Are Preconceptions Postconceptions? Evidence on Motivated Political Reasoning[1]

Matthew Lilley[2]
Duke University, Economics Department

Brian Wheaton[3]
UCLA, Anderson School of Management

February 15, 2024

Abstract: How pervasive a phenomenon is motivated political reasoning, and can individuals be de-biased from engaging in it? To answer these questions, we run a survey experiment wherein one treatment group receives an empirical fact plausibly relevant to some normative belief and another treatment group receives this same fact, albeit presented as a hypothetical. The former group is then asked what their actual normative belief is while the latter is asked what their normative beliefs would be if the hypothetical was true. The experiment repeats this structure for a variety of political issues. We find that respondents claim to be quite open-minded, reporting hypothetical normative beliefs quite different from control group beliefs. But when the information is presented as a true fact, respondents' beliefs are instead statistically indistinguishable from the control group. In other words, despite claiming they would change their minds in response to new facts, respondents do no such thing. We further show that these effects are driven by cases where the fact does not comport with the respondent's ideology – evidence that the patterns we find do reflect motivated reasoning. Providing the information treatment to individuals who already stated their hypothetical normative beliefs – thereby tying their hands – ameliorates the extent of motivated reasoning on that specific normative question. However, when asked a different but very closely-related normative question instead, motivated reasoning persists undiminished. In a follow-up survey, motivated reasoning returns for the constrained question – though still remains less strong than the unconstrained one. All in all, these results suggest that motivated reasoning is a pervasive and powerful phenomenon.

1 Introduction

Motivated reasoning describes the tendency to uncritically accept new information that accords with one's worldview while critically analyzing and discarding that which does not.

[1] We are grateful to Alberto Alesina, Robert Barro, Ed Glaeser, Larry Katz, David Laibson, and Gautam Rao for helpful advice and suggestions. We also thank the participants of the Behavioral lunch seminar at Harvard, the Political Economy lunch seminar at Harvard, the economics departmental seminar at University of Georgia, and the GEM brownbag at UCLA.
[2] Contact: matthew.lilley@duke.edu
[3] Contact: brian.wheaton@anderson.ucla.edu

Allegations of motivated reasoning are particularly common in political domains, with actors across the political spectrum routinely accused of ignoring inconvenient facts to justify their beliefs. Conclusively demonstrating the existence of motivated reasoning, however, is more challenging. A key difficulty is that one needs a notion of how much individuals *should* update their beliefs in response to information. It may be perfectly rational to for individuals to discard information that is distant from their priors. More distant information may seem more implausible, opening the door for rational distrust of information. To deal with this issue and obtain a notion of how much individuals should update their beliefs in response to information, we employ a simple and intuitive approach: we ask them. That is, we ask individuals what their beliefs *would* be if they received a given piece of information.

More concretely, we ran a 2000-participant survey experiment through the survey panel company Prolific. The experiment involved a variety of different issues: affirmative action, climate change, crime rates in Democratic- vs. Republican-run cities, economic mobility, gun control, the Olympics, racial bias in policing, taxation of the top 1%, and transgender participation in sports. In the survey experiment, for each issue, we ask respondents to guess the true value of some empirical fact. For example, on the issue of climate change, the empirical fact pertains to the number of years out of the last 20 which were the hottest on record, according to NASA. On the issue of racial bias in policing, the empirical fact pertains to the percentage of individuals shot and killed by police over a five-year period who were black, according to the *Washington Post* database on police-involved shootings.

Then, for each issue, the survey splits into one of three arms. Respondents randomized into the control arm are simply asked for their normative beliefs on some matter plausibly influenced by the aforementioned empirical fact. Respondents randomized into the information treatment arm are asked for their normative beliefs, given that the true value of the empirical fact is x^* (which they are told is actually true). Respondents randomized into the hypothetical treatment arm are asked what their normative beliefs would be, if they hypothetically learned that the true value of the empirical fact was x^* (which they are not told is actually true or false). For example, for the issue of climate change, the normative beliefs questions pertain to (i) whether the respondent

believes climate change is occurring and (ii) whether strong action should be taken by government to combat it. For the issue of racial bias in policing, the normative beliefs questions pertain to (i) whether police are systemically racist against black people and (ii) whether police forces should have their funding reduced and re-allocated to social services.

The hypothetical treatment arm features an additional layer. Respondents in the hypothetical treatment arm of a given issue will receive the true empirical fact at the end of the survey and then be asked their actual normative beliefs. Because hypothetical normative beliefs are only elicited for one version of the normative beliefs question (either (i) or (ii)), each respondent should be directly constrained for one version of the normative beliefs questions and not for the other. This allows us to investigate whether tying respondents' hands by eliciting their conditional beliefs prior to information revelation is a way of de-biasing them from engaging in motivated reasoning – and whether any such de-biasing extends to highly related questions on which respondents' hands are not tied.

Finally, one week after the initial survey wave, we ran a follow-up survey. In this follow-up, we once again asked individuals their normative beliefs on all of the issues they encountered in the initial survey wave. This allows us to assess whether the motivated reasoning and/or the de-biasing persisted a week after the initial survey.

Our analysis of the survey results reveals evidence of strong, significant, and fairly persistent motivated reasoning. Specifically, in the hypothetical treatment arm, respondents report being quite open-minded and willing to change their beliefs in response to new information. Their normative beliefs are claimed to be reasonably flexible as a function of (hypothetical) information. On average, in response to the hypothetical information that corresponds to the truth, respondents report hypothetical normative beliefs statistically and meaningfully different from control group beliefs. On the contrary, respondents in the information treatment arm – who are presented with the true information and told it is indeed true – report beliefs statistically indistinguishable from control group beliefs. The gap between the behavior of the hypothetical treatment arm and the information treatment arm is suggestive of substantial motivated reasoning.

We next look within the hypothetical arm, comparing the hypothetical normative beliefs and the ex-post normative beliefs also provided by respondents in the hypothetical arm – after they receive the true information at the end of the survey. Here, too, we find a significant gap, though it is driven by the version of the normative beliefs question for which the respondent is unconstrained – the version for which they were not asked to provide their hypothetical beliefs. On the version for which respondents are constrained, motivated reasoning is substantially reduced, and the gap between hypothetical beliefs and ex post beliefs is much smaller. However, in our follow-up survey one week later, we find that motivated reasoning increases substantially even for the constrained question – though it is still not as strong as for the unconstrained question.

To further ascertain whether this behavior is actually motivated reasoning, we distinguish between favorable and unfavorable information from the perspective of a given respondent. Favorable information is that which comports with the ideological preferences of the respondent; unfavorable information does not. Our choice of political issues is broad, and our choice of empirical facts purposefully includes both some right-coded and some left-coded facts. For example, the fact that 19 out of the past 20 years are the hottest on record, per NASA, is severely underestimated by Republicans and only somewhat underestimated by Democrats. The fact that 26.4% of people shot and killed by police between 2015 and 2020 were black, per the *Washington Post*, is severely overestimated by Democrats and only somewhat overestimated by Republicans. We show that the lack of updating in response to information is driven most strongly by unfavorable information – consistent with the definition of motivated reasoning. This is the case in both the context of the comparison between the hypothetical treatment arm and the information treatment arm and the context of the comparison between the hypothetical beliefs and ex post beliefs within the hypothetical treatment arm.

Taken as a whole, we argue that these results provide strong evidence that motivated reasoning is a pervasive phenomenon in political contexts. It is possible to de-bias individual from engaging in motivated political reasoning in response to new information by eliciting their conditional beliefs beforehand, but this approach is only effective in a very narrow and specific manner. It requires doing so in each of the precise normative domains the information may affect

– and requires that the individual in question remembers said elicitation. Apart from this, motivated political reasoning appears to be the norm, not the exception.

2 Literature Review

At its roots, our work is founded on the notion that individuals strive to identify with – and to be accepted as – members of groups. A large sociological literature exists on the nature of group identity. This literature is reviewed in a number of articles in the *Handbook of Self and Identity*, notably Stets and Burke (2003) and Hogg (2003). Essentially, the overarching takeaway from this literature is that (most) humans seek to identify themselves with broader groups; to this end, they attempt to behave in ways that allow them to view themselves as good members of such groups.

