10-7-2024

Political Ideology, Emotion Response, and Confirmation Bias

David L. Dickinson

Received: 5 March 2024 | Accepted: 9 September 2024

DOI: 10.1111/ecin.13253

ORIGINAL ARTICLE

Economic Inquiry

Political ideology, emotion response, and confirmation bias

David L. Dickinson[1,2,3]

[1]Appalachian State University, Boone, North Carolina, USA

[2]IZA, Bonn, Germany

[3]ESI, Chapman University, Orange, California, USA

Correspondence
David L. Dickinson, Appalachian State University, Boone, NC 28607, USA.
Email: dickinsondl@appstate.edu

Funding information
Walker College of Business, Appalachian State University

Abstract

Motivated reasoning can serve to help resolve emotional discomfort, which suggests emotion as a likely moderator of such reasoning. This paper addresses a gap in the literature by examining emotion and confirmation bias in the political domain. Results from two preregistered studies, which involved over 900 unique participants, document a confirmation bias across distinct dimensions of belief and preference formation. Also, ideologically dissonant information significantly worsens self-reported emotion. With some exceptions, the evidence generally supports the hypothesis that negative emotion moderates the strength of the bias, which highlights the importance of emotion response in understanding and potentially counteracting confirmation bias.

KEYWORDS

cognitive reflection, confirmation bias, deliberation, emotion, motivated reasoning

JEL CLASSIFICATION

C91, D89, D91

1 | INTRODUCTION

This paper reports on two preregistered studies designed to examine how emotion moderates confirmation bias. Study 1 replicates existing evidence and contributes novel self-reported data on emotion states using a classic political confirmation bias task. Study 1 also administered a second novel task examining distinct emotion and confirmation bias effects on normative preferences. In Study 2, we examined factual political statements and analyzed Bayesian updating —noisy binary signals regarding a statement's veracity were either ideology-consonant, ideology-dissonant, or neither. A main contribution of this paper is to examine how one's emotional state responds to consonant versus dissonant information and how the response affects or moderates (i.e., influences) confirmation bias. Our viewpoint is that self-reported emotion can reveal cognitive dissonance, and thus emotion response is an important factor in understanding

Abbreviations: BDM, Becker–DeGroot–Marschak; *DissDegree*, degree of dissonance; ERQ, emotional regulation; GLS, generalized least squares; OSF, open science framework.

Managing Editor: Tim Salmon

beliefs, preference formation, and motivated reasoning. This viewpoint is consistent with a theoretical framework, which is outlined below, that captures both the deliberative and emotional inputs to confirmation bias.

2 | BACKGROUND

One of the likely contributing elements to a growing political divide is a form of motivated reasoning known as confirmation bias. As individuals preferentially expose themselves to information or assess political arguments in ways that align with their existing beliefs, polarization of viewpoints results. In general, decision biases result from lack of deliberation or attempts to economize on cognitive input to a decision. In contrast, confirmation bias has been shown to increase with deliberation or with a more reflective style of thinking. Both theory and evidence support this claim (Cotton, 1985; Dickinson, 2020; Festinger, 1957; Frey, 1986; Frimer et al., 2017; Jones & Sugden, 2001; Kahan, 2012; Knobloch-Westerwick et al., 2020). One theoretical approach suggests that this bias is a rational response to existing or anticipated cognitive dissonance, which is partly aimed at avoiding regret (Charpentier et al., 2018; Kobayashi & Hsu, 2019; Nicolle et al., 2011). Other frameworks that predict confirmation bias suggest that utility is derived from holding beliefs that support the validity of one's position (e.g., Drobner & Goerg, 2024; Gentzkow et al., 2015).

The experience of cognitive dissonance is associated with negative affective states (Van Veen et al., 2009), and so one's emotional response to new information may play a critical role in understanding this bias and how it may affect various components of preferences and belief formation. In short, confirmation bias is a form of implicit emotion regulation strategy. Evidence confirms that motivated reasoning produces neural activation patterns that differ from when there is no emotional stake in the issue (Westen et al., 2006), and so it is natural to consider that emotion response is a behavioral marker that helps identify the extent to which one's utility may depend on the validity of an ideological position.

Political issues are an important and emotionally charged domain for belief formation and information processing where confirmation bias has been previously documented (Allcott & Gentzkow, 2017; Bail et al., 2018; Bakshy et al., 2015; Bakshy et al., 2015; Bauer et al., 2022; Dickinson, 2020; Hart et al., 2009; Hill, 2017; Knobloch-Westerwick et al., 2020; Lord et al., 1979; Taber & Lodge, 2006). Public opinion research indicates that political ideology affects how individuals see the world,[1] and ideology itself is not a simple concept to define given that individuals may identify with a particular tribe (e.g., the liberal tribe, or the Republican tribe) but also view themselves along an ideological continuum.[2] The political persuasion literature suggests that targeting one's partisan or social identity is an effective way to modify one's attitude (Druckman, 2022). Political governance may differ from other professions where expert advice can most clearly help update beliefs regarding questions of a technical nature, such as skilled trades.[3] The somewhat less technical nature of political belief formation may be precisely what gives space for motivated reasoning to take hold more than in other arenas.

Misinformation afflicts political systems, and biased views of even factual information can be important in preserving one's ideological worldview (Jerit & Zhao, 2020). Thus, political psychology helps in our understanding of how directional goals are created and how political information is processed. Flynn et al. (2017) note that partisanship and prior opinion fuel directionally motivated reasoning, and political misperceptions are rooted in motivated reasoning. Others have suggested a dual-process theory of political beliefs that implies automatic processing, such as one's emotional or affective response, is important for understanding beliefs (Lodge & Taber, 2013). Indeed, emotion is a legitimate target for persuasion (see Druckman, 2022) precisely because it leverages the power of cognitive dissonance—it is often easier to keep beliefs aligned with one's emotional reaction rather than go through the cognitive effort to sort out whether the dissonant message or signal may justify changing one's beliefs.[4] The present study fits into this literature in its efforts to study how emotion response may help predict the extent to which information processing goes off-course.

Though we examine how ideology affects preference and belief formation, ideology may have other important and diverse downstream effects.[5] Our focus in this paper will be the updating of one's preferences and beliefs regarding political issues and facts in the face of arguments and new information. The closest study of which we are aware to ours is Bauer et al. (2022) who also study baseline beliefs, information exposure, and belief updating. While related, our methodologies differ and thus provide complementary evidence on these important questions. Importantly, our study contributes by reporting data on participant emotion states at key points, which allows us to examine the extent to which emotion states may be important in understanding these confirmation bias phenomena.[6]

3 | FRAMEWORK

A general framework can capture the idea of how belief distortion may arise if there is direct utility experienced by one's political ideology (or ideology in general). For example, in the spirit of Engelmann et al. (2024) and Drobner and Goerg (2024), one might consider that a component of utility is derived directly from one's belief regarding the validity of an ideological position, $\widehat{b} \in [0,1]$, which may differ from objective reality, b. However, there is also an instrumental utility component that values accuracy of beliefs—experimentally, this is implemented by means of an incentivized Becker–DeGroot–Marschak (BDM)-type elicitation procedure. Then, there is also a component that recognizes that belief distortion imposes a cognitive cost. Formally, utility that includes these three components may be written as follows:

$$U = \alpha\widehat{b} + \frac{1}{2}\left(1 + 2b\widehat{b} - \widehat{b^2}\right)P - \gamma\left(b - \widehat{b}\right)^2$$

Here, α is the direct utility given to $\widehat{b}$, P represents the utility payoff experienced due to belief accuracy (with functional form implied by the BDM elicitation procedure, which means this material utility component is maximized with full belief accuracy, $\widehat{b} = b$), and γ is the utility cost of belief distortion. Note that $\widehat{b} = b$ maximizes the utility value of belief accuracy and minimizes the cognitive costs of belief distortion. Maximizing utility with respect to $\widehat{b}$ leads to the following first-order condition:

$$\frac{\partial U}{\partial \widehat{b}} = \alpha + bP - \widehat{b}P + 2\gamma b - 2\gamma\widehat{b} \equiv 0$$

From this, one's optimal choice of beliefs relates to objective reality, b, in the following way:

$$\widehat{b^*} = b + \frac{\alpha}{P + 2\gamma}$$

Thus, the convergence of beliefs to reality occurs as the payoffs to belief accuracy increase or as the cognitive costs of distorting reality increase. However, it is also clear that a direct utility component to one's belief will lead one to subjectively inflate the validity of one's ideological position (i.e., $\widehat{b^*} > b$). Drobner and Goerg (2024) consider this very explicitly in a Bayesian belief-updating task,[7] which is similar to our Study 2 task, but such a framework can also be useful to understand how a utility-motivation to one's ideological belief will produce a general belief distortion that can affect a range of outcomes. This is not to say that all scenarios fit this framework neatly, but because beliefs matter for preferences and choices, it is useful to consider how such distortions may arise. Regarding the specific focus of this paper, we consider that α is reflected in one's emotion response to belief consonant or dissonant arguments, information, signals, etc. A stronger emotion response in ideology-relevant situations is consistent with a larger (unobservable) direct utility component of beliefs, α, regarding the validity of one's ideological position.

4 | METHODS (COMMON TO BOTH STUDIES 1 AND 2)

Both studies were administered online using the Prolific platform (Palan & Schitter, 2018; Peer et al., 2017) to recruit participants. Emotion reports were elicited at key points and included several dimensions of positive (happy, excited, surprised, satisfied) and negative (angry, irritated, confused, regret, disgust) emotion. A main hypothesis tested in each study (and each task) is that ideologically dissonant information will elicit a relatively more negative emotion response in the participant. We also hypothesized whether the relative negative versus positive emotion response to dissonant statements or information signals can predict the extent to which preferences or beliefs are updated (or not). We used the Open Science Framework to preregister plans for both studies (recruitment approach, sample sizes, hypotheses, and analysis plans) prior to any data collection (Study 1 at https://doi.org/10.17605/OSF.IO/YHVSB; Study 2 at https://doi.org/10.17605/OSF.IO/NQRJD).

