bad news (b), or unknown (u) at the initial acceptance phase, and $n_r \in \{g, b, u\}$ the analogous knowledge at the recall phase. Our key assumption is that $\mathcal{P}(a = 1|n_a = g) > \mathcal{P}(a = 1|n_a = u) > \mathcal{P}(a = 1|n_a = b)$. It may also be the case that recall depends on n_r, but our key comparisons hold this fixed. So, for example, someone who learns that the piece of news is good after the acceptance phase remembers it with probability $\mathcal{P}(a = 1|n_a = u)\mathcal{P}(r = 1|n_r = g)$, which is strictly less than the probability of remembering for someone who knew it was good knows at the acceptance phase, which is $\mathcal{P}(a = 1|n_a = g)\mathcal{P}(r = 1|n_r = g)$.

B Ballot Proposition Information

B.1 Colorado Proposition 131 (2024)

Title: Top-Four Primary and Ranked-Choice Voting Initiative

Colorado Proposition 131, was on the ballot in Colorado in 2024. Proposition 131 would have amended state election law to establish a top-four primary system, where all candidates seeking that office appear on one ballot regardless of party affiliation or non-affiliation, for elections including the U.S. Senate, U.S. House of Representatives, governor, and attorney general.

Rather than having separate primaries for Democrats and Republicans, under this measure there would be one combined primary, where the top four vote-earners for each office would have advanced to the general election where ranked-choice voting (RCV) would be used to determine the winner.

Here is how the ranked-choice voting (RCV) would work in the general election. Voters can rank as many candidates as they want in order of their preference. Votes would be tallied in rounds, until one candidate has a majority of the first place votes. If no candidate has a majority, the candidate with the fewest first-place votes is eliminated before the next round. Ballots for eliminated candidates would be reassigned to the next ranked choice. This process would repeat until a candidate achieved a majority. Voters would not be required to rank all candidates.

Endorsements:

- U.S. Sen. John Hickenlooper (Dem)
- Gov. Jared Polis (Dem)
- State Representative Matt Soper (Rep)
- Former U.S. Representative Ken Buck (Rep)

Opposition:

- U.S. Representative Lauren Boebert (Rep)
- Former State Representative Dave Williams (Rep)
- U.S. Senator Michael Bennet (Dem)
- U.S. Rep. Diana DeGette (Dem)

Reasons (Positive):

- "Prop 131 is about giving voters more voice, choice and power in our elections. It's about giving us the power to vote our true preferences. ... It's about making candidates represent all of us. It's about making our leaders produce better results on the issues that we care about."

Reasons (Negative):

- "Just like the origin of this proposal, it's backed by big money. So in the open primary, what you're going to find is those who have access to those resources, and those who are interested in protecting their corporate interests, they're going to fund the candidates that best align with them."

Outcome: the measure was defeated

- 46.47 percent voted Yes
- 53.53 percent voted No

B.2 Texas Proposition 12 (2023)

Title: Abolish Galveston County Treasurer Amendment (2023)

The Texas Constitution provides that the office of county treasurer may be abolished via a constitutional amendment. The amendment abolished the Galveston County treasurer and authorized the county to employ or contract a qualified person or designate another county officer to fulfill the functions previously performed by the treasurer.

Hank Dugie, elected in 2022, was the Galveston County treasurer at the time of the election. In his 2022 campaign, he called for eliminating the office.

Endorsements: Only a couple of endorsers are mentioned for this proposition. However, it received bipartisan support at in the Texas House of Representatives and in the State Senate when it was put to a vote as a legislatively referred constitutional amendment.

- Galveston County Treasurer Hank Dugie (R)
- Voted in favour in the Senate by: (HJR 134)
  - Sen Carol Alvarado (D)
  - Sen Sarah Eckhardt (D)
  - Sen Phil King (R)
  - Sen Tan Parker (R)

Opposition: Again, only one opposer was mentioned. However, it received bipartisan criticism in the Texas House of Representatives and in the State Senate when it was put to a vote.

- Grayson County Treasurer Gayla Hawkins (R)
- Voted against in Senate by: (HJR 134)
  - Sen Pete Flores (R)
  - Sen Roland Gutierrez (D)
  - Sen Royce West (D)

Reasons (Positive):

- Galveston County Treasurer Hank Dugie (R): “I believe county treasurer positions were created back in the 1800s and since then, county government and financial technology has really evolved and improved. In Galveston County, we do not need an elected treasure to keep our money safe. We have a system of checks and balances outside of the county treasure that will be able to maintain that for the taxpayers."

Reasons (Negative):

- True Texas Project: “The current Treasurer campaigned on a promise to eliminate his position, which prompted this legislative action. Since one less government position means less government, we initially supported this amendment. However, we then heard from many conservative activists in the Galveston area who said they don’t want the position to be dissolved because there will be no more accountability to the office and it will be handed to cronies."
- Cass County Treasurer Melissa Shores (R): "The county treasurer’s office is very important to the function of the county as a whole. It is also an important part of checks and balances. It’s possible that the wrong person in office can have a negative impact, but that’s not an excuse to abolish the office."

Outcome: the initiative passed with simple majority.

- 52.9 percent voted Yes
- 47.1 percent voted No

B.3 Arizona Proposition 131 (2022)

Title: Create Office of Lieutenant Governor Amendment

Proposition 131 created the position of lieutenant governor in Arizona. Previously, Arizona was one of five states without a lieutenant governor.

Under Proposition 131, the state’s lieutenant governor is elected on a joint ticket with the governor. As of 2022, 26 states elected the governor and lieutenant governor on a joint ticket. The ballot measure required gubernatorial candidates to select running mates at least 60 days before

the general election, although the legislature is permitted to prescribe a different date. The first election for a joint governor and lieutenant governor ticket is on November 3, 2026.

Proposition 131 required that if the incumbent governor dies, resigns, or is removed from office, the lieutenant governor would succeed to the governor's office. Previously, the secretary of state succeeded to the governor's office in these situations.In Arizona's history, the secretary of state had succeeded the office of governor six times.

Endorsement: The proposition was officially endorsed by the Republican party but it received bi-partisan support in the house and senate.

- State Sen. Sean Bowie (D)
- State Sen. Javan Daniel Mesnard (R)
- Votes in favour in State Senate: (SCR 1024)
  - Sen Nancy Barto (R)
  - Sen Christine Marsh (D)

Opposition: There was no clear opposition to the measure. However, there were members from both national paries that voted against the amendment in the house and senate.

- Votes against in State Senate: (SCR 1024)
  - Sen Theresa Hatathlie (D)
  - Sen Stephanie Stahl Hamilton (D)
  - Sen Michelle Ugenti-Rita (R)
- Votes against in House of Representatives: Rep Judy M. Burges (R)

Reasons (Positive):

- Dr. Kelli Ward, chairwoman of the Republican Party of Arizona; Yvonne Cahill, secretary of the Republican Party of Arizona: "Some key points are this proposition does not expand

government nor does it create another agency since the governor will appoint this individual to the executive branch. Additionally, the continuity of the government is important. Arizona is one of only a few states nationwide without a Lt. Governor position. In the vacancy or absence of the governor, the Lt. Governor would fill the role and responsibilities. A clear succession line is important and should stay within the elected party of power from the previous election cycle."