This intuition is enriched by foundational models within economics of identity or ego utility, like those of Akerlof and Kranton (2000, 2010), Brunnermeier and Parker (2005), and Mobius et al (2014). In the context of such models, identity considerations enter directly into the utility function. Individuals may receive direct utility from behaving in a way consistent with their identity – or the identity they aspire to. For those with identity utility associated with their political party or ideology, deviations from rational beliefs can be utility-maximizing in this context, as they induce a direct, first-order improvement in identity utility but only a second-order loss in instrumental utility stemming from worsened decision-making due to the deviation from rational beliefs. Consequently individuals may prefer to accept only information that comports with their political identity, while discarding information that does not.

A literature of dozens of papers within social psychology attempts to identify the extent of such deviations in the political context – often termed motivated political reasoning. This literature is reviewed by Ditto et al. (2019). Most papers within this literature consist of experiments that measure respondents' assessments of a piece of information's quality after treating respondents with something that colors their assessment of the partisanship of the information source – without directly coloring its reliability. However, as discussed extensively by Tappin, Pennycook, and Rand (2020), this literature tends to suffer from a foundational issue: information that is distant from one's political identity/preferences is also likely to be distant from one's priors. And it is

perfectly rational under Bayes' Rule to (largely) discard information that is distant from one's priors. Phrased differently, the discrepancy between a piece of information and one's priors should indeed rationally inform one's assessment of that information in contexts where there is uncertainty about the information's reliability. While the many studies in this literature find evidence of motivated reasoning, Tappin, Pennycook, and Rand essentially argue that the bar to detecting motivated reasoning is set too low within this literature, and consequently perfectly rational behavior may be mis-characterized as motivated political reasoning.

In the realm of behavioral economics, our work relates most closely to that of Thaler (2020). Thaler provides evidence for the existence of motivated political reasoning. Thaler elicits the median point on individuals' belief distributions for a variety of issues. Respondents should be equally likely to expect the truth to be below/above the median, so Thaler treats respondents by telling them their median is either "too high" or "too low" and asks them to state whether they believe this information is true or false, finding that respondents are more likely to state that info favoring their party is true but no more likely to state that info that's actually true is true.

The work of Chopra, Haaland, and Roth (2022) is also related. They find in a survey experiment that fact-checking a left-leaning news sources reduces demand for it by highly-ideological Democrats while boosting demand amongst moderate Democrats – consistent with motivated political reasoning.

There is also some work within behavioral economics on motivated reasoning outside the realm of politics. Motivated reasoning about one's own ability is one such topic. Thaler (2021) examines motivated reasoning in the context of (over)confidence about one's own ability, finding stronger motivated reasoning of such a form amongst men than amongst women. Oprea and Yuksel (2021) study how motivated reasoning about one's own ability can be amplified in the context of social exchanges.

Our contribution, firstly, is to identify the existence of motivated political reasoning in what we think is a more robust approach to the aforementioned concerns than has been done to date. Instead of making any a priori assumptions about how much individuals should update their beliefs in response to a piece of information, we simply ask individuals how much they *would* update in

response to information x^*. We then compare that to how much they *do* update in response to x^*. Unlike most past studies, our approach also focuses directly on the extent of belief updating as the outcome. Additionally, our approach is careful to distinguish between reporting "true information" and "true information according to source X." It is a subtle, often-overlooked, yet important point that individuals may distrust any given source, and consequently, eliciting beliefs without making this distinction may lead to individuals reporting as priors what they think the actual truth is, whereas they are treated with what the source says the actual truth is. Because these are fundamentally two different distributions, this may lead to motivated reasoning being detected when it is not actually occurring. Our approach avoids this issue.

Furthermore, woven into our approach is a potential tool for de-biasing; this is our second main contribution. We show that tying individuals' hands by eliciting their conditional beliefs before information revelation can be an effective way of de-biasing them from motivated reasoning. Put more simply, we have people commit to how they would respond to information before they receive it. However, motivated reasoning is so strong and persistent that this tool works if and only if the conditional beliefs are elicited on exactly the domain that ex post beliefs are later elicited. Highly related but technically distinct beliefs will not be de-biased.

3 Analytic Framework

3.1 Theory

In the standard rational framework, beliefs affect utility only instrumentally. Having correct beliefs helps individuals optimize their decision-making. We draw on models of identity or ego utility – such as in Akerlof and Kranton (2010), Brunnermeier and Parker (2005), and Mobius et al (2014) – whereby beliefs affect utility both instrumentally and also directly. Individuals receive direct identity or ego utility from holding and professing certain beliefs. Slightly more formally, let an individual hold beliefs $\theta \in \mathbb{R}$ about some political issue. The individual has an identity bliss point of θ^{ID} – the beliefs he or she would like to hold absent any instrumental concerns. Meanwhile, given information set $\boldsymbol{\mathcal{I}}$, let the rational beliefs be denoted by $\theta_{\boldsymbol{\mathcal{I}}}^{\mathrm{RE}}$. Utility is governed by a function of the form $U(c(\theta), \theta)$, whereby the first argument corresponds to instrumental

concerns and therefore $c(.)$ is decreasing in the distance between rational beliefs (or truth) and chosen beliefs. The second argument corresponds to identity/ego utility. The canonical result from the aforementioned papers is that there will be a distortion of the individual's beliefs θ away from $\theta_{\mathcal{I}}^{\mathrm{RE}}$ toward θ^{ID}. This $\Delta\theta$ produces first-order gain in identity utility and only second-order instrumental loss. Thus some amount of distortion is utility-maximizing from the individual's perspective.

To understand this distortion better, however, we need some notion of θ^{ID}. As previously noted, a large sociological literature exists on the nature of group identity. This literature is reviewed in a number of articles in the *Handbook of Self and Identity*, notably Stets and Burke (2003) and Hogg (2003). Essentially, the overarching takeaway from this literature is that (most) humans seek to identify themselves with broader groups; to this end, they attempt to behave in ways that allow them to view themselves as good members of such groups. "A scientist, for example, may act in ways that make it clear to herself, as well as to others, that she is careful, analytical, logical, and experimentally inclined. She may engage in a variety of actions and interactions to convey these images" (Stets and Burke 2003). Similarly, a politically-inclined liberal (or conservative) may want to take actions consistent with his or her group identity and therefore experience disutility from professing conservative (or liberal) beliefs. However, we conjecture, politically-inclined people of any persuasion may also want to view themselves as open-minded – and take actions to portray that image – so long as it doesn't conflict too harshly with the preceding goal. Thus we hypothesize that many people may profess willingness to change their political views in the face of new information – but fail to do so when they are actually given new information. To test this hypothesis, we turn to a pre-registered survey experiment.

3.2 Survey Experiment

We prepared an incentivized survey experiment on the Qualtrics survey platform. The experiment was pre-registered with the AEA RCT Registry, and it was distributed to 2000 participants through the survey panel company Prolific. Prolific maintains a panel of US respondents, and we made use of their nationally-representative sample. Research comparing the

quality of various survey panels has marked Prolific's as amongst the best – if not the best – on a variety of metrics (cf. Peer et al. 2022 and Douglas et al. 2023), significantly better than Amazon Mechanical Turk. The survey opens by asking a variety of demographic questions and other questions useful for classification (e.g., age, sex, race, party ID, ideology, etc.).

The experiment involved a variety of different issues: affirmative action, climate change, crime rates in Democratic- vs. Republican-run cities, economic mobility, gun control, the Olympics, racial bias in policing, taxation of the top 1%, and transgender participation in sports. Each respondent proceeds through the structure of the survey issue-by-issue for five randomly-selected issues from the above list of nine. Within a given issue, the respondent may be randomly sorted into the control arm, the information treatment arm, or the hypothetical treatment arm. (This randomization is re-done within each issue, so the typical respondent is in the control arm for some issues, information for others, and hypothetical for others.)

In all arms, the respondent is asked to guess the true value of some empirical fact that is related to the issue in question. For example, for the taxation issue, respondents are asked to guess what percentage of total income tax receipts are paid by the top 1% of earners. For the crime issue, respondents are asked to guess the murder rate in the two largest Republican-run cities (given the murder rate in the two largest Democratic-run ones).