Both studies were reviewed and approved by the author's human subjects review board. For Study 1, we recruited roughly equal numbers of self-identifying political conservatives and liberals residing in the U.S. who were at least 18 years of age. Study 2 recruited first from those who had participated in Study 1, and then the Study 2 sample size was increased as specified in the preregistration plan. All participants were compensated a fixed amount for participation in either online study, which was consistent with the Prolific fair-pay standards. The subjective preference measures elicited in the Study 1 tasks were not amenable to additional variable pay incentives for different decisions made. Study 2, however, examined beliefs regarding the truthfulness of factual statements, and so it provided additional compensation to participants based on an incentive compatible belief elicitation mechanism describe later in the Study 2 methodology. Full survey details for Studies 1 and 2 are given in Supporting Information S4: Appendix C. In what follows below, we organize the paper by examining details of Studies 1 and 2 separately prior to a general discussion of our overall key findings.

5 | STUDY 1: EMOTION AND PREFERENCE EFFECTS OF CONSONANT VERSUS DISSONANT MESSAGES

In study 1, a sample of $n = 650$ U.S. participants who self-identified as politically conservative ($n = 327$) or liberal ($n = 323$) were recruited on the Prolific platform, with $n = 611$ (females: $n = 306$, 50%) passing the attention checks (conservatives: $n = 302$, 49%). The Study 1 survey administered two distinct tasks: Task 1 was a classic confirmation bias task aimed at replicating Taber and Lodge (2006), which measured self-selected exposure to information sources and perceived strength of arguments regarding the issue of "gun control." We then modified Task 1 to elicit reports regarding one's emotional state at key junctures to test novel hypotheses related to emotion response. A distinct and original second task (Task 2) elicited beliefs and preferences regarding policy importance and priorities on 12 distinct policy issues. For each issue, preference elicitation was followed by an information statement, which was randomized to be either aligned or misaligned with one's current beliefs or preferences. Emotions were elicited after the information statement, and then preferences were re-assessed. This allowed us to test the emotion moderation hypothesis in a distinct domain of subjective preference measures across a wide set of policy issues.[8]

5.1 | Study 1: Survey methods

The study first administered Informed Consent, followed by demographic and political preference measure questions, the Emotional Regulation Questionnaire (ERQ) (Gross & John, 2003), and Likert scale baseline emotion state measures for each of several positive and negative states (happy, excited, surprised, satisfied, angry, irritated, confused, regret, disgust). From the ERQ, we specifically considered cognitive reappraisal style in one of our hypotheses below. Cognitive reappraisal is a proactive emotion regulation strategy where one changes how one thinks about a situation to regulate its emotional impact, rather than actively suppressing it (Gross & John, 2003), and it is generally considered a healthy form of explicit emotion regulation. Following elicitation of these measures we administered two tasks aimed at the assessing confirmation bias and emotion moderation effects. The first task (Task 1) has been studied previously in the literature, but a novel Task 2 was also administered in Study 1 to extend and complement our understanding of confirmation bias.

Task 1 of study 1 was a classic confirmation bias task (Taber & Lodge, 2006), which includes two dimensions along which one may express a confirmation bias. We focused our attention on the issue of "gun control," and the first portion of the task was a selective information exposure task where participants were required to view six statements or arguments on the issue of gun control. Participants were allowed to select viewing statements from a set of four sources, two of which were clearly on each side of the partisan divide.[9] A confirmation bias test examines whether one selectively samples more statements aligned with one's political ideology or not. For this portion of Task 1, however, we did not elicit emotion reports following the selection or viewing or these statements. In a second portion of Task 1, participants were all required to view a standard set of six arguments on the issue of gun control: three arguments favored gun control and three arguments were against gun control. Following each argument, participants rate on a 9-point scale how weak or strong they felt the argument was, as distinct from whether they agreed with the argument. Participants also gave emotion reports following each argument-strength rating. The emotion reports included with the argument-strength portion of Task 1 allows for us to test the emotion moderation hypothesis in this classic confirmation bias task, and so we focus on this argument-strength dimension of Task 1.

After Task 1 (i.e., the classic confirmation bias task), the survey presented participants with Task 2 (i.e., the novel "information manipulation" task) that examined 12 current political issues in randomized order: environment protection, drug addiction, COVID-19, gun rights, crime reduction, climate change, border security, federal budget deficits, rights of illegal immigrants, police brutality, and ethics in government. Task 2 was included within Study 1 because it was intended to complement Task 1 in the sense that it also examines subjective preferences in the political domain. Yet, it extends beyond Task 1 in two important ways: First, it examined preferences across a variety of policy issues, as opposed to just a singular issue as was done in Task 1. Secondly, it introduced a randomized information treatment with the intent of identifying causal impacts of ideologically consonant or dissonant information on these subjective preference ratings.

In Task 2, two questions were asked of each issue: "On a scale of 0–100, please tell us your view of *how big of a problem POLICY ISSUE X is* (0 means not a problem at all, 100 means huge problem, in your opinion)"; "On a scale of 0–100, please tell us your view of *how many resources the Federal Government should be devoting, out of its limited resources (people, budget funds), compared to what it is devoting* to address this problem of *POLICY ISSUE X* (0 means the government *should* be devoting a lot fewer, and 100 means the government *should* be devoting a lot more resources, in your opinion)." Thus, the preference measures elicited captured the participant's perception of how big of a problem the issue is and her preferred resource prioritization for that issue. Because the outcomes measures are explicitly personal preferences, the tasks were not amenable to variable payoff incentives and participants were compensated a flat rate in Study 1. We will complement the results from Study 1 with Study 2, which we present later in this paper.

Critical to Task #2 of Study 1 was a randomized information treatment. After preference elicitation on each policy issue, a randomized information statement was displayed that either amplified ($Amplify = 1$ treatment) or minimized the issue's importance ($Amplify = 0$ treatment)—see Supporting Information S4: Appendix C for details on all information statements.[10] Following the information treatment, the respondent self-reported emotion states, and policy views/preferences were re-assessed.[11]

5.2 | Study 1: Hypotheses

Hypothesis 1 (task 1). When presented belief-dissonant arguments in the classic confirmation bias task, participants will report relatively more negative emotion states.

Hypothesis 2 (task 1). Negative emotion states will predict a stronger classic confirmation bias in perceived argument strength.

Hypothesis 3 (task 2). When viewing information dissonant with one's policy viewpoints, participants will report relatively stronger negative emotion states.

Hypothesis 4a (task 2). Dissonant information will have a lesser effect than consonant information on updating one's policy preferences.

Hypothesis 4b (task 2). Negative emotion will moderate the confirmation bias relationship between information and policy preferences.

Hypothesis 4c (task 2). A stronger cognitive reappraisal (emotion) regulation style will lessen the effect of negative emotion on the discounting of belief-dissonant information.

5.2.1 | Study 1: Key variables and analysis

For brevity, we describe key variables and their scales here, and complete details are given in Supporting Information S2: Appendix A regarding the specific construction of each variable. Any other variable constructed for purposes of hypotheses tests will be described in discussing the results specific to that hypothesis test.

Key independent variables used for analysis of Tasks 1 and 2 outcomes in Study 1 were the following. *Liberal Score* $\in [1, 9]$, which measured one's self-reported political ideology from 1 = Very Conservative to 9 = Very Liberal

(5 = Middle-of-the-Road). Emotion or *Mood* (as we will use for variable names, for simplicity) is always operationalized as a composite of self-reports along the nine dimensions mentioned previously (happy, excited, surprised, satisfied, angry, irritated, confused, regret, disgust). Each emotion state was reported on a 7-point scale, where 1 = zero level of this emotion and 7 = maximum level of this emotion. The average of the 4 positive and, separately, 5 negative affective states were used as a *Positive Mood* and *Negative Mood* composite measures, respectively. As noted previously, our use of the term "mood" in these variable names is for expositional purposes and not meant to imply we are capturing long-lived dispositions of the participant (see footnote 6). The difference in these compositive measures was also used as a "net mood" (i.e., net emotional state) measure that we call *Relative Negative Mood* = *Negative Mood* − *Positive Mood*, which is defined in the [−6, +6] interval. Emotion reports at key points in Task 1 and Task 2 were used to test the hypothesis that emotion moderates the confirmation bias. These emotion reports were also taken at the beginning of the Study as a baseline control for what emotion-state one brings to the experiment, for which we control in the statistical analysis.

Additional variables of interest for both tasks in Study 1 included *Control Variables* for age, female (indicator), minority (indicator), and education level (See Supporting Information S2: Table A1 for their full descriptions and summary data). Some empirical specifications also included *Additional Controls* variables for one's cognitive reappraisal style, political preference and characteristic measures, self-reported sleep (shown to affect confirmation bias, see Dickinson, 2020), and baseline emotion states elicited at the start of the study. These independent variable measures were elicited in the Study 1 survey prior to administration of the relevant decision task.

5.2.2 | Task 1—Classic confirmation bias task: Key variables

In Task 1 of Study 1, *Pro-Gun Views* ∈ [−24, +24] assessed one's viewpoints on the issue of gun control. *Anti-Gun Mood* ∈ [−18, +18], examined one's overall emotion elicited in direct response to a set of arguments regarding gun control (see Supporting Information S2: Figures A1 and A2 for the distributions of *Liberal Score* and *Anti-Gun Mood* in the Study 1 sample). A key outcome variable for Task 1, *Perceived Argument Strength* ∈ [−24, +24], was constructed such that higher (lower) values indicate that one viewed arguments favoring gun rights (gun control) as stronger.

5.2.3 | Task 2—Information manipulation task: Key variables

Key independent variables for Task 2 included the following. *Amplify* (=0 or 1), which was the indicator variable identifying the information statement treatment. The variable, *Degree of Dissonance* ∈ [0, 100], describes the level of issue-specific dissonance between *Preference Ratings* measure and randomly selected information statement received. Specifically, *Degree of Dissonance* = *Preference Rating* when the information treatment minimizes the importance of the policy issue (*Amplify* = 0), and *Degree of Dissonance* = 100 − *Preference Rating* when *Amplify* = 1. The key outcome measure in Task 2 was *Preference Ratings* ∈ [0, 100], which was a subjective report given on each of the two dimensions described above for this novel preference measure task (i.e., issue importance and resource prioritization preference).