- State Sen. J.D. Mesnard (R-17); State Sen. Sean Bowie (D-18): "Fixing this gap in Arizona's officers is a nonpartisan issue. It's about good governance. That's why Republican and Democratic legislators came together to champion Prop 131, and why it passed the legislature with broad, bipartisan support. It requires each gubernatorial nominee to select a running mate to serve as Lt. Governor, similar to how the President selects a Vice President, with both names appearing on the ballot. To ensure the new office doesn't "grow government," the law requires the Lt. Governor to occupy an existing high-level executive position within the governor's administration, such as Chief of Staff or agency director."

Reasons (Negative): There was no clear argument against the proposition since Arizona is one of the few states not to have this position. However, this issue was raised in the past in 2010 in the form of Proposition 111 when it was rejected. The opposition cited the concerns regarding the nomination rules. Hence, note that the wording of the previous proposition was different and the rules of nominating the lieutenant were of greater concern. Proposition 131 heard those concerns and made adjustments. Therefore, critiques of the two measures are not exactly comparable. The argument cited below, however, related to the creation of a new position in general which might apply to the Proposition 131 as well.

- "The name change alone is undesirable as it will allow the Lieutenant Governor to place himself/herself in the position of a Governor in waiting, much as the U.S. Vice President, tending largely to public relations and ceremonial duties. Although the many duties of Secretary of State will remain, for now, it won't be long before the position will be declared too high profile and important to be burdened with petty administrative duties, such as issuing notary certificates and registering trade names. Duties will be quickly spun off to other Departments".[31]

[31] https://apps.azsos.gov/election/2010/info/PubPamphlet/Sun_Sounds/english/prop111.htm

Outcome: The measure was approved.

- 55.16 percent voted Yes
- 44.84 percent voted No

B.4 Maine Question 1 (2021)

Title: Electric Transmission Line Restrictions and Legislative Approval Initiative

Maine Question 1, the Legislative Approval of Certain Electric Transmission Lines Initiative, was on the ballot in Maine as an indirect initiated state statute on November 2, 2021.

Question 1 was designed to stop the New England Clean Energy Connect (NECEC), a 145-mile long, high-voltage transmission line project that would transmit around 1,200 megawatts from hydroelectric plants in Quebec to electric utilities in Massachusetts and Maine. Construction of NECEC began after the project received a presidential permit on January 15, 2021. The ballot initiative prohibited the construction of high-impact electric transmission lines in the Upper Kennebec Region, retroactive to September 16, 2020, thus prohibiting Segment 1 of NECEC.[12] Segment 1 was permitted to begin construction on May 13, 2021.

The ballot initiative also required a two-thirds vote of each state legislative chamber to approve high-impact electric transmission lines. Question 1 defined high-impact electric transmission lines as those that are (a) 50 miles in length or more, (b) outside of a statutory corridor or petitioned corridor, (c) not a generator interconnection transmission facility, or (d) not constructed to primarily provide electric reliability.

Endorsements:

- State Sen. Richard Bennett (R)
- State Sen. Russell Black (R)
- State Rep. Seth Berry (D)
- State Rep. Nicole Grohoski (D)

Opposition:

- Gov. Janet T. Mills (D)
- U.S. Secretary of Energy Jennifer Granholm (D)
- Former Governor Paul LePage (R)

Reasons (Positive):

- Former Sen. Thomas Saviello (R-17): "Mainers know they're being lied to by these two foreign corporations, and they know that this project will forever change our state's character, environment and economy in ways that will not benefit us."
- State Rep. Jennifer Poirier (R-107): "... I have strongly opposed the CMP Corridor because I am concerned about what we are giving up, all for 38 permanent jobs. We are not an extension cord for Massachusetts and the CMP Corridor is a terrible deal for Maine."

Reasons (Negative):

- Dana Connors, president of the Maine State Chamber of Commerce: "It is a project that provides for access from Quebec to our New England grid, passing through Maine, that offers economic opportunity in terms of employment, tax base, catalysts for other things that depend on power to come to our state."
- Hope Pollard, president of the Maine Chapter of Associated Builders and Contractors: "We believe the November referendum jeopardizes our future. If the ballot question passes, this will likely kill the project and take away all the jobs, customer savings, tax reductions, and other benefits that go along with it. But, what's more, it puts politicians in charge of how you pay for and what electricity sources are available to Mainers. This could hamper the development of new electricity sources, including in-state wind and solar and keeps out low-cost energy solutions. ... If project opponents get their way, it also paints a bleak portrait of Maine's future — this is a project that secured all its federal, state, and local approvals, based on the laws in effect at the time. If those laws can be changed and applied retroactively, who will want to invest in Maine?"

Outcome: the ballot measure was approved

- 59.20 percent voted Yes
- 40.80 percent voted No

B.5 Arkansas Issue 2 (2020)

Title: Change State Legislative Term Limits Amendment

Arkansas Issue 2, the State Legislative Term Limits Amendment, was on the ballot in Arkansas as a legislatively referred constitutional amendment on November 3, 2020.

As of 2019, Arkansas legislators could serve up to 16 years throughout their lifetimes in the House or Senate. This measure changed term limits of state legislators to twelve consecutive years with the opportunity to return after a four-year break. The 12-year limit applied to anyone elected in 2021 or after.

Those first elected to the legislature before 2021 kept the state's existing term limit of 16 years except that they were eligible to run for election again after four years had passed under the amendment.

Endorsements:

- State Senator Alan Clark (R)
- State Representative Jim Dotson (R)
- Vote in favour in House of Representative (SJR 15):
  - State Rep Nicole Clowney (D)
  - State Rep Megan Godfrey (D)

Opposition:

- Vote against in House of Representative (SJR 15):

- Rep Stan Berry (R)
- Rep Frances Cavenaugh (R)
- Rep Fredrick J. Love (D)
- Rep Milton Nicks, Jr. (D)

Reasons (Positive):

- Arkansas State Representative and measure sponsor Jim Dotson (R-93): "The purpose of term limits is to limit power and advantages of incumbency. So if you have an incumbent who is running against someone who is not an incumbent, they obviously have a built-in advantage. After this resolution — if it is adopted and approved by the voters — is passed, after 12 years someone loses that advantage of incumbency." Dotson also said, "Really, it's trying to balance out and make it where we don't have a complete and total exodus by cutting everybody off that currently is serving, and then having a new fresh start on those subsequent years.

Reasons (Negative):

- Arkansas State Representative Vivian Flowers (D-17): "Those members who are currently serving would get to operate under the current law and serve 16 years — up to 16 years — while everyone else in the state would have to be limited to 12 years, thereby giving us in this chamber right now a definitive advantage over everybody else in the state?"

Outcome: The measure was passed

- 55.38 percent voted Yes
- 44.62 percent voted No

B.6 South Dakota Constitutional Amendment B (2020)

Title: Deadwood Sports Betting Legalization Amendment

South Dakota Constitutional Amendment B was on the ballot in South Dakota as a legislatively referred constitutional amendment on November 3, 2020. The measure amended the South Dakota Constitution to authorize the South Dakota State Legislature to legalize sports betting within the city limits of Deadwood, South Dakota. Going into the election, in Deadwood, blackjack, craps, keno, poker, roulette, and slot machines were legal. Gambling in Deadwood was legalized after the approval of citizen-initiated Amendment B of 1988. Like other authorized forms of gambling within the city, all net municipal proceeds was set to be dedicated to the historic restoration and preservation of Deadwood.