Following this, in the control arm, respondents are simply asked for their normative beliefs on some matter (plausibly) influenced by the aforementioned empirical fact. For example, for the taxation issue, the normative beliefs questions are (i) whether the respondent thinks taxes should be raised on the top 1% and (ii) whether the respondent thinks the top 1% excessively use loopholes to evade taxes. For the crime issue, respondents are asked (i) whether they think the policies of Democratic mayors create crime problems and (ii) whether they prioritize rehabilitation over punishment in criminal justice.

Respondents randomized into the information treatment arm are asked for their normative beliefs, given that the true value of the empirical fact is x^* (which they are told is actually true). Respondents randomized into the hypothetical treatment arm are asked what their normative beliefs would be, if they hypothetically learned that the true value of the empirical fact was X – for

three different values of X (none of which are they told is actually true or false). One of the three different values of X presented is always x^*. These different values are presented on different pages so as to more closely mirror the structure with which the information treatment is delivered. This structure allows us to compare the extent to which individuals claim they will revise their beliefs when provided with new information (hypothetical treatment arm relative to control) to the extent to which they actually do (information treatment arm relative to control).

The hypothetical treatment arm features an additional layer at the end of the survey. Respondents in the hypothetical treatment arm of a given issue will receive the true empirical fact at the end of the survey and then be asked their actual normative beliefs. Because hypothetical normative beliefs are only elicited for one version of the normative beliefs question (either (i) or (ii)), each respondent should be directly constrained for one version of the normative beliefs questions and not for the other. This allows us to investigate whether tying respondents' hands by eliciting their conditional beliefs prior to information revelation is a way of de-biasing them from engaging in motivated reasoning – and whether any such de-biasing extends to highly related questions on which respondents' hands are not tied.

We note that the guess individuals are asked to make about an empirical fact and the actual/hypothetical empirical fact itself are both clearly articulated as being about the data *as reported by a given source*. For example, individuals are asked to guess the number of years NASA reports are hottest on record, and they are told the number of years that NASA reports are hottest on record. Individuals are asked to guess what this percentage is, according to the *Washington Post* database, and they are told the percent of people shot and killed by police who are black, according to the *Washington Post* database. It is a subtle, often-overlooked, yet important point that individuals may distrust any given source, and consequently, we are careful to avoid creating asymmetries by asking individuals what their guess about the true fact is while instead giving them the fact as per a certain source. Both the guess and the fact – actual and hypothetical – are about the source's data. This reduces the likelihood of spuriously detecting motivated reasoning.

3.3 Data Processing

Because we collected our data through a randomized survey experiment, it is possible to identify treatment effects with simple regressions comparing belief outcomes between the control, information, and hypothetical groups. Having said that, to enable running regressions that pool all of the issues we examine together, it is first necessary to transform the data. Specifically, we normalize the normative beliefs outcome such that the information should drive the normative beliefs outcome higher the hypo group higher, and therefore motivated reasoning can be measured as insufficient upward movement in beliefs in the information treatment group relative to the hypothetical treatment group. We do this in two different ways, the latter of which builds on the former.

In the first approach (a.k.a. V1), we do this at the issue level. For example, in the case of climate change issue, the vast majority of respondents reported a guess (about the number of years out of the past 20 which were the hottest on record, according to NASA) which was below the true number, 19. The normative beliefs questions for the climate change issue pertain to (i) whether the respondent believes climate change is occurring and (ii) whether the respondent thinks strong action should be taken against climate change, even if costly. Responses to these questions are entered on a 5-point scale between "Strongly Disagree" [1] and "Strongly Agree" [5]. Therefore, for the typical person, the information (that an indicator of climate change is worse than expected) should be associated with an upward movement in beliefs. No transformation is required for this issue. Conversely, in the case of the police racism issue, the vast majority of respondents reported a guess (about the percentage of people shot by police who were black, according to the *Washington Post*) which was above the true number, 26.4%. The normative beliefs questions for the police racism issue pertain to (i) whether police are systemically racist and (ii) whether police forces should have their funding cut and reallocated to social services. Therefore, for the typical person, the information (that an indicator of police bias is not as bad as expected) should be associated with a downward movement in beliefs. The beliefs variable must be rotated (6 minus beliefs).

In the second approach (a.k.a. V2), we do the normalization at the issue-by-individual level. This represents a slight refinement of the previous approach. Consider again the climate change example. The vast majority of people did guess a number lower than 19. However, a small number of people guessed 20. For these individuals, the information suggests that an indicator of climate change is not as bad as expected. This should be associated with a downward movement in beliefs. Therefore, these individuals within the climate issue should have their beliefs variable rotated (6 minus beliefs).

3.4 Identification

Having completed these transformations, it is possible to run simple regressions geared toward identifying motivated reasoning. For all of these regressions, we drop the hypothetical normative beliefs responses corresponding to untrue hypotheticals. Because we are interested in comparing how individuals say they *would* respond to information to how they actually do respond to the information, we must focus on the true hypothetical. For the first regression, we compare hypothetical beliefs in the hypothetical treatment group and actual beliefs in the information treatment groups to actual beliefs in the control group.

$$Y_{iqs} = \alpha + \beta\, Info_{is} + \gamma\, Hypo_{is} + \delta X_{iqs} + \varepsilon_{iqs}, \quad (1)$$

where Y_{iqs} denotes the (hypothetical or actual) normative beliefs reported by respondent i on normative beliefs question q of issue s. $Info_{is}$ is an indicator for whether respondent i is in the information treatment group for issue s. $Hypo_{is}$ is an indicator for whether respondent i is in the in hypothetical treatment group for issue s. X_{iqs} is a vector of controls. ε_{iqs} is the error term. In this setting, the extent of motivated reasoning is discernable by comparing β to γ. Because our survey is a randomized experiment, β and γ can be interpreted as causal treatment effects.

To more succinctly obtain one coefficient representing the extent of motivated reasoning, it is possible to further trim the sample, dropping the control group and instead running the following specification:

$$Y_{iqs} = \alpha + \beta\, Info_{is} + \delta X_{iqs} + \varepsilon_{iqs}, \quad (2)$$

The omitted group is now the hypothetical group, and β tells us the extent of motivated reasoning occurring: the more negative the value β, the stronger the extent of motivated reasoning. Because this specification provides more succinct results, we will mostly rely on it for our main results. However, corresponding versions of specification (1) will be provided in the appendix.

It is worth noting that, in order to verify what we are observing is actually motivated political reasoning, as clasically understood, we need to verify that there is asymmetric updating in response to ideologically-favorable versus ideologically-unfavorable information. We can define an indicator variable *Unfavorable* that is equal to 1 when the information does not comport with the respondent's ideology. By design, some of the empirical facts in our survey are more favorable to right-wing respondents, whereas others are more favorable to left-wing respondents. More precisely, the fact is closer to the median guess of right-wing respondents in some cases (racial bias in policing, taxation of the 1%, transgender participation in sports, and affirmative action) and closer to the median guess of left-wing respondents in others (climate change, economic mobility, crime in Democratic- vs Republican-run cities, and gun control). The list is consistent with what we marked as left-coded and right-coded treatments in our AEA RCT pre-registration. To be more concrete, the fact that 19 of the past 20 years have been the hottest on record, according to NASA, is more favorable to left-wing respondents because it is closer to their median guess. The fact that 26.4% of people shot and killed by police in recent years were black, according to the *Washington Post*, is more favorable to right-wing respondents because it is closer to their median guess. Therefore, we can cross-reference the partisan valence of the empirical fact for each issue with the party ID variable we collected in our survey to create the *Unfavorable* indicator. (Results are highly similar if, instead of party ID, we use the ideology variable or combine the two into a political score variable.) This allows us to run the following specification:

$$Y_{iqs} = \alpha + \beta\, Info_{is} + \eta\, Unfavorable_{is} + \rho\, Info_{is} * Unfavorable_{is} + \delta X_{iqs} + \varepsilon_{iqs}, \qquad (3)$$

In this specification, a negative and statistically-significant ρ tells us that the result we are finding is consistent with the classical definition of motivated reasoning, whereby the respondent is more critical of and more likely to discard unfavorable information versus favorable information.