All estimations testing the Study 1 hypotheses included sensitivity analysis with full results shown in Supporting Information S2: Appendix A. Specifically, simple or binary regressions were complemented with specifications that included demographic and additional controls. We summarize the key results by way of coefficient plots and relegate the full estimation results to Supporting Information S2: Appendix A for brevity. The pre-registration plans identified several other a priori hypotheses that are not as relevant to the focus of this paper. These additional preregistered hypotheses, which were mostly supported by the data, are described and test results given in Supporting Information S3: Appendix B.

5.3 | Study 1: Results

Our sample of $n = 611$ participants from Study 1 includes $n = 307$ females (50%), $n = 302$ self-reported Republicans (49%), and the average age of the participants was 34.11 years (±13.56 years standard deviation, range 18–72). *Relative Negative Mood* ∈ [−6, +6] (defined above) at baseline for the Study 1 sample was −1.76 (±1.71, range −6 to +3.6), which implied an average net positive emotion state prior to the administration of the decision tasks.[12]

5.3.1 | Tests of H1–H2 (classic confirmation bias task)

Results of our H1 test are shown in Figure 1, which shows robust support for the hypothesis. We estimated linear regression models using two distinct controls for one's views on gun control. In each instance, we also estimate models with and without additional controls and summarize results with coefficient plots (Figure 1—full results of all estimations are in Supporting Information S2: Table A3). The dependent variable used in the H1 test, *Anti-Gun Mood,* captures the effect of dissonant arguments on viewpoints regarding gun control by combining the relative negative-over-positive affective state reported after viewing pro-gun arguments with the relative positive-over-negative affective state reported after viewing anti-gun arguments (i.e., reverse scoring the net emotion state measure as needed). That is, *Anti-Gun Mood* captures the combined argument-consonant positive emotion and argument-dissonant negative emotion states reported by the participant so that *Anti-Gun Mood* reflects a singular coherent composite measure.

Estimation results shown in Figure 1 coefficient plots indicate the viewpoints favoring gun control (i.e., lower levels of *Pro-Gun Views*) and stronger liberal political ideology (i.e., higher values of *Liberal Score*) predict a significantly higher level of *Anti-Gun Mood*. This effect is robust to the inclusion of control variables. In other words, we find a consistently estimated effect that one's viewpoint (or ideology, which proxies for viewpoints on political issues) promotes an emotion response to support that viewpoint. This supports H1 and establishes the basis for the emotion-mechanism hypotheses that follow.

Hypothesis 2 tests whether emotion moderates the degree of confirmation bias.[13] Figure 2 summarizes these H2 results via coefficient plots, with full results in Supporting Information S2: Appendix A. Evidence for a general confirmation bias result would be a statistically significant negative effect of *Liberal Score* on *Perceived Argument Strength*, which we find ($p < .01$ in all instances, as seen in Supporting Information S2: Table A4). On the left-panel in Figure 2 we see that, holding ideology constant, the degree of one's *Anti-Gun Mood* response predicts a main effect of a lower perceived strength of pro-gun arguments in the classic confirmation bias task ($p < .01$ in all instances). The right-panel coefficient estimates on the interaction term *Anti-Gun Mood* × *Liberal Score* test whether emotion moderates the estimated confirmation bias, with the negative coefficient estimates supporting H2 ($p < .01$) because they indicate that the main emotion effect is magnified for those more liberal participants.

5.3.2 | Tests of H3–H4 (information manipulation task—A novel confirmation bias task)

These tests utilize the panel data set created from Task 2 of Study 2. Recall that participants assessed 12 distinct policy issues before and after a randomized information treatment that provided dissonant or consonant information on the issue, and dissonance was measured by *Degree of Dissonance* defined above. To establish the connection between this component of the study and the previous gun control confirmation bias task, we first tested H3 to evaluate the hypothesis that information more dissonant with one's perceived importance and prioritization of the issue would significantly increase one's relative negative emotion state (as was supported by results above with H1). Here, the dependent variable in the regressions is defined as one's average *Relative Negative Mood* (i.e., subtracting average positive from negative emotion dimensions reports) after viewing the informational statement randomly shown for that political issue. We have 12 observations per participant, and the models estimated are random effects generalized least squares (GLS) models with errors clustered at the participant level. Because we define the dependent variable as a relative *negative* emotion report, H3 is supported if the estimated coefficient on *Degree of Dissonance* is positive. As shown in Figure 3 coefficient plots, in all cases the data support H3 with respect to both preference dimensions ($p < .01$ in all cases).

Hypothesis 4 is evaluated next, where we again use random effects GLS estimations with standard errors clustered at the participant level. The dependent variable is the log of one's post-treatment preference rating (*Perceived Problem* or *Resource Prioritization*), which we model as a function of the log of one's pre-treatment preference rating, an indicator for the information statement received (*Amplify*), the *Degree of Dissonance* of the information statement with one's preference rating, and an interaction between *Degree of Dissonance* and *Amplify*. Preference ratings of 0 on the 0–100 scale were replaced with "1" prior to the log transformation. The baseline specification, to which we then add control variables for the sensitivity analysis is:

$$\log\left(Rating_{post}\right) = \alpha + \beta_1 \log\left(Rating_{pre}\right) + \beta_2 Amplify + \beta_3 Dissonance\ Degree + \beta_4 (Dissonance\ Degree \times Amplify) + error \quad (1)$$

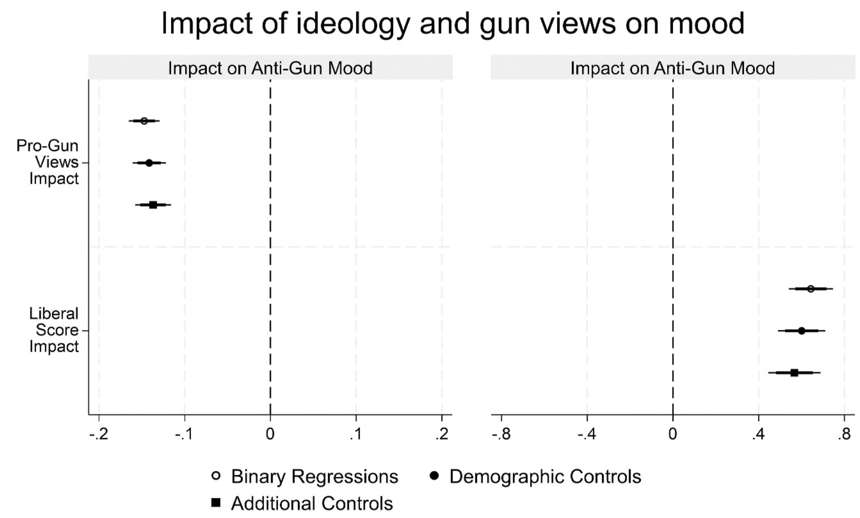

FIGURE 1 Study 1 Hypothesis 1 test—emotion effect of partisan arguments (see Supporting Information S2: Table A3 for full estimation results). Thick (thin) lines show the 95% (99%) confidence interval around the point estimates for the 1-tailed preregistered hypotheses regarding the effect of one's *Pro Gun Views* (left side) or *Liberal Score* (right side) on her Anti-gun mood measure. The *Anti-gun Mood* measure is an average of one's net anti-gun emotion state averaged across the 3 pro-gun and, separately, the 3 anti-gun arguments to create a singular metric that averages one's net negative emotion state after seeing pro-gun arguments and one's net positive emotion state after seeing anti-gun arguments. As defined, the coefficient estimates all support H1 by indicating that *Anti-gun Mood* increases in one's degree of liberalism but decreases in one's degree of pro-gun views. Note the different *x*-axis scales used due to the different ranges of the dependent variables used—*Pro-Gun Views* $\in [-24, +24]$, *Liberal Score* $\in [1, 9]$.

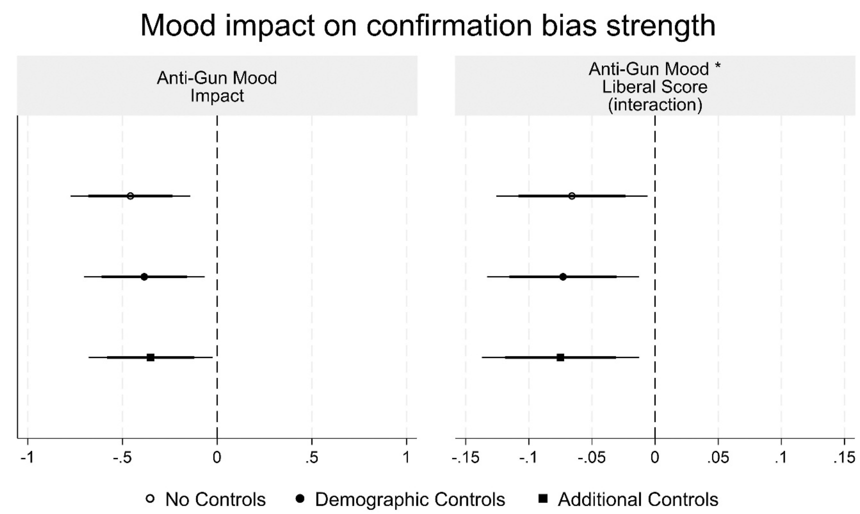

FIGURE 2 Study 1 Hypothesis 2 test—emotion moderation of confirmation bias on perceived argument strength (see Supporting Information S2: Table A4 for full results). Thick (thin) lines show the 95% (99%) confidence interval around the point estimates for the 1-tailed preregistered hypotheses regarding the effect of one's relative *Anti-Gun Mood* and the *Liberal Score* × *Anti-Gun Mood* interaction on perceived argument strength. Note the different *x*-axis scales in figure. The main *Liberal Score* effect (not shown above) was negative and statistically significant across all models ($p < .01$ in all instances), which shows clear evidence of a confirmation bias effect whereby one's political ideology affects how strong one views the strength of a dissonant argument.