Endorsements:

- State Senator Bob Ewing (R)
- Votes in favour in State House of Representatives (Senate Joint Resolution 501)
  - State Rep Caleb Finck (R)
  - State Rep Linda Duba (D)
  - State Rep and Minority Whip Oren L. Lesmeister (D)

Opposition:
Votes against in State House of Representatives (Senate Joint Resolution 501)

- State Rep David L. Anderson (R)
- State Rep Thomas J. Brunner (R)
- State Rep Ryan Cwach (D)
- State Rep Michael P. Saba (D)

Reasons (Positive):

- Deadwood Gaming Association: "We must compete with surrounding gaming jurisdictions' offerings. Sports wagering is doing extremely well in Iowa, and it will be starting in March in Montana and in May in Colorado. Sports wagering is currently happening in South Dakota ILLEGALLY! When given the option, we believe South Dakotans want to place their sports wagers in a safe, legal and regulated environment."
- "While some do not gamble or support gaming, I have always supported gaming in Deadwood as it is permissive. No one has to participate. Whether one chooses to wager is solely their own decision... If passed will add another option for people to support and wager on sporting events. Deadwood will then be on a level playing field competing with other states that allow sports wagering. The tax dollars raised by Deadwood gaming are enjoyed by local cities, schools and the State of South Dakota... The whole state enjoys the dollars spent by tourist traveling across South Dakota on their way to Deadwood."

Reasons (Negative):

- South Dakota State Representative Steven Haugaard (R): "It could take over your life. It's not in the best interests of the state to expand any aspects of gambling.

- "Sports can already be an obsession. It shouldn't be a training ground for young people to develop a gambling addiction. The few dollars that would come from sports betting pales in comparison to the damage it causes. We are the second most gambling dependent state in the nation. We don't need to make that worse. No one should take advantage of vulnerable people"

Outcome: The measure was passed.

- 58.47 percent voted Yes
- 41.53 percent voted No

C Prompt for Information Usage Analysis

To code whether pieces of information are used in arguments (as well as argument quality for paying bonuses), we tried several queries and models with the ChatGPT API. We also hand-coded 400 argument-information pairs, 200 sampled at random and 200 of which over-sampled pairs for which an initial set of automatic codings tended to disagree. After several iterations using gpt-4o and gpt-5.1 with several levels of "reasoning," the model/query paring which achieved the highest correlation with both the hand-coding and an indicator for whether the piece of information was actually shown was gpt-5.1 with reasoning set to "none" and the following query:

```
You are a cautious human coder.
Someone may or may not have read PHRASE ('words')
and then wrote an argument ARG ('arg').
Your general goal is to figure out if
they read and paid attention to PHRASE when making ARG
In particular, determine whether they used any specific wording
or the key ideas in PHRASE when writing ARG
Also rate ARG quality (clarity+persuasiveness)
on 0-100 using anchors:
0-19: unclear/incoherent;
20-39: weak/unclear;
40-59: fair;
60-79: clear/somewhat persuasive;
80-90: very clear+persuasive; 91-100: exemplary.
Output JSON ONLY with keys:
label (0,0.5,1), confidence (0..1), explanation (<=15 words),
quality_score (integer 0..100), quality_reason (<=15 words).
Rules: 1=key idea or specific phrasing used; 0=absent; 0.5=ambiguous.
```

The correlation between this measure and the hand-coding is 0.65. This is somewhat low but likely driven by the fact that the task is somewhat ambiguous: e.g., in some cases the argument uses similar wording but does not draw on the key idea in the information, and in others a related idea is used in the argument but it is different enough to make coding challenging. Further, the correlation between this coding and whether the information was actually shown (among the subset hand-coded) is 0.16, only marginally lower than the correlation between the hand-coding and whether the information was shown (0.17).

D Additional Results

D.1 Study 1

Table A1 shows regressions predicting participants answer to a question about what share of endorsers of the proposition are Democrats. The first column reports a manipulation check which only includes an indicator for whether the participant saw endorsements from Democrats, as well as proposition fixed effects. The treatment increases this belief by around 30% (relative to seeing Republican endorsers). The second column shows a regression with the participant's own prior belief, an indicator for being a Republican, and an interaction between the two. The positive coefficient on the prior indicates that, among Democrats, those who think the proposition is a good idea at the outset tend to think most endorsers are Democrats. However, for Republicans, the effect of the prior is effectively zero. This is consistent with Democrats assuming propositions they like must be supported by other Democrats. (The lack of a relationship among Republicans may be because the question is asked in terms of the share of Democrats.)

Table A1: Study 1, Manipulation Check and Beliefs about Democratic Support

	(1)	(2)	(3)
D Treatment	30.38^{***}		30.24^{***}
	(0.77)		(0.75)
Prior		0.34^{***}	0.33^{***}
		(0.02)	(0.01)
Republican		18.70^{***}	18.33^{***}
		(1.59)	(1.37)
Prior × Republican		-0.33^{***}	-0.32^{***}
		(0.03)	(0.02)
Fixed effects	Prop	Prop	Prop
N	8622	8622	8622

Notes: Estimates from linear regressions of the estimated share of endorsers who are Democrats, in percentage points. Standard errors (clustered by participant) in parentheses. $*p < .05$, $**p < .01$.

Table A2 shows regressions predicting the assessment of the quality of the information/reason. Column (1) includes the prior belief that the participant would vote for the proposition, an indicator for the reason being for the proposition, and an interaction between the two. Because of the interaction term, the coefficient on the prior means that among those reading against reasons, those who are more favorable to the proposition at the outset rate the reason as less persuasive. The

Table A2: Study 1, Predictors of Information Quality Ratings

	(1)	(2)	(3)
Prior	-0.07^{***}		-0.06^{**}
	(0.02)		(0.02)
Positive	-28.71^{***}		-28.68^{***}
	(1.33)		(1.32)
Prior × Positive	0.55^{***}		0.55^{***}
	(0.02)		(0.02)
Know Congruent		1.06	1.17
		(0.67)	(0.63)
Know Noncongruent		-4.97^{***}	-4.81^{***}
		(0.68)	(0.63)
Fixed effects	Prop	Prop	Prop
N	8622	8622	8622

Notes: Estimates from linear regressions of the assessment of argument quality on a scale from 0 to 100. Standard errors (clustered by participant) in parentheses.
* $p < .05$, ** $p < .01$.

interaction term indicates that this flips for pro reasons: those reading these are more likely to rate them as persuasive if they generally agree at the outset.

Column two contains two indicators for whether they know the argument was "congruent" or not with their party endorsement. That is, for those who receive a copartisan endorsement, pro arguments are coded as congruent and against arguments are coded as incongruent. The base category here is participants who see the argument before the endorsement, hence congruence is unknown when they answer this question. The incongruent coefficient shows that learning an outparty endorsement decreases the assessment of argument quality by nearly 5%, a modest but highly significant effect. The second coefficient shows that receiving a copartisan endorsement leads to an increase of argument quality by around 1%, though this is not significant. Contrasting the two, among those who know the congruence of the argument, those who know it is congruent rate the argument as around 6% more persuasive than those who know it is incongruent.

Table A3 shows the main specification from table 1 with alternative ways of computing standard errors. Column 1 replicates the main text version with clustering by participant (also including p values). Column 2 has no clustering and column 3 reports heteroskedasticity-consistent standard errors. Across all versions, the main effects ("positive" and "copartisan") are statistically significant at $p < .05$, while the key interaction term ("Copartisan x End first") is not.