It is possible to run an alternative version of any of the above specifications that, instead of focusing on comparing the information group to the hypothetical group, focuses on comparing the hypothetical belief answers to the ex-post beliefs answers within the hypothetical group. As noted previously, individuals in the hypothetical group for any issue are later told what the true information is and asked both normative beliefs questions. This specification tests whether motivated reasoning exists even in this setting where individuals' hypothetical normative beliefs (at least on one version of the normative beliefs question) were previously elicited:

$$Y_{iqs} = \alpha + \beta\, HypoExPost_{is} + \delta X_{iqs} + \varepsilon_{iqs}, \quad (4)$$

The sample for this regression consists exclusively of the hypothetical group – but adds the ex post normative beliefs responses after these individuals are actually provided the information. Here, as in (2), β tells us the extent of motivated reasoning. Note that there are two versions of each normative beliefs question, only one of which these individuals were previously asked their hypothetical beliefs on. Thus it is possible to split the analysis by q = {Constrained} or q = {Unconstrained} in order to determine the efficacy of the hypothetical elicitation as a tool for de-biasing in each of these contexts.

We also note that each specification – (1) through (4) – can be run with varying numbers of demographic and other controls, so we run versions that both include and exclude such controls. In all specifications, we cluster our standard errors at the respondent level.

4 Results

Before turning to our regressions, we begin with some descriptive notes on the data. To get a sense of the ideological distribution – and as a sanity check of sorts – in Figure 1 we plot the distribution of ideology by political party of the respondent. Here ideology is measured by taking the average of normative beliefs (normalized so that a higher number corresponds to more right-leaning beliefs) across issues in the control group. As one would expect, the Democratic distribution is to the left of the Republican distribution, with the distribution of Independents centered between the two. There appears to be a decent amount of ideological variation in our dataset, and the groups line up as expected.

In Figure 2, we plot hypothetical normative beliefs as a function of which hypothetical piece of information the respondent is considering. Because multiple issues are pooled in this figure, the beliefs are rotated such that a higher value of the information should correspond to a higher value of the beliefs outcome. Also, beliefs and information are both residualized on hypothetical fixed-effects. The plot thus shows that, on average, respondents claim to be to new empirical information on the topics we consider. While many people claim their beliefs are sensitive to new information, there are also people in our data who report that their beliefs are not. Essentially, these are people who are saying the empirical fact in question is not particularly relevant to their beliefs. This is not (necessarily) motivated reasoning and will not be measured as such in our regressions. In order to be a motivated reasoner, it is first necessary to claim that one *will* adjust one's beliefs in response to new information.

Next we turn to our regression analysis. Figure 3 shows the results of specifications corresponding to regression equation (1). The specification in the top row uses the V1 beliefs outcome variable and features no issue-by-question FEs or demographic controls. The second specification adds the issue-by-question FEs and the third additionally adds demographic controls (age by issue FEs, sex by issue FEs, race by issue FEs, income group by issue FEs, education by issue FEs, and 7-point ideology by issue FEs). The specifications in the fourth through sixth rows repeat this pattern, albeit with the V2 beliefs outcome. The results are highly consistent across the specifications: respondents receiving the hypothetical treatment report significantly and meaningfully higher beliefs than the control group, whereas respondents receiving the information treatment report little to no meaningful differences from control. The gap between the two sets of coefficients is indicative of motivated reasoning.

Indeed, Figure 4 shows the results of analogous specifications corresponding instead to regression equation (2). Here the coefficients can more directly be interpreted as the extent of motivated reasoning. A more negative coefficient reflects insufficient updating relative to what was claimed in the hypothetical, evidence of motivated reasoning. Strong evidence to this effect is apparent across all the specifications.

In Figure 5, we turn to regression equation (4). The specification in the top row includes ex-post responses to both the constrained and unconstrained normative beliefs questions. Recall that the hypothetical treatment elicits respondents' hypothetical beliefs for one of the two normative beliefs questions on an issue. Therefore, when these respondents are later provided with the true information, their answer is constrained on one of the issue's normative beliefs questions but not (directly) constrained on the other. Comparing the top row in Figure 5 to the top row in Figure 4 reveals noticeably less motivated reasoning in this context, however. That said, rows two and three of Figure 5 reveal that this is almost entirely driven by the constrained question. On the constrained question, little to no motivated reasoning is occurring; tying respondents' hands appears to be effective at de-biasing them. However, on the unconstrained question, almost as much motivated reasoning is occurring as in Figure 4. Little to none of the de-biasing carries over to a very closely related – but technically different – normative beliefs question. Rows four through six repeat these specifications for the V2 beliefs outcome.

In Figure 6, we turn to regression equation (3). This is where we interact the indicator that the respondent received *Unfavorable* news with the information treatment. The mechanism suggested by the theory implies that this group should be driving the lack of updating. Indeed, this is important for ascertaining that what we are detecting is indeed motivated reasoning. Surely enough, this is precisely what we find in the data. The first three rows correspond precisely to equation (3) – with varying inclusion of control variables. The second three rows correspond to a version of equation (3) that compares the ex post normative beliefs to hypothetical normative beliefs within the hypothetical group. The result is actually a bit stronger statistically in these latter regressions. Note that, in this figure, all specifications use the V2 beliefs outcome. This is because the *Unfavorable* indicator variable is necessarily defined at the individual-by-issue level. The V2 beliefs outcome is also defined at the individual-by-issue level, whereas the V1 beliefs outcome is defined at the issue level – a mismatch that would lead to measurement error in this context.

In the interest of transparency, we run the specifications corresponding to regression equations (2) and (4) issue-by-issue. The results are shown in Figure 7. This cuts our sample size in ninths and therefore substantially undercuts the amount of statistical power we have to identify

motivated reasoning for any given issue. Having said that, we still find statistically-significant evidence of motivated reasoning on 5 out of 9 issues. In other words, our results do not appear to be driven by any one issue.

As noted previously, we ran a follow-up survey one week after the main survey, asking individuals once again their normative beliefs. Figure 8 shows the results of specifications that are instead estimated over the follow-up sample. Motivated reasoning persists quite strongly across the board. There is, however, some evidence that tying respondents' hands continues to de-bias them even one week out. The gap between the extent of motivated reasoning amongst those who were in the information treatment group versus the hypothetical treatment group is noticeably larger in the constrained context than the unconstrained context; this difference is significant in the specification using the V1 beliefs outcome. Less motivated reasoning is occurring for the question on which conditional beliefs were elicited before providing the information, suggesting some effects of our de-biasing may actually persist.

5 Conclusion

We study the phenomenon of motivated political reasoning and whether individuals can be de-biased from engaging in it. We do this by conducting a survey experiment wherein individuals in a hypothetical treatment arm are asked what their normative beliefs would hypothetically be if they learned a certain piece of information was true, whereas individuals in an information treatment arm are asked the same normative beliefs questions but actually given that information and told it is true. We find that individuals claim to be quite open-minded; they claim their beliefs would change in response to the information. In actuality, no such thing happens. We show that this effect is driven by receipt of ideologically-unfavorable information, i.e., information that does not comport with the respondent's worldview. In other words, consistent with the definition of motivated reasoning, respondents avoid updating their beliefs in response to unfavorable information.

Additionally, our work provides the first evidence (that we are aware of) that it is possible to de-bias individuals from engaging in motivated reasoning. Our method entails tying individuals'

hands by eliciting their conditional beliefs in advance of information provision – or, put more simply, having people commit to how they would respond to information before they receive it. Having said that, our general finding is that motivated reasoning is a pervasive and persistent phenomenon. Our de-biasing only appears to be effective under specific conditions. The conditional beliefs must be elicited on exactly the domain that ex post beliefs are later elicited. Highly related but technically distinct beliefs will not be de-biased. Therefore we think a fruitful avenue for future work could be to uncover other methods of de-biasing, hopefully methods that are more powerful and therefore broader in their effects.

References

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. The quarterly journal of economics, 115(3), 715-753.