Here, we use log(*Rating*) given that *Dissonance Degree* is constructed from the raw $Rating_{pre}$ measure as noted above.[14] Figures 4–6 show the results of our H4a–H4c tests in the form of coefficient plots of the key coefficient estimates that test H4, β_3 and β_4 (see Supporting Information S2: Appendix A for full results).[15] In Figure 4, the positive and statistically significant coefficient estimates on the main *Dissonance* measure, β_3, along with the interaction term

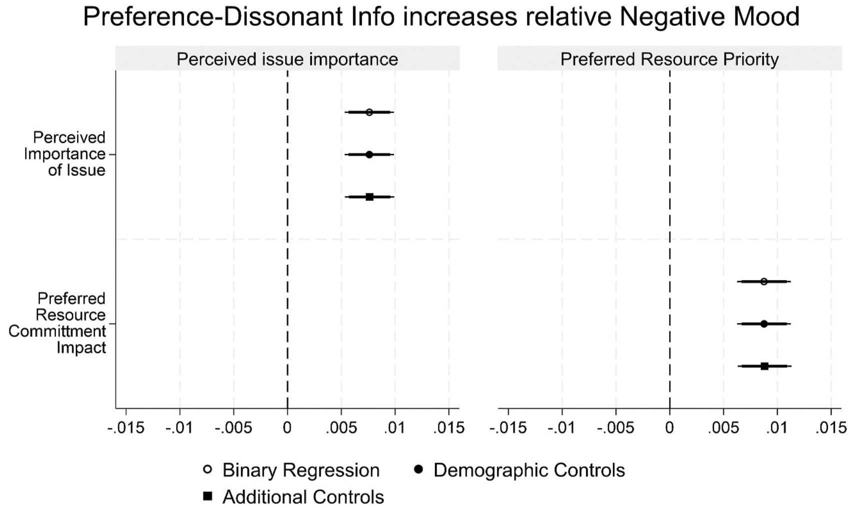

FIGURE 3 Hypothesis 3 tests (panel data estimations) (see Supporting Information S2: Table A5 for full results). The panel data set is comprised of 12 observations per participant capturing emotion and preference measures on 12 distinct political issue ($n = 7332$ for the estimations, with robust standard errors clustered at the subject level). The thick (thin) lines show the 95% (99%) confidence interval around the point estimates for the 1-tailed preregistered hypotheses regarding the effect of the dissonance information, *DissDegree* $\in [0, 100]$, on *Rel Neg Mood* after seeing the information. Perceived importance of the issue above defines *DissDegree* with respect to the preference measure that asks about how big of a problem one feels the issue is.

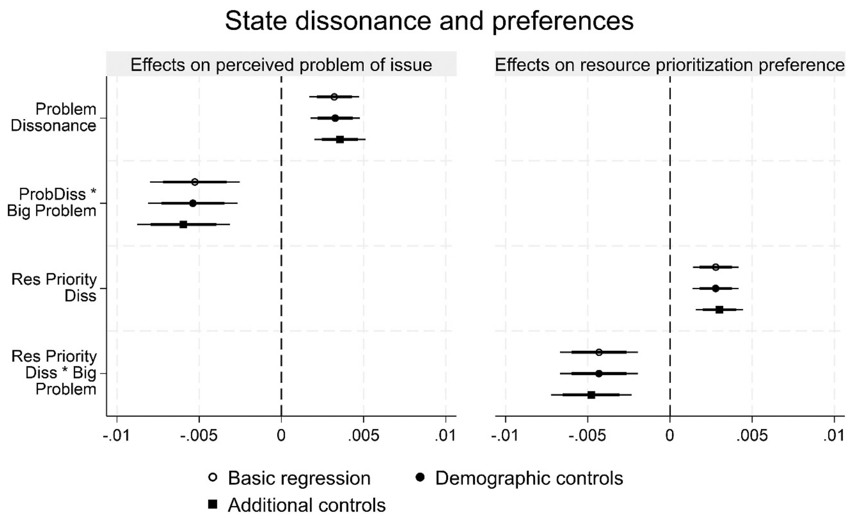

FIGURE 4 Hypothesis 4a test (see Supporting Information S2: Table A6 for full results). The thick (thin) lines show the 95% (99%) confidence interval around the point estimates for the 1-tailed preregistered hypotheses. Dependent variable for the estimations is the log of the preference rating (perceived *Size of Problem* of the issue [Left panel] or *Preferred Resource Prioritization* for the issue).

coefficient, $\beta_{4,}$, have the following interpretation. When *Amplify* $= 1$, the combined effect of increased statement *Dissonance* and the *Amplify* $\times$ *Dissonance* interaction is negative. This implies a marginal reduction in one's preference rating that is separate from the main positive effect of *Amplify* $= 1$ (not shown in Figure 4). The combined β_3 and β_4 effects are significantly different from zero (β_2 test, $p < .01$ in each instance) in the direction of the hypothesis (i.e., $\beta_3 - \beta_4 < 0$ is the alternative hypothesis). That is, we estimated that one's preference ratings update to a lesser degree when ideology is more dissonant with an *Amplify* $= 1$ statement.[16] This result supports H4a that updating preference measures occurs to a lesser degree for dissonant statements, though the effect sizes are small. The magnitude of the

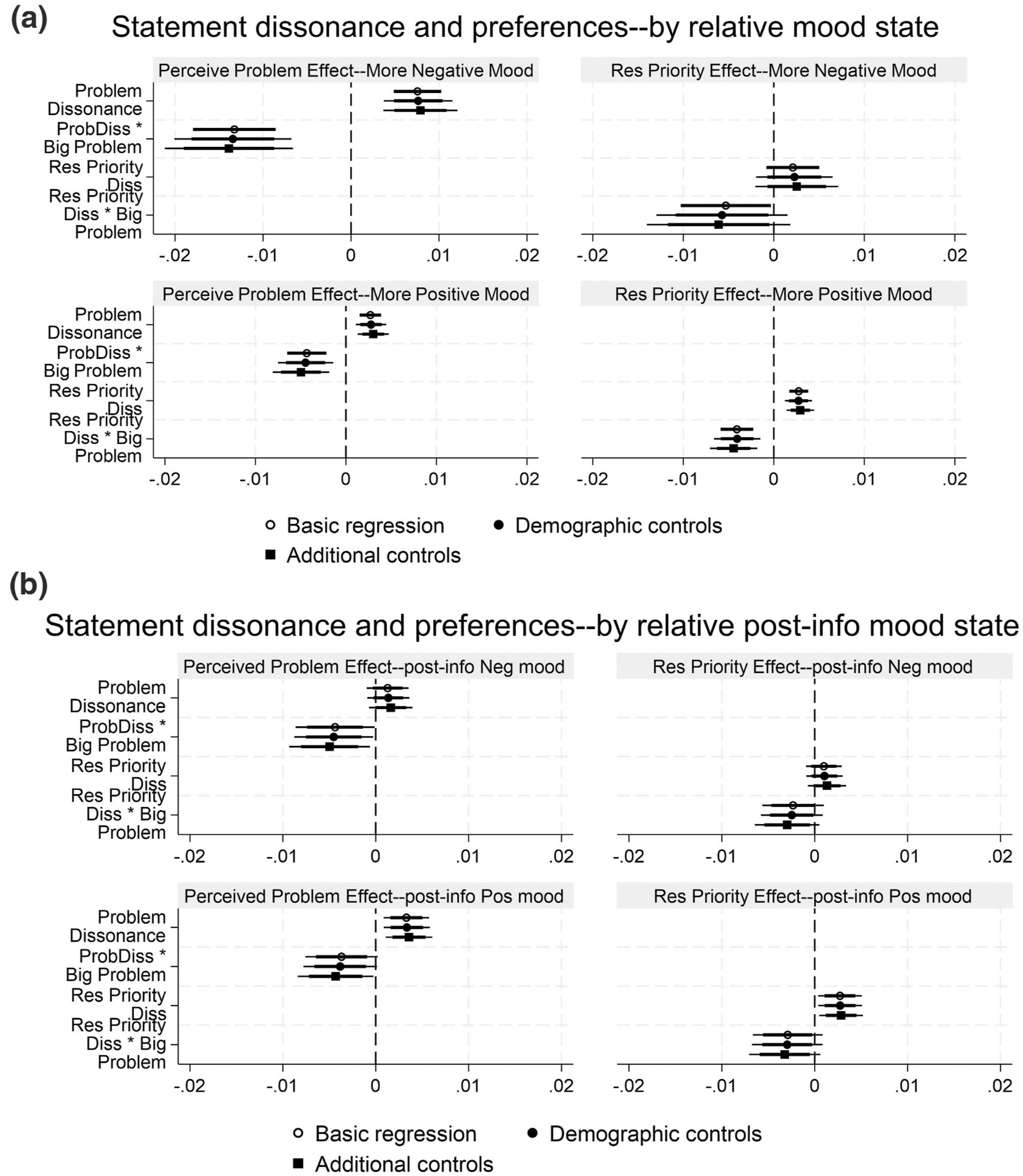

FIGURE 5 Hypothesis 4b test. Panel (a): By baseline emotion state (see Supporting Information S2: Tables A7 and A8 for full results). The thick (thin) lines show the 95% (99%) confidence interval around the point estimates for the 1-tailed preregistered hypotheses. The dependent variable for the estimations is the log of the preference rating. Trials where positive > negative baseline mood: $n = 6216$ (518 participants). Trials where negative > positive mood: $n = 912$ (76 participants). We excluded $n = 18$ participants where baseline positive = negative mood. Panel (b): By post-treatment emotion state (see Supporting Information S2: Tables A9 and A10 for full results). Trials where positive > negative mood post-info: $n = 3219$ (562 participants). Trials where negative > positive mood: $n = 3349$ (532 participants). Unlike with baseline emotion state, with post-information emotion state a given participant may present trials with both relatively negative and relatively positive emotion responses, depending on the political issue for that trial. We excluded $n = 803$ trial-level observations (10%) from this analysis where post-treatment positive = negative emotion state measures.