Table A3: Study 1, Predictors of Final Support, Alternative Standard Errors Clustering

	(1)	(2)	(3)
Prior	0.70^{***}	0.70^{***}	0.70^{***}
	(0.01)	(0.01)	(0.01)
	0.000	0.000	0.000
Positive	4.04^{***}	4.04^{***}	4.04^{***}
	(0.65)	(0.64)	(0.64)
	0.000	0.000	0.000
End first	0.18	0.18	0.18
	(0.80)	(0.78)	(0.81)
	0.822	0.818	0.824
Positive × End first	0.68	0.68	0.68
	(0.92)	(0.90)	(0.90)
	0.460	0.452	0.452
Copartisan	12.49^{***}	12.49^{***}	12.49^{***}
	(0.74)	(0.64)	(0.66)
	0.000	0.000	0.000
Copartisan × End first	-0.78	-0.78	-0.78
	(0.92)	(0.91)	(0.91)
	0.397	0.388	0.392
Fixed effects	Prop + Part	Prop + Part	Prop + Part
Clustering	Participant	IID	HC
N	8622	8622	8622

Notes: Estimates from linear regressions of the assessment of argument quality on a scale from 0 to 100. Standard errors in parenthesis, and p values below.

Table A4 shows our key specification subsetted to each ballot proposition. The coefficients on receiving positive information and a copartisan endorsement are broadly similar and significant for most propositions. The coefficient on our key interaction term (Copartisan × End First) is negative for 5/6 propositions, and never statistically significant.

Table A5 presents a version of Table 1 where the dependent variable is the logit transformation of the probability of voting for the proposition. (We treat those answering 0 as a probability of .005 and those answering 1 as .995.) The general conclusions are the same as the linear version: being randomized to see a "for" argument and copartisan endorsement both increase the probability of saying one would vote for the proposition, but the order does not matter.

Table A6 presents a similar table but where the outcome and prior are the probability that another member of the same party would vote for the proposition. The main difference here is that there is a larger effect of a copartisan endorsement, and a slightly smaller coefficient on receiving

Table A4: Study 1, Predictors of Final Support, by Proposition

	AR2	AZ131	CO131	ME1	SDB	TX12
Prior (self)	0.70^{***}	0.75^{***}	0.78^{***}	0.62^{***}	0.83^{***}	0.78^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Positive	4.30^{**}	3.56^{**}	6.40^{***}	2.81	5.54^{***}	2.25
	(1.51)	(1.27)	(1.28)	(1.56)	(1.31)	(1.49)
Copartisan	11.15^{***}	12.72^{***}	12.89^{***}	12.88^{***}	7.42^{***}	11.95^{***}
	(1.49)	(1.26)	(1.28)	(1.59)	(1.33)	(1.48)
End first	-2.42	1.06	-0.23	4.15^{*}	-0.62	-1.40
	(1.91)	(1.68)	(1.73)	(2.12)	(1.78)	(1.71)
Positive $\times$ End first	2.49	0.22	1.64	-4.04	1.37	4.04^{*}
	(2.09)	(1.79)	(1.86)	(2.24)	(1.84)	(2.02)
Copartisan $\times$ End first	1.97	-2.32	-0.55	-1.82	-2.43	-1.47
	(2.10)	(1.79)	(1.86)	(2.24)	(1.84)	(2.02)
Fixed effects	None	None	None	None	None	None
N	1521	1521	1521	1521	1521	1521

Notes: Estimates from linear regressions of the final posterior probability of supporting the proposition (self). Standard errors in parentheses.
* $p < .05$, ** $p < .01$.

a for argument. The key interaction term remains negative and not statistically significant.

In Tables A7, we run our main preregistered specification for the final support evaluation on various subgroups. The first two columns restrict to Democrats and Republicans. One interesting difference here is that Democrats respond to both of our key treatments more strongly, particularly a copartisan endorsement. However, the key interaction term between a copartisan endorsement and this coming first is still negative. Next we subset to people who score above the median social partisanship score from Huddy, Mason and Aarøe (2015). The main difference here is that those high on this scale respond less to the reasons (though interestingly, the coefficients on the copartisan endorsement are very close). Again, the key interactive coefficient is negative and not signifncant. Finally, we subset to those who start with a prior between 20 and 80 ("Mod. Prior") and those outside this range ("Ext Prior"). There are no major differences between these groups, including on our key interaction term.

Table A8 reports the key recall regressions with alternative clustering of the standard errors, first for correct recall and then for answering "Don't remember." The key coefficients remain insignificant.

Next we analyze the recall questions about the party and jobs of people who endorsed or op-

Table A5: Study 1, Predictors of Final Support, Logit Scale

	(1)	(2)	(3)	(4)	(5)
Prior (self, logit)		0.78^{***}	0.78^{***}	0.77^{***}	0.73^{***}
		(0.01)	(0.01)	(0.01)	(0.01)
Positive	0.39^{***}	0.34^{***}	0.30^{***}	0.30^{***}	0.28^{***}
	(0.05)	(0.03)	(0.05)	(0.05)	(0.05)
Copartisan	0.82^{***}	0.77^{***}	0.77^{***}	0.78^{***}	0.82^{***}
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)
End first			-0.04	-0.04	-0.03
			(0.06)	(0.06)	(0.06)
Positive × End first			0.08	0.09	0.06
			(0.06)	(0.06)	(0.07)
Copartisan × End first			0.00	-0.01	-0.01
			(0.06)	(0.06)	(0.07)
Fixed effects	No	No	No	Prop	Prop+Part
N	8622	8622	8622	8622	8622

Notes: Estimates from linear regressions of the final posterior probability of supporting the proposition (self), logit transformed. Prior is logit transformed as well. Standard errors (clustered by respondent) in parentheses
. * $p < .05$, ** $p < .01$.

posed the proposition. Here we replace *Endorsements First* with *Reason First*. Our pre-registered motivated-memory prediction is that reason-congruent endorsements are better remembered when the reason appears first. However, this test is less clean for the party-recall question, since participants who come to support a proposition may infer that their own party endorsed it.

Tables A9 and A10 present the results. The only significant coefficient is that individuals are more likely to correctly recall the party when the truth is that the endorser is a copartisan.

Table A6: Study 1, Predictors of Final Belief about Others' Support

	(1)	(2)	(3)	(4)	(5)
Prior (other)		0.57^{***}	0.57^{***}	0.57^{***}	0.47^{***}
		(0.01)	(0.01)	(0.01)	(0.01)
Positive	3.99^{***}	3.54^{***}	3.03^{***}	3.03^{***}	2.87^{***}
	(0.55)	(0.45)	(0.64)	(0.63)	(0.66)
Copartisan	20.40^{***}	20.16^{***}	20.89^{***}	20.91^{***}	21.33^{***}
	(0.70)	(0.62)	(0.76)	(0.76)	(0.80)
End first			0.53	0.46	0.56
			(0.79)	(0.79)	(0.78)
Positive × End first			1.02	1.07	0.85
			(0.87)	(0.87)	(0.92)
Copartisan × End first			−1.46	−1.48	−1.29
			(0.86)	(0.85)	(0.91)
Fixed effects	No	No	No	Prop	Prop+Part
N	8622	8622	8622	8622	8622

Notes: Estimates from linear regressions of the final posterior probability of another respondent supporting the proposition (self).. Standard errors (clustered by respondent) in parentheses.
* $p < .05$, ** $p < .01$.