Akerlof, G. A., & Kranton, R. E. (2010). Identity economics: How our identities shape our work, wages, and well-being. Princeton University Press.

Brunnermeier, M. K., & Parker, J. A. (2005). Optimal expectations. American Economic Review, 95(4), 1092-1118.

Chopra, F., Haaland, I., & Roth, C. (2022). Do people demand fact-checked news? Evidence from US Democrats. Journal of Public Economics, 205, 104549.

Ditto, P. H., Liu, B. S., Clark, C. J., Wojcik, S. P., Chen, E. E., Grady, R. H., ... & Zinger, J. F. (2019). At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives. Perspectives on Psychological Science, 14(2), 273-291.

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. Plos one, 18(3), e0279720.

Hogg, M. A. (2003). Social identity. In Leary, M.R. and Tangney, J. (eds), Handbook of self and identity, 462-479.

Möbius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2014). Managing self-confidence. NBER Working paper, 17014.

Oprea, R., & Yuksel, S. (2022). Social exchange of motivated beliefs. Journal of the European Economic Association, 20(2), 667-699.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. Behavior Research Methods, 1.

Stets, J. E., & Burke, P. J. (2003). A sociological approach to self and identity. In Leary, M.R. and Tangney, J. (eds), Handbook of self and identity, 23-50.

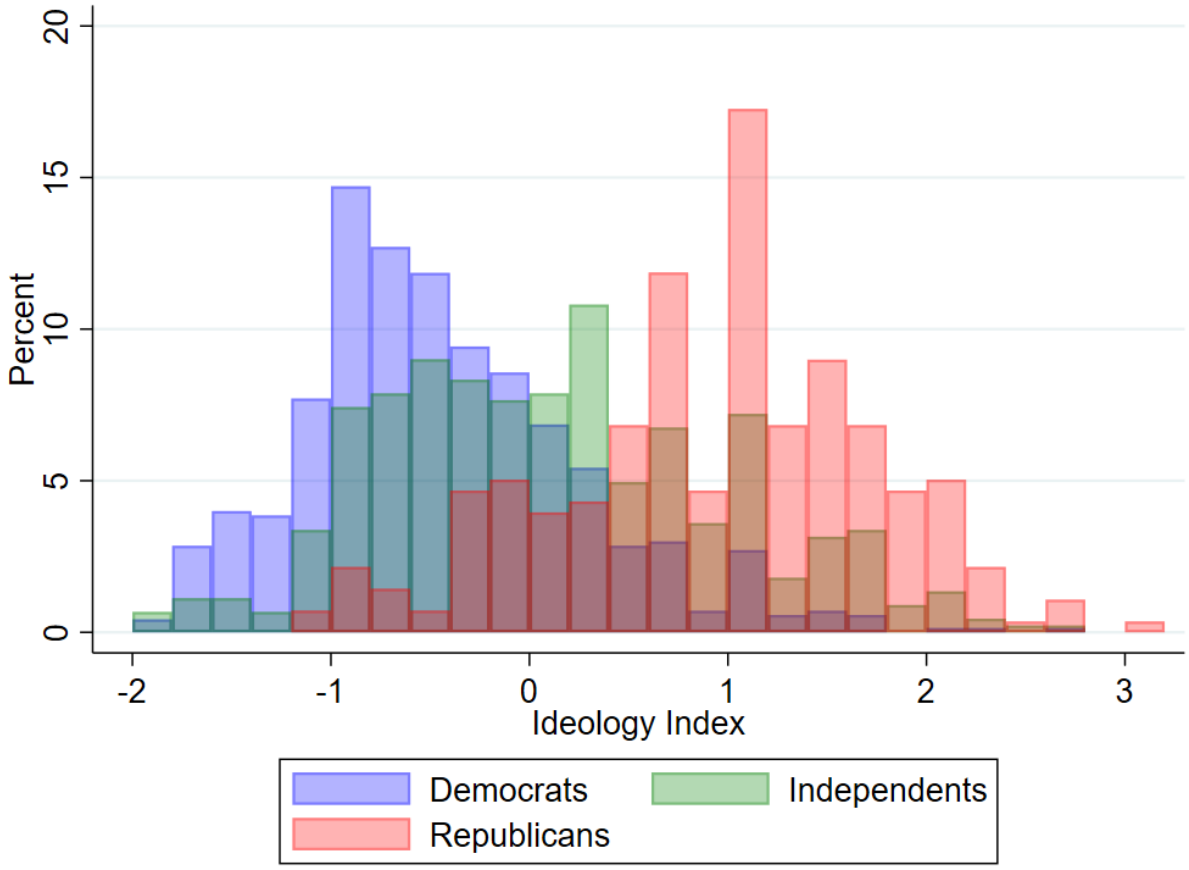
Tappin, B. M., Pennycook, G., & Rand, D. G. (2020). Thinking clearly about causal inferences of politically motivated reasoning: Why paradigmatic study designs often undermine causal inference. Current Opinion in Behavioral Sciences, 34, 81-87.

Thaler, M. (2020). The fake news effect: Experimentally identifying motivated reasoning using trust in news. arXiv preprint arXiv:2012.01663.

Thaler, M. (2021). Gender differences in motivated reasoning. Journal of Economic Behavior & Organization, 191, 501-518.