coefficients in the log-linear regression are such that in the left panel of Figure 4, for example, a statement that has a 30-unit higher *Dissonance Degree* predicts an approximate 9%–10% less updating.[17]

Hypotheses H4b and H4c evaluate whether H4a is moderated (i.e., increased) by one's negative emotion state (H4b) or the lack of cognitive reappraisal (H4c). The results in Figure 5 offer qualified support for H4b. In the panel a results of Figure 5, the left-side compares the same β_3 and β_4 estimates of the *Perceived Problem* preference model for the subset of those reporting more negative emotion state (top-left of panel Figure 5a) compared to more positive emotion state (bottom-left of panel Figure 5a), where emotion state is defined by one's *baseline* emotion state measures. For the

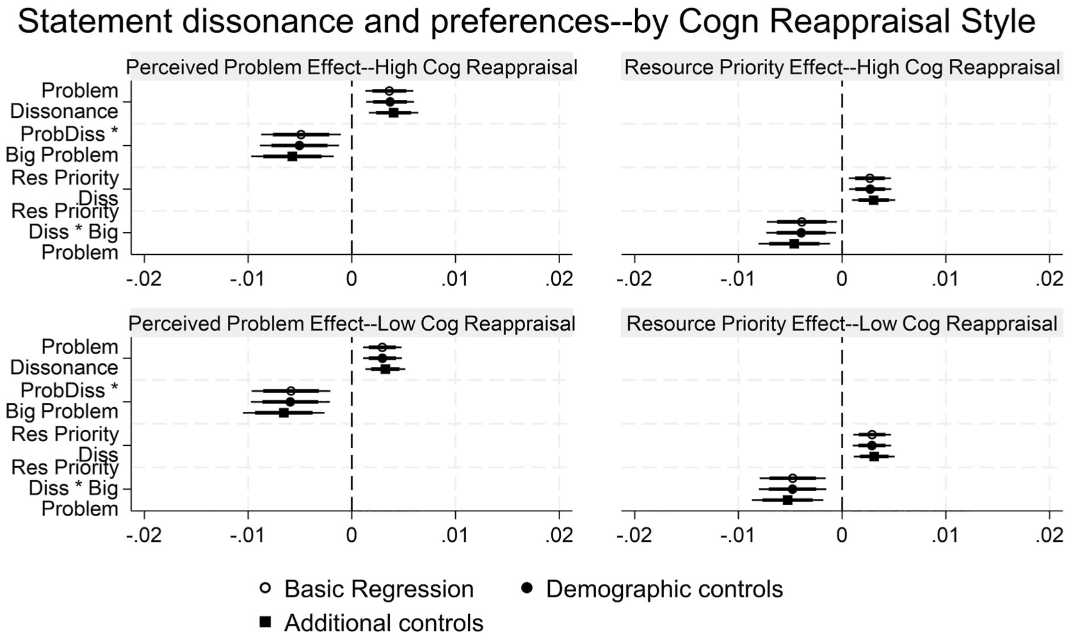

FIGURE 6 Hypothesis 4c test (see Supporting Information S2: Tables A11 and A12 for full results). The thick (thin) lines show the 95% (99%) confidence interval around the point estimates for the 1-tailed preregistered hypotheses. Dependent variable for the estimations is the log of the preference rating (perceived *Size of Problem* of the issue [Left panel] or *Preferred Resource Prioritization* for the issue). There was $n = 3348$ trial-level observations of below-median cognitive reappraisal (low cognitive reappraisal: on $n = 279$ clusters at level of participant) and 3984 observations above the median (high cognitive reappraisal: on $n = 332$ participant clusters).

Perceived Problem measure, we find that those with more negative baseline mood/emotion state have a stronger moderation effect of dissonant information than those with more positive baseline mood/emotion state. The panel a (Figure 5) comparison in the upper-right and lower-right shows the same results for the *Resource Priority* measure, and here we fail to estimate a significant difference. So, H4b is supported for one of the preference measures in this task when using baseline emotion state. However, in panel b, which defines emotion using the *post-information statement* emotion state reports, the results fail to support H4b. Finally, test results in Figure 6 also fail to support H4c that cognitive reappraisal style moderates the confirmation bias predictions. Thus, H4a is supported, but failure to support H4b and H4c tests leads us to conclude that dissonant information causes the same degree of reduced preference updating regardless of emotion state or explicit emotion regulation style.

5.4 | Study 1: Discussion

Overall, our Study 1 results supported the preregistered hypotheses regarding dissonant information and relative emotion state. We also found support in Task 1 that negative emotion predicts (i.e., moderates) a stronger confirmation bias. Our novel information manipulation task (Task 2) allowed us to examine confirmation bias along two preference dimensions, and we found support for the hypothesis that the dissonant information promotes a more negative emotion state. Furthermore, our data supported the hypothesis that dissonant informational statements have a lesser effect on the updating of one's preference ratings on a variety of political issues. This Task 2 finding therefore contributes novel evidence to complement our understanding of confirmation bias by documenting its effect in limiting the extent to which one updates preferences in the direction of new information. The emotion moderation hypothesis was not robustly supported in the Task 2 data—support here depended on the preference measures considered and the use of baseline emotion, rather than post-statement emotion (or emotional response), to test the hypothesis. Finally, we generally found no support for the preregistered hypothesis that those who explicitly regulate emotion via cognitive reappraisal would display a lesser confirmation bias in Task 2.

One critique of the novel information updating Task 2 is that it examines preference ratings along dimensions that may not be clearly defined. This critique also applies to the classic confirmation bias Task 1 regarding its use of perceived argument strength as a key outcome measure. Subjective ratings are, of course, not necessarily comparable across participants, and one may have little incentive to report sincere ratings. However, it is worth highlighting that our pattern of results would not be expected unless bias occurs in policy issue preference ratings (or perceived argument strength assessments) *and* emotion reports in systematic fashion.[18] The panel nature of the Task 2 data set also implies that the estimation results leverage within-subject variation in the data, which means the randomly assigned information treatment effects reported benefit from this within-subject variation.

Preference ratings differ from probability estimates and so Task 2 of Study 1 does not lend itself to a more proper Bayesian updating analysis. Study 2 (discussed next) was aimed at addressing these concerns to provide further complementary evidence on the question of emotion and confirmation bias. That is, armed with evidence from Study 1 regarding confirmation bias and emotion effects in environments where subjective preferences cannot be assessed against any objective standard, Study 2 further complements our research by examining incentive compatible probability judgments in a factual domain where confirmation bias implies objective harmful effects on belief-updating.

6 | STUDY 2: EMOTION AND INFORMATION SIGNAL IN A BAYESIAN TASK

In Study 2, a sample of $n = 503$ U.S. participants (females: $n = 294$, 58%) who self-identified as politically conservative ($n = 266$, 53%) or liberal ($n = 237$) were recruited on the Prolific platform.[19] We preregistered plan to recruit from the set of Study 1 participants, but were only able to recruit $n = 193$ ($n = 111$ Conservative) return participants for Study 2. Preregistration plans then specified to recruit new participants to reach a total sample of $n = 500$. The new Prolific participants were balanced across self-reported political party ($n = 310$ total, $n = 155$ Republican, $n = 155$ Democrat). While there is not perfect alignment between Republican (Democrat) and conservative (liberal) ideology identification, recall that our key analysis is based on a separate *Liberal Scale* measure that is common to all participants (and identical to Study 1).[20]

The survey flow for Study 2 included a Consent page, followed by baseline Likert scale emotion reports. Political ideology was then elicited using the *Liberal Score* measure described previously. Instructions were then given for the Bayes task, followed by an attention check, and then participants were administered the Bayes task (discussed more below). The survey ended with final emotion state reports after viewing images of Donald Trump and Joe Biden (see Study 1).

Notably, because the original intent was to recruit up to 500 participants for Study 2 from the original set of over 600 participants in Study 1, demographics or preference measures were not collected for Study 2. As a result, we do not have as robust a set of control variables for our estimation models of Study 2 data, other than age and sex.[21] Notably, Study 2 complements the Study 1 evidence by examining the core questions (i.e., confirmation bias, emotion response, and emotion moderation) in the arena of belief formation regarding political factual statements. Thus, a critique regarding the subjective outcome measures used in the Study 1 tasks is avoided in Study 2 using an incentive compatible belief-elicitation procedure that was not possible in the Study 1 tasks.

6.1 | Study 2: Survey methods

The task administered for Study 2 followed the design in Hill (2017), which elicited incentivized beliefs regarding the truthfulness of several factual political statements after the presentation of noisy signals regarding the statement's truth. The set of statements balanced factually true and false statements, and whether the statement's truthfulness or falseness would benefit conservatives or liberals (see Table 1).[22] Probabilistic beliefs regarding the truthfulness of each statement were elicited using an incentive compatible crossover scoring mechanism.[23] As in Hill (2017), for each statement a baseline belief was elicited, followed by the presentation of a noisy signal that was accurate 75% of the time, on average, and this was common knowledge. Beliefs were elicited after the noisy signal and this process repeated for a total of 4 noisy signals and posterior belief reports.

Practice trials, which were not incentivized, helped ensure a participant was familiar with the belief elicitation process and how the sequence of noisy signals would be presented. The practice statement was unrelated to political facts—it was a statement regarding the average temperature on Mars. After the practice trials, the Bayesian elicitation task was administered for each of the 4 statements, which were presented in randomized order, such that we have a panel with 16 observations (4 per statement) of posterior beliefs per participant that can be modeled as a function of the prior trial belief and the observed signal. Our design innovation was that emotion reports were elicited upon first

TABLE 1 Factual political statements (study 2).