Table A7: Study 1, Predictors of Final Support, Heterogeneity

	Dem	Rep	H Partisan	L Partisan	Mod. Prior	Ext. Prior
Prior (self)	0.72^{***}	0.65^{***}	0.67^{***}	0.71^{***}	0.63^{***}	0.74^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Positive	4.65^{***}	3.02^{***}	1.55	5.77^{***}	3.76^{***}	4.30^{**}
	(0.87)	(0.91)	(0.95)	(0.85)	(0.72)	(1.42)
Copartisan	15.50^{***}	8.25^{***}	12.25^{***}	11.85^{***}	10.41^{***}	15.03^{***}
	(1.03)	(0.97)	(1.13)	(0.93)	(0.82)	(1.56)
End first	0.06	0.18	−0.10	0.58	0.21	1.06
	(1.08)	(1.08)	(1.16)	(1.04)	(0.90)	(1.81)
Positive × End first	1.15	0.40	1.10	0.58	−0.10	−0.60
	(1.24)	(1.25)	(1.32)	(1.21)	(1.05)	(1.94)
Copartisan × End first	−1.24	−0.39	−1.02	−1.02	−0.19	−1.95
	(1.24)	(1.26)	(1.36)	(1.18)	(1.06)	(1.98)
Fixed effects	Prop+Part	Prop+Part	Prop+Part	Prop+Part	Prop+Part	Prop+Part
N	4548	4578	4116	4968	6038	3088

Notes: Estimates from linear regressions of the final posterior probability of supporting the proposition (self). Standard errors (clustered by respondent) in parentheses.
*$p < .05$, ** $p < .01$.

Table A8: Study 1, Predictors of Recall, Alternative Standard Errors Clustering

	Correct Recall			Don't Remember		
	(1)	(2)	(3)	(4)	(5)	(6)
Party Congruent × E First	−0.49	−0.49	−0.49	1.10	1.10	1.10
	(2.90)	(2.89)	(2.89)	(2.32)	(2.31)	(2.30)
	0.867	0.866	0.866	0.635	0.633	0.631
Party Congruent	−3.62	−3.62	−3.62	−0.47	−0.47	−0.47
	(2.09)	(2.05)	(2.07)	(1.61)	(1.64)	(1.62)
	0.084	0.077	0.081	0.773	0.776	0.774
E First	1.84	1.84	1.84	−1.08	−1.08	−1.08
	(2.06)	(2.07)	(2.04)	(1.62)	(1.65)	(1.63)
	0.372	0.376	0.369	0.507	0.515	0.509
Self Congruent	1.20	1.20	1.20	2.35	2.35	2.35
	(1.52)	(1.52)	(1.52)	(1.23)	(1.21)	(1.22)
	0.429	0.430	0.431	0.055	0.053	0.055
Fixed effects	Q + Part	Q + Part	Q + Part	Q + Part	Q + Part	Q + Part
Clustering	Participant	IID	HC	Participant	IID	HC
N	4309	4309	4309	4309	4309	4309

Notes: Estimates from linear regressions of the assessment of argument quality on a scale from 0 to 100. Standard errors in parenthesis, and p values below.

Table A9: Predictors of Recall, Endorsers' Party

	(1) Correct	(2) Correct	(3) Correct	(4) DR	(5) DR	(6) DR
Party Congruent x R First		0.39	2.51		0.61	−1.05
		(3.05)	(2.80)		(2.97)	(2.30)
Party Congruent	0.18	−0.01	0.16	0.90	0.58	−1.54
	(1.50)	(2.11)	(1.91)	(1.50)	(2.05)	(1.62)
R First		−0.98	−3.96		1.16	3.40*
		(2.19)	(2.02)		(2.10)	(1.62)
Self Congruent	4.50**	4.50**	2.78*	−0.84	−0.82	−0.24
	(1.50)	(1.50)	(1.37)	(1.44)	(1.45)	(1.14)
Fixed effects	No FE	No FE	Q + Part	No FE	No FE	Q + Part
N	4304	4304	4304	4304	4304	4304

Notes: Estimates from linear regressions of indicators for the answer to recall questions being correct or "don't remember," multiplied by 100. Standard errors (clustered by respondent) in parentheses.
* $p < .05$, ** $p < .01$.

Table A10: Predictors of Recall, Endorsers' Job

	(1) Correct	(2) Correct	(3) Correct	(4) DR	(5) DR	(6) DR
Party Congruent x R First		1.93	1.91		−0.06	−0.83
		(2.81)	(2.71)		(3.09)	(2.47)
Party Congruent	−0.73	−1.71	−0.42	1.11	1.14	−0.82
	(1.38)	(1.95)	(1.87)	(1.53)	(2.19)	(1.74)
R First		−0.23	−1.66		−0.11	2.03
		(2.06)	(1.99)		(2.25)	(1.79)
Self Congruent	2.05	2.07	−0.20	−2.98	−2.98	−1.18
	(1.48)	(1.48)	(1.39)	(1.53)	(1.53)	(1.24)
Fixed effects	No FE	No FE	Q + Part	No FE	No FE	Q + Part
N	4304	4304	4304	4304	4304	4304

Notes: Estimates from linear regressions of indicators for the answer to recall questions being correct or "don't remember," multiplied by 100. Standard errors (clustered by respondent) in parentheses.
* $p < .05$, ** $p < .01$.

D.2 Study 2

We first present results of placebo tests predicting the prior support with our key treatments, which only affect what participants see after reporting this prior. Table A11 shows regressions predicting the prior support. Column 1 shows that those who were randomized to argue for the proposition started with a somewhat lower prior on average, though this difference is not significant. Columns 2 shows very little imbalance in whether the position was known at the outset. Columns 3 shows some imbalance across the 2 $\times$ 2 randomization. In particular, among those who do not know the position at the outset ("Known" is 1), those assigned to argue for have a modestly lower prior, which is statistically significant. This would be potentially problematic if it was driven by differential attrition, e.g., if people who know they are randomized to argue against their own prior are more likely to drop out. Column 4 shows this is not the case: when we include every participant-proposition where a prior was reported (before treatment status), including those who eventually dropped out of the survey, the coefficients are nearly identical. Table A12 shows the analogous test for the prior belief that others would support the proposition, which does not exhibit any significant imbalance.

Table A13 shows the key specification for the final evaluation subsetted by proposition. The key coefficient (For $\times$ Known) is positive for all three and of a similar magnitude, though the statistical significance varies by proposition.

Table A14 shows the results of our main specification with alternative different ways of com-

Table A11: Study 2, Predictors of Prior Support

	(1)	(2)	(3)	(4)
Intercept	58.65^{***}	57.77^{***}	59.20^{***}	59.26^{***}
	(0.64)	(0.63)	(0.90)	(0.89)
For	-1.50		-2.78^{*}	-2.69^{*}
	(0.89)		(1.26)	(1.23)
Known		0.21	-1.12	-1.34
		(0.89)	(1.28)	(1.25)
For × Known			2.58	2.54
			(1.79)	(1.75)
R^2	0.00	0.00	0.00	0.00
Sample	Used	Used	Used	All
N	4506	4506	4506	4684

Notes: Estimates from linear regressions of prior probability of supporting the proposition (self), in percentage points. Standard errors (clustered by respondent) in parentheses.
* $p < .05$, ** $p < .01$.