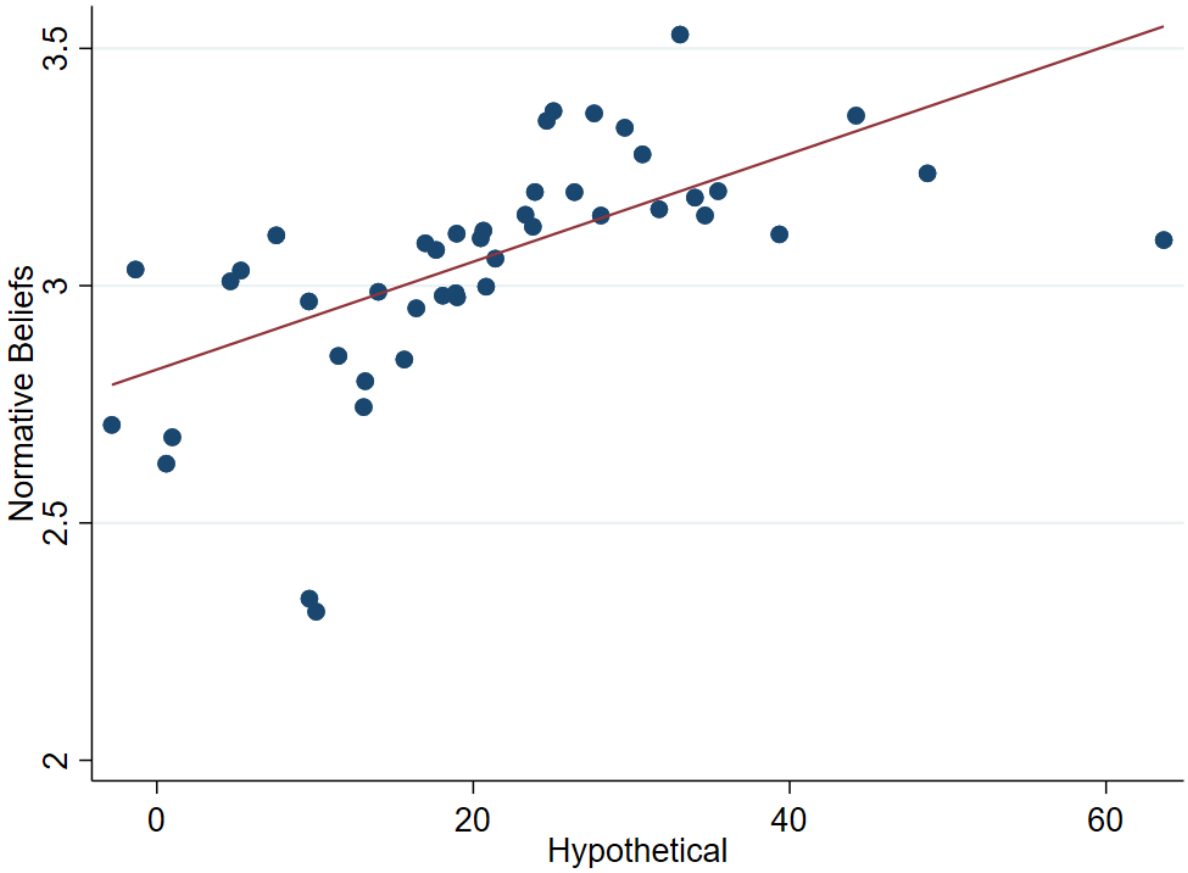
Figures

Figure 1: Ideology Distribution



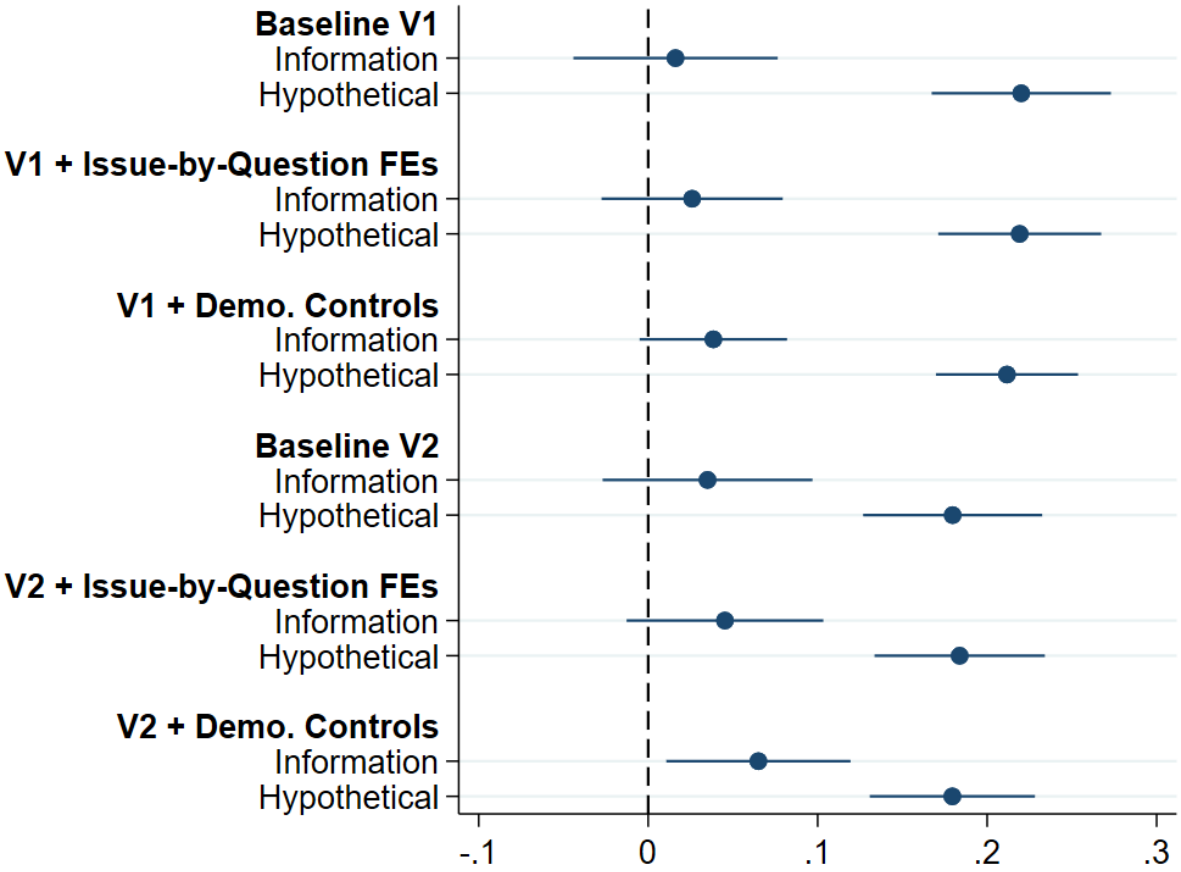
Note: This figure displays the average ideology (across issues) of respondents in the control group, as determined by their responses to the normative beliefs questions.

Figure 2: Binscatter – Response of Conditional Beliefs to Hypotheticals, Pooled Across Issues



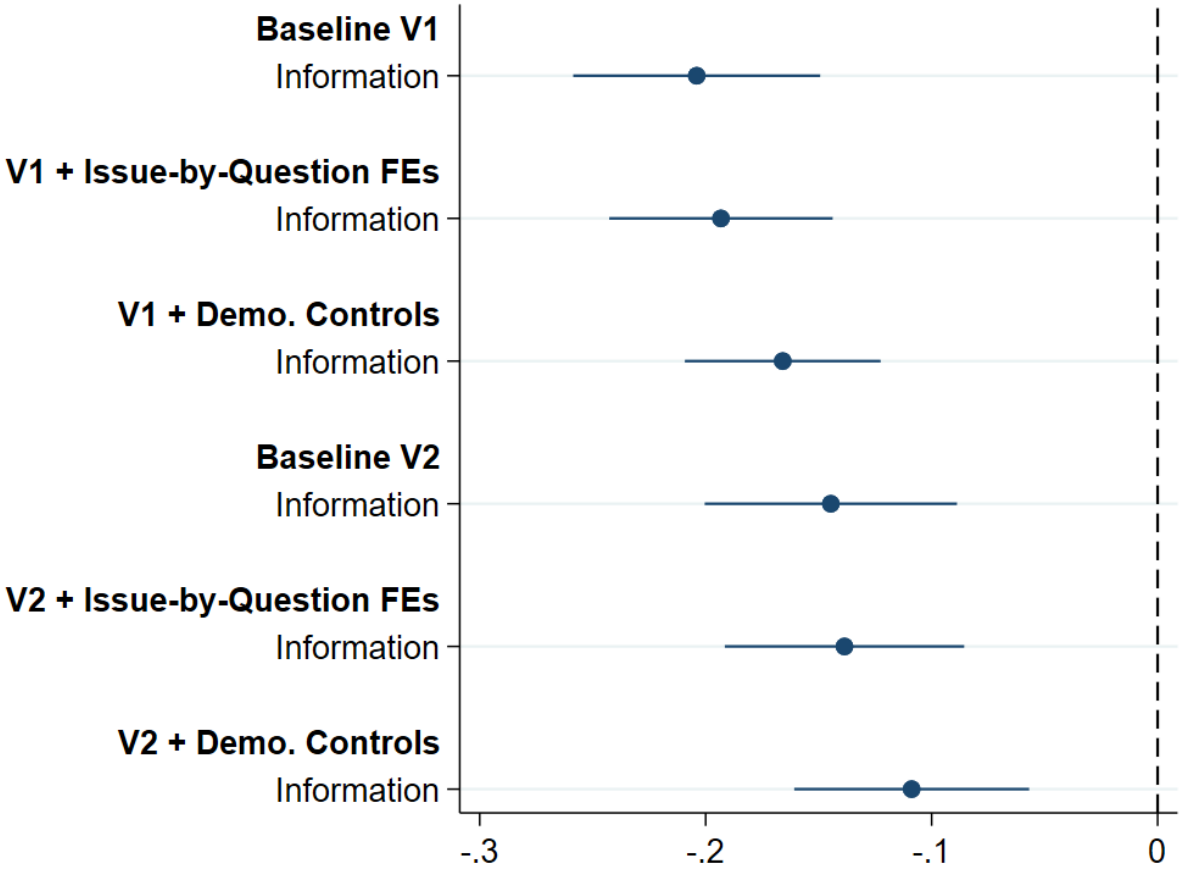
Note: This figure displays the effects of hypothetical information on normative beliefs. Normative beliefs are on a 1 to 5 scale ("Strongly Disagree" to "Strongly Agree"), rotated as necessary for certain issues so that a higher hypothetical piece of information should correspond to higher normative beliefs. Hypothetical information and normative beliefs are both residualized on an issue fixed-effect in order to pool the data across issues.