Statement	Statement text	Truthfulness & ideology effect
Practice trials	The average temperature on Mars, as a whole, is −81°F	TRUE: Benefits no one
1	Median household incomes grew, and both the poverty rate and jobless rate of racial minorities fell from 2017 through 2019 while Donald Trump was president	TRUE: Benefits conservatives (its truthfulness)
2	There were more U.S. COVID-19 deaths during the first 9 months of the pandemic when Donald Trump was president than there were in the first 9 months of the Joe Biden presidency	FALSE: Benefits conservatives (its falseness)
3	A careful analysis of the 2020 U.S. Presidential election data found no evidence that systematic voter fraud harmed incumbent President Trump and helped elect Joe Biden to the presidency	TRUE: Benefits liberals (its truthfulness)
4	Between 2009 and 2016 when Barack Obama was president, European countries were less confident that the U.S. would do the right thing regarding world affairs than between 2001 and 2008 when George W. Bush was president	FALSE: Benefits liberals (its falseness)

Note: See supplemental information for sourcing on each statement.

viewing each political statement, as well as after the 4 trials for that statement. In each trial, the participant was reminded of his/her prior-trial probability estimate and 20-s timer counted down for timely responses (see Hill, 2017, and Supporting Information S4: Appendix C that shows the Study 2 experiment instructions).

6.2 | Study 2: Hypotheses

Hypothesis 1. Beliefs will update in the direction predicted by Bayes rule.

Hypothesis 2. Confirmation Bias presence.

Hypothesis 2a. Baseline beliefs will reflect one's political ideology.

Hypothesis 2b. New information signals not favorable to one's political ideology will predict a lesser level of belief updating compared to favorable signals.

Hypothesis 3. Dissonant statements (re: political ideology) will elicit more negative emotion.

Hypothesis 4. Additional dissonant noisy signals will increase one's relative negative emotion.

Hypothesis 5. More negative emotion post-statement will predict the degree of belief-updating.

6.2.1 | Study 2: Key variables and analysis

We briefly describe key variables of Study 2 here, with full details given in Supporting Information S2: Appendix A. The key outcome measures for Study 2 H1–H5 include emotion state reports and the beliefs elicited from the participant, which we define in our specifications as the logit of one's belief regarding a statement's truthfulness. As in Study 1, political ideology was captured by one's self-reported *Liberal Score* $\in [1, 9]$ and emotion states were also measured at the beginning of the study so that we could control for one's emotion state brought into the experiment. *TRUE* and *FALSE* statement indicator variables were used to separately estimate the effect of true versus false signals (i.e., our specifications will omit the constant term). *Statement Dissonance Level* $\in [1, 9]$ defines how dissonant a given statement is with one's *Liberal Score* ideology (i.e., if true, would the statement be dissonant with one's ideology). Finally, *Dissonant versus Consonant Signal* indicator variables identify if the noisy signal was aligned or misaligned with one's ideology.

Importantly, we consider that a signal may be considered *neither* categorically consonant nor dissonant if one's ideology is not firmly conservative or liberal (i.e., when $4 \leq \textit{Liberal Score} \leq 6$, $\textit{DissSignal} = \textit{ConsSignal} = 0$). Other key variables describing baseline beliefs and the likelihood of a statement's truth are defined below when discussion the empirical specification.

Empirical model specification:

Bayes rule defines the posterior probability of a statement being TRUE at time t, $\widehat{p}_t$, as a function of the prior probability at time $t-1$ of TRUE, $\widehat{p}_{t-1}$, and the information received in time t via the observed binary truthfulness signal, s_t, = S ∈ {0, 1}. Using odds ratios, and the likelihood ratio, LR_S, to capture new information, Bayes rule can be written as:

$$\frac{\widehat{p}_t}{1-\widehat{p}_t} = \frac{\widehat{p}_{t-1}}{1-\widehat{p}_{t-1}} \cdot LR_S \tag{2}$$

A log-odds (logit) specification is common in the literature (e.g., Coutts, 2019; Hill, 2017; Holt & Smith, 2009) and we maintain an interest in separately identifying the contribution of True ($S = 1$) and False ($S = 0$) signals, such that we specify the model as

$$\text{logit}(\widehat{p}_t) = \text{logit}(\widehat{p}_{t-1}) + I\{s_t = 1\} \cdot \ln(LR_1) + I\{s_t = 0\} \cdot \ln(LR_0) \tag{3}$$

Here, $I\{\cdot\}$ *are* indicator functions identifying true or false signals. This follows Coutts (2019) and produces the baseline empirical model:

$$\text{logit}(\widehat{p}_t) = \delta\text{logit}(\widehat{p}_{t-1}) + \beta_1 I\{s_t = 1\}\ln(LR_1) + \beta_0 I\{s_t = 0\}\ln(LR_0) + e_{it} \tag{4}$$

The signal strength we use in our design is 3/4, and so one should update beliefs regarding the statement being true by $ln\left(\frac{3/4}{1/4}\right) = \ln(3)$ when receiving a true signal and by $ln\left(\frac{1/4}{3/4}\right) = \ln\left(\frac{1}{3}\right)$ when receiving a false signal. The emotion hypotheses are tested using modifications of this baseline specification where we consider both the consonance or dissonance of the TRUE versus FALSE signal and one's initial statement-specific emotion state. Recall that, as defined, a signal can be either consonant, dissonant, or neither, such that Equation (5) below is well-specified.

$$\begin{aligned}\text{logit}(\widehat{p}_t) = {} & \delta\text{logit}(\widehat{p}_{t-1}) + \beta_1 I\{s_t = 1\}\ln(LR_1) + \beta_0 I\{s_t = 0\}\ln(LR_0) + \gamma \textit{DissSignal} \cdot \ln(LR_S) \\ & + \varphi \textit{ConsSignal} \cdot \ln(LR_S) + e_{it}\end{aligned} \tag{5}$$

6.3 | Study 2: Results

Our sample of $n = 503$ participants for Study 2 included $n = 294$ females (58%), $n = 266$ self-reported conservatives or Republicans (53%), and the average age of the participants was 37.65 years (±14.32 years standard deviation, range 18–79). Baseline *Relative Negative Mood* was −1.96 (±1.65, range −6 to +5), which implied an overall average net *positive* affective state at baseline.

6.3.1 | Tests of H1 and H2 (Bayesian updating and confirmation bias)

The test of hypothesis 1 required estimating model (4) above, which we did both with and without controls for age, gender, and baseline emotion state at the start of the survey. Errors were clustered at the participant level in these estimations, and estimation results are shown in Table 2. Consistent with previously reported findings (Coutts, 2019; Hill, 2017), our results indicate a type of cautious Bayesian updating. Specifically, both priors and new information significantly affect beliefs in the predicted direction ($\delta > 0$, $\beta_1 > 0$, $\beta_0 > 0$: $p < .01$ in each case). However, a perfect Bayesian would yield an estimated $\delta = \beta_1 = \beta_0 = 1$ in Equation (4). We find $0 < \delta < 1$, $0 < \beta_1 < 1$, $0 < \beta_0 < 1$, such that

TABLE 2 Belief updating and Bayes rule.

Dependent variable: Logit posterior belief	Model (1)	Model (2)
Regressor		
Age	—	0.001 (0.001)
Female (=1)	—	−0.016 (0.026)
Relative Negative Mood (baseline)	—	−0.000 (0.007)
Logit prior beliefs	0.852 (0.014)**	0.852 (0.014)**
Log-likelihood TRUE signal	0.588 (0.034)**	0.566 (0.050)**
Log-likelihood FALSE signal	0.588 (0.028)**	0.609 (0.041)**
R-squared	.774	.774
Observations (total)	8048	8048

Note: Models are estimated by linear regression with no constant term (given specification with TRUE and FALSE signal indicators on log-likelihood) with robust standard errors clustered at the subject level (n = 503 subjects total).

*p < .05, **p < .01 for the preregistered 1-tailed test on the prior beliefs and log-likelihood regressors (otherwise, 2-tailed tests).

the new information signal does not update one's beliefs to the extent predicted by Bayes rule ($p < .01$ for both the test $\beta_0 = 1$ and $\beta_1 = 1$ tests).

Hypothesis 2 examined whether a confirmation bias was present in one's baseline beliefs (H2a) and/or in one's degree of belief updating (H2b). For H2a, estimations were carried out separately for each statement such that we regressed baseline beliefs (before receiving any signals regarding the statement's accuracy) on the political ideology measure, *Liberal Score*. Results in Table 3 show that a higher *Liberal Score* predicts lower baseline estimates of the statement's truthfulness for Statements 1 and 4 (in each case, the statement being true would benefit conservatives), but a higher *Liberal Score* predicts higher baseline estimates of the truthfulness of Statements 2 and 3 (where truth would benefit liberals). Thus, the data support H2a.

Table 4 estimates Equation (5) above to test H2b. We also conducted sensitivity analysis by estimating specifications that included $\ln(LR_S)$ interactions with either dissonant or consonant signals (but not both, see models (3)–(6) in Table 4). Our estimation results in Table 4 show consistent evidence that a signal consonant with one's ideological beliefs (whether a "TRUE" or "FALSE" signal) is given significantly more weight one's belief updating. Dissonant signals are not weighted less, but together these results are consistent with H2b that dissonant signals will be given *relatively* less weight in one's belief updating than belief-consonant signals.

6.3.2 | Tests of Hypotheses 3 and 4 (emotion effects of dissonant statements and signals)

Hypothesis 3 is examined in regression results shown in Table 5. Here, we have four observations per participant for the relative negative emotion state from self-reports given immediately after viewing each political statement. We regressed *Rel Neg Mood* on the *Statement Dissonance Degree* to test H3. We estimated a specification with controls for age, gender, and baseline emotion (start of the study), and we also included a model with a *Female* × *Statement Dissonance Degree* interaction given some recent findings suggesting motivated reasoning in some context may differ by sex (Thaler, 2021). The Table 5 results show robust support for H3, which states that ideologically dissonant statements elicit more negative emotion states.