Table A12: Study 2, Predictors of Prior Beliefs about Others' Support

	(1)	(2)	(3)	(4)
Intercept	60.05^{***}	60.27^{***}	60.54^{***}	60.48^{***}
	(0.42)	(0.41)	(0.59)	(0.58)
For	0.09		-0.53	-0.47
	(0.58)		(0.82)	(0.81)
Known		-0.34	-0.98	-0.99
		(0.58)	(0.84)	(0.82)
For × Known			1.25	1.17
			(1.17)	(1.14)
R^2	0.00	0.00	0.00	0.00
Sample	Used	Used	Used	All
N	4506	4506	4506	4684

Notes: Estimates from linear regressions of prior probability of supporting the proposition (self), in percentage points. Standard errors (clustered by respondent) in parentheses.
* $p < .05$, ** $p < .01$.

Table A13: Study 2, Predictors of Final Support by Proposition

	ar2	co131	sdb
Prior	0.81^{***}	0.85^{***}	0.89^{***}
	(0.02)	(0.01)	(0.01)
For	8.77^{***}	5.34^{***}	8.06^{***}
	(1.27)	(1.23)	(1.22)
Known	-3.35^{*}	-3.23^{*}	-2.10
	(1.44)	(1.35)	(1.39)
For $\times$ Known	4.22^{*}	5.26^{**}	2.59
	(1.88)	(1.71)	(1.79)
Share pro	3.60^{*}	12.08^{***}	18.98^{***}
	(1.49)	(1.33)	(1.47)
Fixed effects	None	None	None
N	1502	1502	1502

Notes: Estimates from linear regressions of final support by proposition, in percentage points. Standard errors in parentheses.
* $p < .05$, ** $p < .01$.

puting standard errors. The first column is the version in the main text which clusters by participant. The second column does no clustering, and the third uses heteroskedasticity-consistent standard errors. The standard errors slightly change but the key coefficient ("For x Known") remains significant at $p < .001$.

Table A15 contains our main alternative analysis, using the posterior reported share of others who would support the proposition (i.e., give themselves at least a 50% chance of voting). This measure was incentivized, and also plausibly tests whether the key randomization affect genuine beliefs about how appealing the proposition is. Across specifications, we consistently find positive and nearly always significant coefficients on "For" and "For x Known," albeit about half of the magnitude of the self-evaluations.

Table A14: Study 2, Predictors of Final Support, Alternative Standard Errors Clustering

	(1)	(2)	(3)
Prior	0.84^{***}	0.84^{***}	0.84^{***}
	(0.01)	(0.01)	(0.01)
	0.000	0.000	0.000
For	7.12^{***}	7.12^{***}	7.12^{***}
	(0.90)	(0.87)	(0.89)
	0.000	0.000	0.000
Known	-3.42^{***}	-3.42^{***}	-3.42^{***}
	(0.95)	(0.89)	(0.93)
	0.000	0.000	0.000
For $\times$ Known	4.70^{***}	4.70^{***}	4.70^{***}
	(1.29)	(1.25)	(1.28)
	0.000	0.000	0.000
Share pro	10.34^{***}	10.34^{***}	10.34^{***}
	(0.95)	(0.96)	(0.95)
	0.000	0.000	0.000
Fixed effects	Resp + Prop	Resp + Prop	Resp + Prop
Clustering	Participant	IID	HC
N	4506	4506	4506

Notes: Estimates from linear regressions of the assessment of argument quality on a scale from 0 to 100. Standard errors in parenthesis, and p values below.

Table A16 contains the same regressions as in Table 3, except replacing the prior and posterior with a logit transformation of these values. This matches more closely with the theoretical model. We replace responses of 0 with $.005$ and 1 with $.995$ before taking the transformation. While the coefficients are harder to interpret, the sign, significance, and relative magnitudes are similar to the untransformed version.

Table A18 shows versions of our preregistered specification, but on subsets of respondents based on how they answer partisan questions. For the first three columns we subset by self-reported partisanship. The coefficients on “for” and “for x known" are marginally higher for Republicans. As we sampled those who previously reported having a partisan preference, we have few independents, so these coefficients are very imprecisely estimated (though still in the predicted directions). Next we compute a social partisanship scale by converting the questions from (Huddy, Mason and Aarøe, 2015) on to a 0-1 scale, taking the mean across the four questions, and doing a median split. Those who are high on the scale have a somewhat larger coefficient on “for”, but a somewhat smaller coefficient on “for x known.” Finally, we do a median split on the out-party feeling

Table A15: Study 2, Predictors of Final Belief about Others' Support

	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	50.38^{**}	51.24^{**}	9.39^{**}			
	(0.67)	(0.79)	(0.96)			
Prior			0.69^{**}	0.70^{**}	0.62^{**}	0.62^{**}
			(0.01)	(0.01)	(0.02)	(0.02)
For	6.08^{**}	4.57^{**}	4.94^{**}	4.90^{**}	4.63^{**}	4.55^{**}
	(0.59)	(0.84)	(0.59)	(0.58)	(0.68)	(0.68)
Known		-1.65^{*}	-0.98	-0.95	-1.26	-1.24
		(0.84)	(0.63)	(0.63)	(0.70)	(0.69)
For × Known		3.03^{**}	2.18^{**}	2.21^{**}	2.25^{*}	2.33^{*}
		(1.15)	(0.83)	(0.82)	(0.95)	(0.95)
Share pro	8.91^{**}	8.83^{**}	8.42^{**}	8.42^{**}	7.52^{**}	7.48^{**}
	(0.90)	(0.90)	(0.65)	(0.65)	(0.74)	(0.74)
R^2	0.04	0.05	0.51	0.52	0.70	0.70
Fixed effects	No	No	No	Prop	Resp	Resp+Prop
N	4506	4506	4506	4506	4506	4506

Notes: Estimates from linear regressions of final posterior probability of supporting the proposition (self), in percentage points. Standard errors in parentheses.
* $p < .05$, ** $p < .01$.

thermometer. Those with low affective polarization (below median out-party feeling thermometer) have a higher coefficient on "for x known." Overall we view these differences as modest, indicating evidence for motivated information processing across all subgroups.

Table A19 shows the same specification for subsets based on the reported prior and argument text. First we look at those with an "interior prior" (i.e., not 0 or 100), and those with a moderate prior (between 20 and 80). In general, our treatments have a larger effect for this group, perhaps due to ceiling effects for those excluded. Next we do a median split on argument length. Those who make longer arguments have a larger coefficient on "for" but a smaller coefficient on "for x known." Finally we subset to those who are randomized to argue in the same direction as their prior belief or against it. The estimates are less precise here but generally similar.

Overall we take this as evidence that the results are not driven by any particular subgroup.