Figure 3: Effect of Information and Hypothetical on Beliefs (Relative to Control)



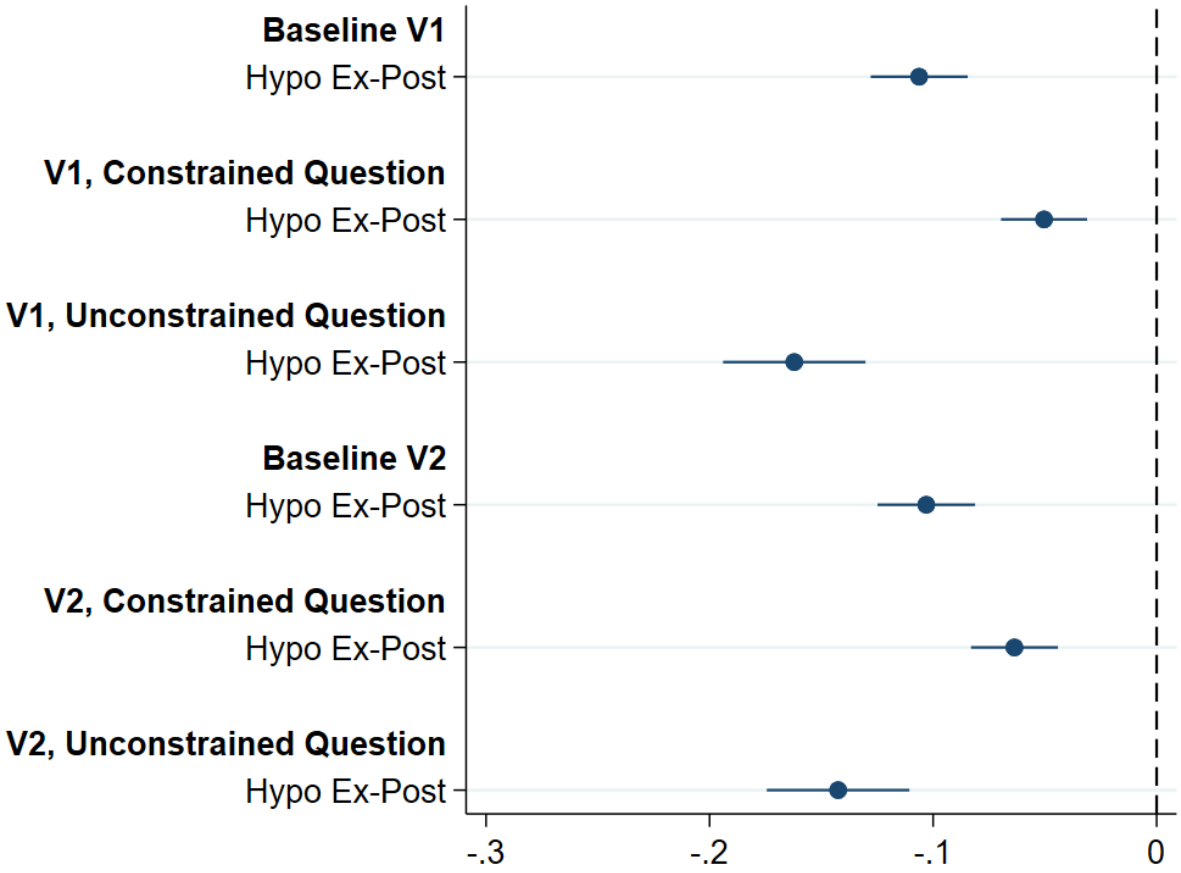
Note: This figure displays the results of regression specifications corresponding to Equation (1). Essentially, it shows the effects of the information treatment and the hypothetical treatment on beliefs, relative to the control group. On average, respondents claim that their beliefs would change substantially if they hypothetically received information x^*, but when they actually receive x^*, their beliefs scarcely change. The top specification corresponds to the V1 beliefs outcome with no RHS control variables. The second specification adds issue-by-question FEs. The third further adds a variety of demographic controls. The remaining three repeat this for the V2 beliefs outcome.

Figure 4: Effect of Information on Beliefs (Relative to Hypothetical)



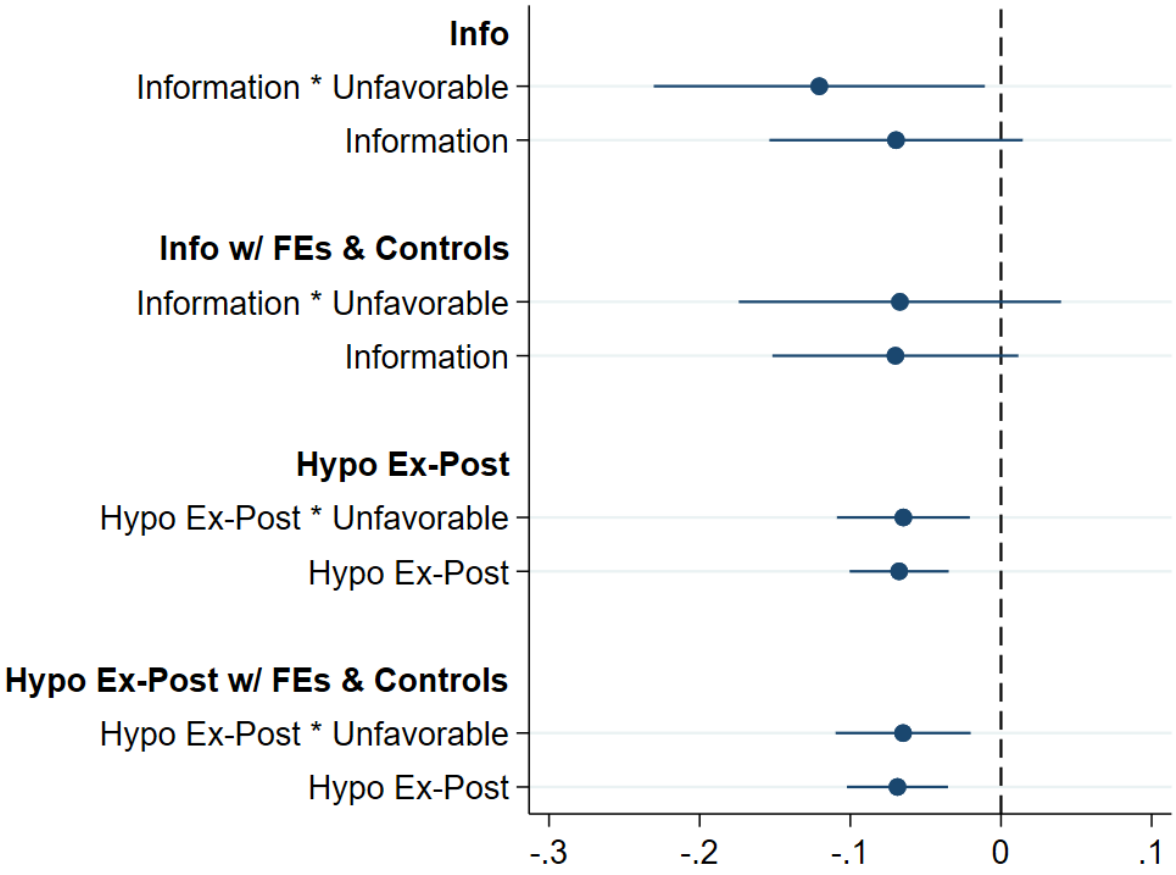
Note: This figure displays the results of regression specifications corresponding to Equation (2). Essentially, it shows the effects of the information treatment on beliefs, relative to the hypothetical treatment. On average, respondents update their beliefs substantially less than they claim they would, indicative of motivated reasoning. The top specification corresponds to the V1 beliefs outcome with no RHS control variables. The second specification adds issue-by-question FEs. The third further adds a variety of demographic controls. The remaining three repeat this for the V2 beliefs outcome

Figure 5: Effect of Post-Hypothetical Information on Beliefs (Relative to Hypothetical)