Hypothesis 4 is a related hypothesis that we tested by examining the change in one's self-reported emotion for a given statement both before the first noisy signal and after the series of four noisy signals have been received. We coded *# Dissonant Signals* to count the total number of dissonant signals received (between 0 and 4) for each statement. The dependent variable subtracts the relative negative emotion measure upon first viewing a statement from one's relative negative emotion measure after the series of four noisy signals. Thus, a positive coefficient estimate on *# Dissonant Signals* supports H4 by indicating that additional dissonant signals made one's emotion state worse. As can be seen in Table 6, the data fail to support H4, except for the case of Statement 4. We have no reason why Statement 4 should differ from the others regarding how dissonant signals affect one's affective state, and so we conclude that our data do not support H4 in general.

TABLE 3 Ideology effects on baseline perceptions of political statements.

Dependent variable = *Baseline belief* of statement's likely truth								
Regressor	*Statement 1*: TRUE (benefits conservatives)		*Statement 2*: FALSE (benefits conservatives)		*Statement 3*: TRUE (benefits liberals)		*Statement 4*: FALSE (benefits liberals)	
Constant term	0.753 (0.029)**	0.660 (0.051)**	0.455 (0.043)**	0.627 (0.057)**	0.529 (0.034)**	0.676 (0.061)**	0.585 (0.028)**	0.515 (0.050)**
Age	—	0.003 (0.001)**	—	−0.004 (0.001)**	—	−0.003 (0.001)*	—	0.000 (0.001)
Female (=1)	—	−0.071 (0.026)**	—	−0.062 (0.029)*	—	−0.042 (0.031)	—	0.069 (0.026)**
Relative Neg Mood (baseline)	—	−0.007 (0.008)	—	−0.003 (0.009)	—	0.008 (0.010)	—	−0.007 (0.008)
Liberal Score	−0.022 (0.005)**	−0.018 (0.005)**	0.018 (0.006)**	0.017 (0.006)**	0.038 (0.006)**	0.036 (0.006)**	−0.024 (0.005)**	−0.024 (0.005)**
R-squared	.032	.069	.020	.051	.072	.087	.043	.058
Observations	503	503	503	503	503	503	503	503

*$p < .05$, **$p < .01$ for the preregistered 1-tailed test on the coefficient estimate of *Liberal Score* (otherwise, 2-tailed tests).

TABLE 4 Updating response to belief-consonant versus dissonant signals.

Dependent variable: Logit posterior belief	Control for consonant and dissonant signals		Control for dissonant signals only		Control for consonant signals only	
Regressor	(1)	(2)	(3)	(4)	(5)	(6)
Age	—	0.001 (0.001)	—	0.001 (0.001)	—	0.001 (0.001)
Female (=1)	—	−0.015 (0.02)	—	−0.016 (0.026)	—	−0.015 (0.026)
Relative Neg Mood (baseline)	—	−0.000 (0.007)	—	−0.000 (0.007)	—	−0.000 (0.007)
Logit prior beliefs	0.849 (0.014)**	0.849 (0.014)**	0.851 (0.014)**	0.851 (0.014)**	0.849 (0.014)**	0.849 (0.014)**
Log-likelihood × TRUE signal	0.553 (0.037)**	0.531 (0.051)**	0.604 (0.036)**	0.584 (0.051)**	0.551 (0.032)**	0.529 (0.042)**
Log-likelihood × FALSE signal	0.549 (0.035)**	0.571 (0.047)**	0.605 (0.030)**	0.625 (0.042)**	0.547 (0.028)**	0.569 (0.042)**
Dissonant signal × log-likelihood	−0.004 (0.048)	−0.003 (0.048)	−0.057 (0.037)	−0.057 (0.037)	—	—
Consonant signal × log-likelihood	0.133 (0.047)**	0.133 (0.047)**	—	—	0.134 (0.037)**	0.134 (0.037)**
R-squared	.774	.774	.774	.774	.774	.774
Observations	8048	8048	8048	8048	8048	8048

Note: Models are estimated by linear regression with no constant term (given specification with TRUE and FALSE signal indicators on log-likelihood) with robust standard errors clustered at the subject level ($n = 503$ subjects total).

*$p < .05$, **$p < .01$ for the preregistered 1-tailed test on the coefficient estimates of the signal type interaction variables (otherwise, 2-tailed tests).

6.3.3 | Test of Hypothesis 5 (emotion affects the confirmation bias in belief updating)

Our final hypothesis for Study 2 tests of whether confirmation bias is a function of emotion state. We test H5 by estimating the model (5) specification for the subsets of those with relatively negative (*Rel Neg Mood* > 0) versus positive

TABLE 5 Dissonant statements and emotion.

Dependent variable = *Relative Negative Mood* after initially viewing statement			
Regressor	(1) Binary regression	(2) Controls regression	(3) Controls + gender interaction
Constant term	−1.422 (0.105)**	−0.632 (0.196)**	−0.644 (0.215)**
Age	—	−0.006 (0.004)	−0.006 (0.004)
Female (=1)	—	0.468 (0.112)**	0.488 (0.200)*
Relative Neg Mood (baseline)	—	0.434 (0.042)**	0.434 (0.042)**
Statement Dissonance Level ∈ [1, 9]	0.112 (0.018)**	0.112 (0.018)**	0.114 (0.027)**
Statement Diss Level × *Female*	—	—	−0.004 (0.036)
R-squared	.018	.166	.166
Observations	2012	2012	2012

Note: Models are linear regressions with errors are clustered at the subject.

*$p < .05$, **$p < .01$ for the preregistered 1 tailed test on the coefficient estimate of *Ideological Dissonance Level* (otherwise, 2-tailed tests).

TABLE 6 Ideologically dissonant noisy signals and negative emotion.

Dependent variable = change in *Relative Negative Mood* across statement trials								
Regressor	*Statement 1*: TRUE (benefits conservatives)		*Statement 2*: FALSE (benefits conservatives)		*Statement 3*: TRUE (benefits liberals)		*Statement 4*: FALSE (benefits liberals)	
Constant term	0.075 (0.059)	−0.118 (0.145)	−0.229 (0.070)**	0.102 (0.192)	−0.046 (0.057)	0.014 (0.139)	−0.171 (0.071)*	−0.407 (0.171)*
Age	—	0.000 (0.003)	—	−0.007 (0.004)	—	−0.001 (0.003)	—	0.004 (0.004)
Female (=1)	—	0.287 (0.092)**	—	−0.121 (0.122)	—	−0.010 (0.090)	—	0.116 (0.111)
# of Dissonant signals received	−0.051 (0.033)	−0.044 (0.033)	−0.035 (0.041)	−0.040 (0.041)	−0.013 (0.031)	−0.012 (0.031)	0.089 (0.03)**	0.087 (0.039)*
R-squared	.005	.024	.002	.008	.0004	.001	.011	.015
Observations	503	503	503	503	503	503	503	503

Note: Baseline emotion state is not used as a regressor here due to our focus on the relative emotion change across the trials of each statement viewed. Models are estimated by linear regression.

*$p < .05$, **$p < .01$ for the preregistered 1-tailed test on the coefficient of *# Dissonant signals received* (otherwise, 2-tailed tests).

(*Rel Neg Mood* < 0) emotion after viewing the statement. Table 7 results show that, for the subsample of those with relatively more positive emotion after viewing the statement, increased weight was given to belief-consonant signals (models (3) and (4) in Table 7).[24] We did not estimate the comparable result in the subsample of those with relatively more negative emotion post-statement (models (1) and (2)). While this is not precisely how H5 was articulated in our preregistration of the hypothesis, this result is somewhat consistent with H5. That is, dissonant signals were weighted *relatively* less than consonant signals in the subsample of those with a more positive post-statement emotion state, which implies that post-statement emotion response influenced the estimated confirmation bias.[25]

Because we defined each signal as *Consonant* = 1 or *Dissonant* = 1 for this analysis *only* for participants with a somewhat clear political ideology direction (i.e., signals were considered neither consonant nor dissonant if *Liberal Score* ∈ [4, 6] on the 1–9 scale), we also examined a less stringent definition of signal indicator variables. The alternative cutoffs considered a signal to be consonant or dissonant for any participant that did not report a *Liberal Score* = 5 (i.e., the middle point that indicated neither a liberal nor conservative identification). Somewhat expectedly, if one runs the same Table 7 analysis with this more accepting (but less sensitive) definition of signal consonance or dissonance, the interaction terms in columns (3) and (4) for Table 7 lose some statistical significance (though they remain significant at

TABLE 7 Emotion effects on the confirmation bias in belief updating.

Dependent variable = logit posterior belief				
	Subsample *Relative Neg Mood* > 0 (trials with relatively more negative reported emotion state)		Subsample *Relative Neg Mood* < 0 (trials with relatively more positive reported emotion state)	
Regressor	(1) Simple	(2) Controls	(3) Simple	(4) Controls
Age	—	0.003 (0.002)	—	−0.000 (0.001)
Female (=1)	—	−0.011 (0.062)	—	0.013 (0.033)
Relative Neg Mood (baseline)	—	0.025 (0.017)	—	−0.000 (0.012)
Logit prior beliefs	0.844 (0.023)**	0.843 (0.023)**	0.834 (0.020)**	0.834 (0.020)**
Log-likelihood × TRUE signal	0.564 (0.063)**	0.480 (0.104)**	0.601 (0.051)**	0.595 (0.065)**
Log-likelihood × FALSE signal	0.717 (0.068)**	0.789 (0.093)**	0.533 (0.043)**	0.539 (0.061)**
Dissonant signal × log-likelihood	−0.133 (0.082)	−0.129 (0.082)	0.015 (0.062)	0.015 (0.062)
Consonant signal × log-likelihood	−0.056 (0.078)	−0.048 (0.079)	0.185 (0.060)**	0.185 (0.060)**
R-squared	.736	.736	.786	.786
Observations	2140 (280 clusters)	2140 (280 clusters)	4824 (440 clusters)	4824 (440 clusters)

Note: Models are estimated by linear regression with no constant with robust standard errors clustered at the subject level ($n = 503$ subjects total, who may appear in one or both samples relative emotion state samples depending on the political statement viewed for a given set of belief assessments).
*$p < .05$, **$p < .01$ for the preregistered 1-tailed test on the coefficient estimates of the signal type interaction variables (otherwise, 2-tailed tests).

the $p < .05$ level and maintain similar magnitude of effect—these results are available on request). In other words, the results are most clear among those with a stronger ideological leaning, as one might expect.