Table A20 runs regressions predicting the belief after all of the information is revealed, but before the argument is made. We control for the belief after the position (if revealed before) and one piece of information, in effect isolating the change in belief as the second to fourth piece of

Table A16: Study 2, Predictors of Final Support, Logit Scale

	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	-0.43^{**}	-0.30^{**}	-0.75^{**}			
	(0.08)	(0.10)	(0.06)			
Prior (logit)			0.86^{**}	0.86^{**}	0.84^{**}	0.85^{**}
			(0.01)	(0.01)	(0.01)	(0.01)
For	0.44^{**}	0.25^{*}	0.45^{**}	0.44^{**}	0.45^{**}	0.45^{**}
	(0.08)	(0.10)	(0.06)	(0.06)	(0.07)	(0.07)
Known		-0.26^{*}	-0.15^{*}	-0.15^{*}	-0.18^{*}	-0.17^{*}
		(0.11)	(0.06)	(0.06)	(0.07)	(0.07)
For × Known		0.39^{**}	0.21^{*}	0.21^{*}	0.25^{*}	0.25^{*}
		(0.15)	(0.08)	(0.08)	(0.10)	(0.10)
Share pro	0.87^{**}	0.86^{**}	0.79^{**}	0.80^{**}	0.71^{**}	0.71^{**}
	(0.11)	(0.11)	(0.06)	(0.06)	(0.08)	(0.07)
R^2	0.02	0.02	0.71	0.72	0.82	0.82
Fixed effects	No	No	No	Prop	Resp	Resp+Prop
N	4506	4506	4506	4506	4506	4506

Notes: Estimates from linear regressions of final posterior probability of supporting the proposition (self), in logit scale. Standard errors in parentheses.
* $p < .05$, ** $p < .01$.

information is revealed.[32] For this stage there are three effective groups: those who do not know their position, those who know they will argue for, and those who know they will argue against. The key variables are indicators for "knew for" and "unknown", treating "knew against" as the base category. Column (2) adds a control for the share of positive information in this phase, and various fixed effects.

Across specifications, we find positive (if not significant) coefficient on "unknown", and a positive and significant coefficient on "knew for." This indicates that the beliefs generally move in the direction of the position when known, even though on average the respondents are seeing the same blend of negative and positive information.

Table A21 presents the key recall specifications with alternative calculations of standard errors. This has little effect on the standard errors or p values.

[32]We cannnot isolate the effect of the first piece of information since our first belief elicitation is the prior and the second is after the position is revealed for some and the first piece of information is revealed.

Table A17: Study 2, Predictors of Final Support, Share of Positive Information

	(1)	(2)	(3)	(4)
Intercept	49.23^{**}	0.44		
	(1.21)	(0.85)		
Prior		0.84^{**}	0.84^{**}	0.84^{**}
		(0.01)	(0.01)	(0.01)
For				6.02^{**}
				(1.74)
Known	-1.06	-0.11	-0.06	-3.19
	(1.69)	(1.05)	(1.24)	(1.78)
For × Known				6.96^{**}
				(2.40)
Share pro	11.98^{**}	12.44^{**}	11.63^{**}	10.56^{**}
	(1.98)	(1.16)	(1.39)	(1.94)
Share pro × Known	0.83	-1.45	-2.09	-0.33
	(2.81)	(1.67)	(2.03)	(2.78)
Share pro × For				2.18
				(2.69)
Share pro × For × Known				-4.55
				(3.77)
R^2	0.02	0.66	0.79	0.81
Fixed effects	None	None	Prop + Resp	Prop + Resp
N	4512	4512	4512	4512

Notes: Estimates from linear regressions of final posterior probability of supporting the proposition (self), scale. Standard errors in parentheses.
* $p < .05$, ** $p < .01$.

Next we look at what predicts the argument length, in characters. Table A22 shows that those who know the argument at the outset write marginally longer arguments, though this is not significant. Those with a moderate prior write shorter arguments. Those with lower or high affective polarization or social partisanship do not make longer or shorter arguments. However, Democrats (the base category in column 6 with party) make slightly longer arguments

Table A23 shows regressions predicting whether each piece of information is used in the argument. As with the memory questions, we multiply the dependent variable by 100 so the coefficients can be interpreted as percentage points. The first column includes all pieces of information (whether shown to the participant or not). Unsurprisingly, congruent pieces of information are far more likely to be used, as are pieces of information which are actually shown.

Table A18: Study 2, Predictors of Final Support, Heterogeneity by Partisanship

	Dem	Rep	Ind	H Partisan	L Partisan	H Aff Pol	L Aff Pol
Prior	0.86^{**}	0.80^{**}	0.87^{**}	0.83^{**}	0.85^{**}	0.87^{**}	0.80^{**}
	(0.02)	(0.02)	(0.07)	(0.02)	(0.01)	(0.01)	(0.02)
For	6.01^{**}	8.85^{**}	4.61	8.71^{**}	5.87^{**}	7.31^{**}	7.26^{**}
	(1.15)	(1.48)	(5.62)	(1.50)	(1.10)	(1.38)	(1.21)
Known	-2.19	-5.72^{**}	0.21	-1.99	-4.55^{**}	-2.55	-4.45^{**}
	(1.22)	(1.50)	(6.23)	(1.51)	(1.21)	(1.40)	(1.30)
For × Known	4.68^{**}	5.08^{*}	2.21	4.08^{*}	5.15^{**}	3.11	6.01^{**}
	(1.69)	(2.03)	(8.09)	(2.02)	(1.66)	(2.01)	(1.69)
Share pro	11.30^{**}	9.41^{**}	0.31	11.46^{**}	9.42^{**}	10.90^{**}	10.60^{**}
	(1.26)	(1.49)	(5.83)	(1.50)	(1.20)	(1.45)	(1.27)
R^2	0.81	0.80	0.78	0.80	0.81	0.83	0.78
Fixed effects	Resp+Prop	Resp+Prop	Resp+Prop	Resp+Prop	Resp+Prop	Resp+Prop	Resp+Prop
N	2574	1818	114	1947	2559	2172	2220

Notes: Estimates from linear regressions of final posterior probability of supporting the proposition (self). Standard errors in parentheses.
* $p < .05$, ** $p < .01$.

Table A19: Study 2, Predictors of Final Support, Heterogeneity by Prior and Argument

	Int Prior	Mod Prior	Long Arg	Short Arg	Against Prior	With Prior
Prior	0.81^{**}	0.76^{**}	0.87^{**}	0.82^{**}	0.94^{**}	0.75^{**}
	(0.02)	(0.03)	(0.02)	(0.02)	(0.05)	(0.03)
For	7.91^{**}	8.18^{**}	10.25^{**}	4.94^{**}	13.06^{**}	13.42^{**}
	(1.02)	(1.45)	(1.29)	(1.45)	(3.47)	(2.17)
Known	-3.94^{**}	-4.74^{**}	-3.22^{*}	-4.93^{**}	-1.54	-2.34
	(1.07)	(1.47)	(1.38)	(1.59)	(1.56)	(1.63)
For × Known	5.60^{**}	5.69^{**}	3.63	6.24^{**}	3.88	2.29
	(1.41)	(1.95)	(1.86)	(2.12)	(2.85)	(1.93)
Share pro	10.87^{**}	10.37^{**}	12.18^{**}	8.58^{**}	10.67^{**}	8.29^{**}
	(1.02)	(1.47)	(1.46)	(1.54)	(1.99)	(1.42)
R^2	0.77	0.75	0.85	0.84	0.87	0.90
Fixed effects	Resp+Prop	Resp+Prop	Resp+Prop	Resp+Prop	Resp+Prop	Resp+Prop
N	3834	2382	2253	2253	2101	2405

Notes: Estimates from linear regressions of final posterior probability of supporting the proposition (self). Standard errors in parentheses.
* $p < .05$, ** $p < .01$.