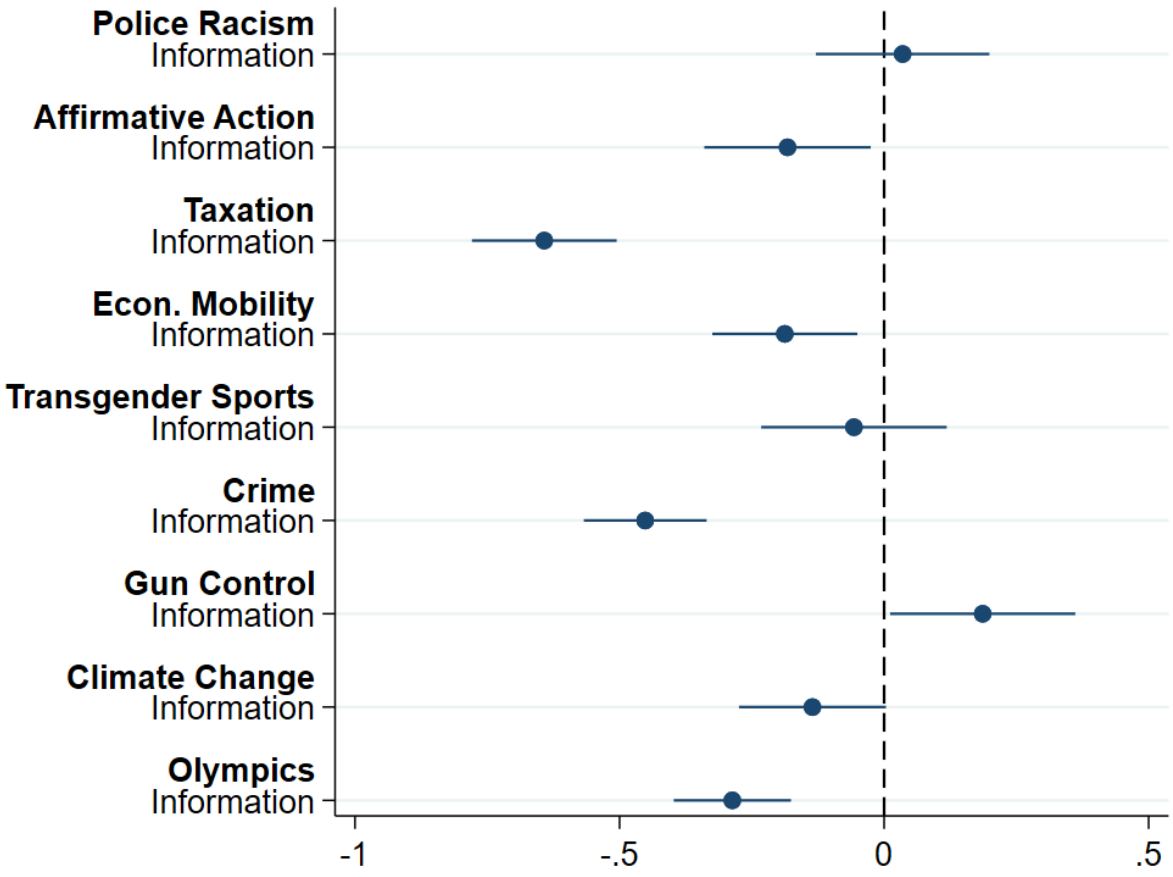
Note: This figure displays the results of regression specifications corresponding to Equation (4). Essentially, it shows the effects of providing information to members of the hypothetical group *after* their hypothetical beliefs were already elicited. Relative to what they claimed they would hypothetically do, these respondents update their beliefs substantially less. This is driven by their responses to the normative beliefs question for which their hypothetical beliefs were not elicited (unconstrained); however, some motivated reasoning is even observed on the question for which their hypothetical beliefs were elicited (constrained). The first three specifications use the V1 beliefs outcome; the latter three use the V2 beliefs outcome. Controls are not included in any of these specifications, but the results scarcely change if they are (since the main comparison here is within-individual).

Figure 6: Effect of Unfavorable Information on Beliefs (Relative to Hypothetical)



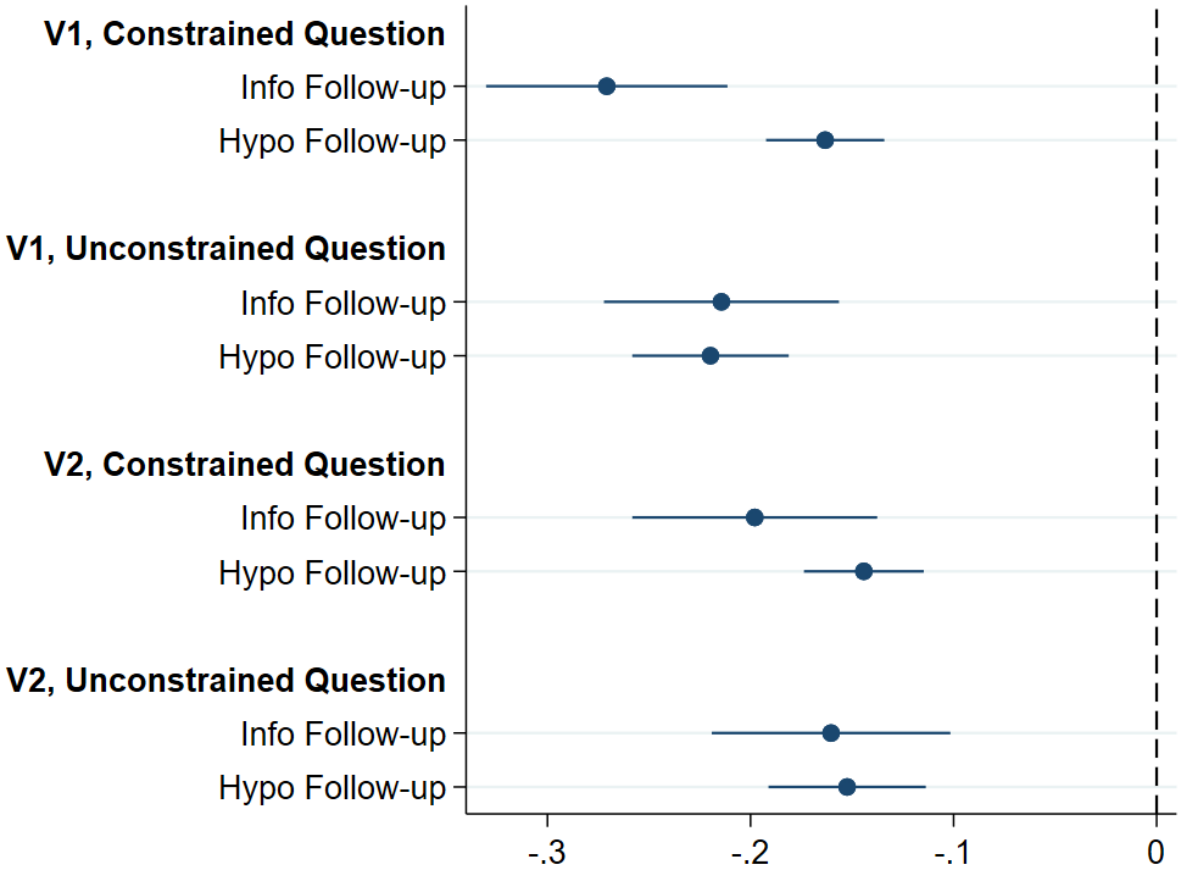
Note: This figure displays the results of regression specifications corresponding to Equation (3). Essentially, the specifications test whether the lack of updating in response to information is stronger for unfavorable information. This appears to be the case. All specifications depicted here use the V2 beliefs outcome, as the *Unfavorable* indicator is defined at the issue-by-individual level, like the V2 beliefs outcome

Figure 7: Effect of Information on Beliefs (Relative to Hypothetical) – Issue by Issue



Note: This figure displays the results of regression specifications corresponding to Equation (2), albeit issue by issue. While running these regressions issue-by-issue leads to a substantial power loss, there is significant evidence of motivated reasoning for the majority of individual issues.

Figure 8: Effect of Information on Beliefs – Follow-Up Survey (Relative to Hypothetical)



Note: This figure displays the results of regressions run on the follow-up survey sample (relative to the initial hypotheticals). Motivated reasoning appears to persist strongly – though it persists somewhat less strongly for those in the hypothetical arm who are asked the question on which they were initially constrained a week earlier (the question for which they were asked their hypothetical beliefs).