Table A20: Study 2, Predictors of Support Dynamics

	(1)	(2)	(3)	(4)	(5)
Intercept	4.45^{**}	-1.15			
	(0.74)	(0.74)			
Pr(Vote) after 1	0.88^{**}	0.87^{**}	0.87^{**}	0.87^{**}	0.87^{**}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Share Positive		11.97^{**}	11.36^{**}	11.95^{**}	11.33^{**}
		(0.73)	(0.85)	(0.72)	(0.84)
Knew for	2.19^{**}	1.83^{**}	2.29^{**}	1.82^{**}	2.28^{**}
	(0.69)	(0.67)	(0.80)	(0.66)	(0.79)
Unknown	0.61	0.41	1.37	0.36	1.30
	(0.61)	(0.59)	(0.70)	(0.59)	(0.70)
R^2	0.72	0.74	0.83	0.74	0.83
N	4506	4506	4506	4506	4506
Fixed effects	None	None	Respondent	Proposition	Resp+Prop

Notes: Estimates from linear regressions predicting the probability of voting for the proposition (self) after all information is revealed, but before the argument is made, in percentage points. Standard errors (clustered by participant) in parentheses.
$^{*}p < .05$, $^{**}p < .01$.

Table A21: Study 2, Predictors of Recall, Alternative Standard Errors Clustering

	Correct Recall			Don't Remember		
	(1)	(2)	(3)	(4)	(5)	(6)
Congruent × Known	1.66	1.66	1.66	-1.42	-1.42	-1.42
	(1.29)	(1.29)	(1.29)	(1.07)	(1.05)	(1.05)
	0.199	0.199	0.199	0.184	0.179	0.179
Congruent	-1.32	-1.32	-1.32	0.15	0.15	0.15
	(0.89)	(0.91)	(0.90)	(0.73)	(0.74)	(0.74)
	0.140	0.147	0.144	0.839	0.843	0.842
Known	-1.28	-1.28	-1.28	0.81	0.81	0.81
	(1.03)	(1.03)	(1.03)	(0.86)	(0.84)	(0.85)
	0.212	0.215	0.212	0.349	0.339	0.343
Fixed effects	Q + Resp	Q + Resp	Q + Resp	Q + Resp	Q + Resp	Q + Resp
Clustering	Participant	IID	HC	Participant	IID	HC
N	18024	18024	18024	18024	18024	18024

Notes: Estimates from linear regressions of the assessment of argument quality on a scale from 0 to 100. Standard errors in parenthesis, and p values below.

Table A22: Study 2, Predictors of Argument Length

	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	312.16^{**}	308.47^{**}		312.51^{**}	309.34^{**}	321.22^{**}
	(6.42)	(6.66)		(7.05)	(7.78)	(6.96)
Moderate prior	-13.81^{**}	-13.87^{**}	-7.62^{*}	-13.81^{**}	-13.65^{**}	-13.02^{**}
	(4.83)	(4.83)	(3.62)	(4.83)	(4.95)	(4.82)
For	-2.33	4.56	-0.51	-2.31	-3.35	-2.56
	(4.13)	(5.96)	(2.92)	(4.12)	(4.21)	(4.12)
Known	5.36	12.50^{*}	3.17	5.35	5.41	5.60
	(4.35)	(5.95)	(3.12)	(4.35)	(4.45)	(4.35)
For × Known		-13.90				
		(8.34)				
Share pro	-11.79	-11.43	-7.75	-11.80	-12.53	-11.83
	(6.85)	(6.86)	(4.89)	(6.85)	(6.98)	(6.84)
H Partisan				-0.83		
				(6.67)		
H Aff Pol					7.80	
					(6.76)	
Independent						-19.23
						(15.42)
Republican						-22.20^{**}
						(6.67)
R^2	0.00	0.00	0.78	0.00	0.00	0.01
Fixed effects	None	None	Prop+Resp	None	None	None
N	4506	4506	4506	4506	4392	4506

Notes: Estimates from linear regressions of argument length (in characters). Standard errors in parentheses. * $p < .05$, ** $p < .01$.

Next we get to more subtle tests of whether knowledge of the argument position affects whether the piece of information ends up being used. In column (2) we add coefficients for whether the position was known, as well as an interaction between this and congruence. We find that the effect of congruence is slightly larger when the position is known. That is, the fact that a piece of information will be useful for the argument has a bigger effect when this is known at the time the information is introduced. This difference is about 1.7% with or without respondent and "piece of information" fixed effects.

For this to be driven by motivated information processing, the congruent x known coefficient should be driven by pieces of information which are actually shown to the participant. Columns (4) and (5) subset to this information, and the point estimates go up (albeit less precisely estimated).

Table A23: Study 2, Predictors of Argument Content

	(1) All	(2) All	(3) All	(4) Shown	(5) Shown	(6) Not	(7) Not
Intercept	6.37***	6.77***		11.34***		8.83***	
	(0.25)	(0.36)		(0.45)		(0.37)	
Congruent	22.48***	21.65***	21.76***	25.76***	25.98***	17.55***	17.51***
	(0.42)	(0.61)	(0.60)	(0.84)	(0.83)	(0.67)	(0.68)
Known		-0.78	-1.47*	-1.14	-2.59**	-0.60	-0.76
		(0.47)	(0.58)	(0.65)	(0.80)	(0.54)	(0.70)
Seen	6.57***	6.56***	5.82***				
	(0.34)	(0.34)	(0.36)				
Congruent × Known		1.67*	1.69*	2.10	2.61*	1.25	1.51
		(0.79)	(0.78)	(1.12)	(1.11)	(0.93)	(0.94)
R^2	0.16	0.16	0.26	0.18	0.33	0.11	0.28
N	36048	36048	36048	18024	18024	18024	18024
Fixed effects	None	None	Resp+Info	None	Resp+Info	None	Resp+Info

Notes: Estimates from linear regressions predicting whether a piece of information is used, multiplied by 100. Standard errors (clustered by participant) in parentheses.
*$p < .05$, **$p < .01$.

The point estimate when subsetting to information not shown (columns 6 and 7) is smaller and not statistically significant.

Table A24 shows regressions predicting getting a correct answer (columns 1-3) or answering "don't remember," with whether the information was used as a predictor. Phrases from information which is used is more likely to be remembered correctly. Importantly, the "used" variable is not randomly assigned, and it is hard to tell whether this relationship is causal. Column (2) shows that this relationship weakens when also controlling for the truth. This is likely because that information which is actually seen is more likely to be used, and participants are more likely to be correct in the memory questions where the answer is that the phrase was seen. Columns 4-6 show that participants are less likely to say they do not remember phrases from information they are coded as having used.

Table A24: Study 2, Predictors of Memory, Seen and Used Information

	(1) Correct	(2) Correct	(3) Correct	(4) DR	(5) DR	(6) DR
Intercept	64.05***	51.16***		21.32***	28.77***	
	(0.58)	(0.80)		(0.60)	(0.80)	
Used	6.77***	1.75	2.97**	-4.93***	-2.03*	-2.45**
	(1.08)	(1.02)	(0.95)	(0.90)	(0.88)	(0.81)
Seen		26.17***	26.92***		-15.12***	-15.84***
		(0.84)	(0.84)		(0.68)	(0.68)
R^2	0.00	0.08	0.30	0.00	0.04	0.33
N	18024	18024	18024	18024	18024	18024
Fixed effects	None	None	Resp+Question	None	None	Resp+Question

Notes: Estimates from linear regressions predicting whether the memory question for a phrase is answered correctly (columns 1-3) or with "don't remember" (columns 4-6). Standard errors in parentheses.
* $p < .05$, ** $p < .01$.