Finally, because the *New* participants in Study 2 contained a higher proportion of female participants (65%) compared to *Original Wave* return participants (49% female), we conducted one final set of estimations to compare whether our results may be moderated by sex as research has suggested (Thaler, 2021). Using split sample estimations, Table 8 results indicate the emotion moderation effects may be restricted to female participants. Here, evidence suggests that for females a negative emotion response promoted discounting of dissonant signals and positive emotion response promoted additional weight on consonant signals. These significant results indicate more clear support of H5 for females than males (models (1) and (2)). Figure 7 helps visualize the key results in Tables 7 and 8, which shows how emotion moderated the impact of signal consonance or dissonance on updating and how this effect was clearer in the female subsample.

6.4 | Study 2: Discussion

Our results from the incentivized belief-elicitation task used Study 2 offer broad support for our hypotheses. The estimation results showed that both prior beliefs and new information matter in forming posterior probability estimates, as suggested by Bayes rule. Consistent with Coutts (2019) and Hill (2017), our results indicate that participants cautiously update beliefs toward the truth. This was the case in our data for both consonant and dissonant signals, which differs from results in Hill (2017) that showed cautious updating only occurred with belief-dissonant signals. In the domain of financially relevant belief-updating, Barron (2021) reported symmetric updating that was approximated by Bayes rule, not asymmetric updating based on signal type. However, in the framework presented earlier it is likely true that the direct utility component of one's beliefs is stronger for an ideological belief than for other types of beliefs, because ideology may even contribute to one's identity (e.g., consider political or religious beliefs). It is this direct utility component of beliefs that promotes an inflated view of one's belief validity that likely sows the seeds for confirmation bias. Though all participants in our Study 2 update beliefs in the correct direction, our findings still suggest that convergence toward the common truth occurs faster when presented with consonant rather than dissonant signals.

A negative emotion response to dissonant statements is robust and supports our hypothesis, although additional dissonant signals regarding the statement's truthfulness did not further amplify one's negative emotions. The data

TABLE 8 Emotion effects on confirmation bias—gender split samples.

Dependent variable = logit posterior belief				
	Males only		**Females only**	
Regressor	**(1) Subsample *Neg Mood* > 0 (neg emotion)**	**(2) Subsample *Neg Mood* < 0 (pos emotion)**	**(3) Subsample *Neg Mood* > 0 (neg emotion)**	**(4) Subsample *Neg Mood* < 0 (pos emotion)**
Logit prior beliefs	0.879 (0.027)**	0.802 (0.036)**	0.826 (0.031)**	0.863 (0.018)**
Log-likelihood × TRUE signal	0.533 (0.112)**	0.655 (0.089)**	0.569 (0.076)**	0.553 (0.056)**
Log-likelihood × FALSE signal	0.690 (0.124)**	0.551 (0.077)	0.731 (0.081)**	0.521 (0.053)**
Dissonant signal × log-likelihood	−0.064 (0.141)	0.015 (0.103)	−0.175 (0.102)*	0.018 (0.078)
Consonant signal × log-likelihood	−0.076 (0.133)	0.183 (0.100)*	−0.040 (0.099)	0.166 (0.065)**
R-squared	.777	.746	.716	.823
Observations	716 (94 clusters)	2216 (187 clusters)	1424 (186 clusters)	2608 (253 clusters)

Note: Models are estimated by linear regression with no constant with robust standard errors clustered at the subject level ($n = 503$ subjects total). *$p < .05$, **$p < .01$ for the preregistered 1-tailed test on the coefficient estimates of the signal type interaction variables (otherwise, 2-tailed tests).

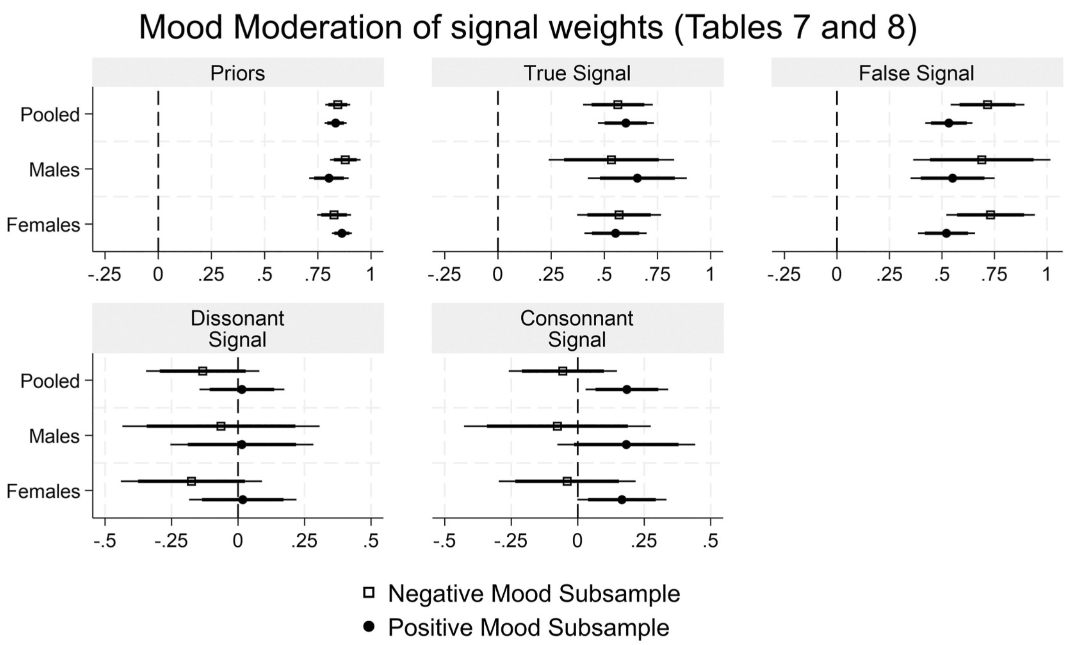

FIGURE 7 Graphical depiction of key Study 2 Results. The thick (thin) lines show the 95% (99%) confidence interval around the point estimates for the 1-tailed preregistered hypotheses.

supported our hypothesis that emotion moderates confirmation bias in this Bayesian setting. Positive emotion after viewing a political statement promoted giving extra weight to belief-consonant signals in one's updating, although those with more negative emotional state did not give significantly less weight to dissonant signals (note: this result was marginal at $p = .06$ significance). The emotion moderation results were stronger for females compared to males in our sample, which suggests emotion is of additional importance among females in understanding motivated reasoning.

These results were found using pooled data from all four political statements considered, although one may ask whether the results differed by statement. Separate estimations on each statement revealed that Statements 1 and 4 lead to findings most like those reported for the pooled data set. Results for Statement 2 estimations, however, show that "False" signals are weighted in one's probability updating, but not "True" signals. And, for Statement 3, "True" signals are the only signal given weight in the subsample of positive emotion response participant. The lack of marginal difference in weight placed on consonant or dissonant signals for some of our statements may suggest that those with *Liberal Score* in the intermediate range (for whom we do not classify any signal as consonant or dissonant) may nevertheless experience a bias in their updating as indicated by the main signal indicator effects. Again, future research can help further explore these nuanced results.

7 | GENERAL DISCUSSION AND CONCLUSION

This paper presented data from two studies examining behavior in 3 different tasks that explored different facets of preferences and belief formation. Each task we studied has its strengths and limitations, but we found evidence of confirmation bias tendencies all three of the different tasks. The findings presented also highlight several consistencies that shed light on our hypotheses regarding dissonance, emotion, and their effect on confirmation bias. Specifically, we reported robust evidence across tasks to support the hypothesis that ideologically dissonant information, arguments, or statements produce a negative affective/emotion response. In general, results across all tasks presented also supported the hypothesis that emotion response influences (i.e., moderates) the confirmation bias. In the classic confirmation bias task, more negative emotion increased the extent to which dissonant arguments were viewed as weak. In the novel information manipulation task, the result was most weak, with a more negative initial emotional state predicting a stronger discounting of dissonant information for just one of the two preference measures. The weaker support for the emotion moderation hypothesis in this task may relate to its focus on a variety of political issues that may differ in how partisan or emotionally charged is the issue. In Study 2, we found that one's emotional state after viewing a political statement is important for understanding how signals regarding the statement's accuracy will be incorporating into one's belief updating process. A relatively more positive emotion response predicted relatively more weight given to belief-consonant signals in updating beliefs, and this result was strongest in females. Perhaps emotion response primes one to differentially update forthcoming information signals, and this priming is stronger in females.[26] Additional research is needed to explore these male-female differences more systematically.

Emotion and affective states are important in understanding motivated reasoning in our data. While this study focused on short-term emotion states, and future research should examine the persistence of the effects we report, people form beliefs and make decisions in the midst of such temporal emotion states. This should underscore the importance of how the messaging of opposing views can alter their influence. In this case, messaging should be careful to not inflame negative emotion (e.g., if one wants to have a more constructive conversation in sharing views with someone from an opposing ideology). Regarding factual information, our design presented binary true or false signals on the factual content of the statements, and so it is unclear whether signals combined with justification would induce a greater or lesser affective response. Results from our information manipulation task in Study 1 suggest that arguments intended to convince one of an opposing viewpoint generated negative emotion responses. As such, while dissonant information signals in Study 2 did not further contribute to negative emotion beyond the initial emotion response to the statement, it may be the case that additional explanations or justifications along with those binary signals may serve to only make emotion states worse (possibly amplifying the effects found in our Study 2 task).

These findings may seem to suggest a no-win situation in dealing with confirmation bias, but our results can inform how one might lessen the effects of this bias. Strategies to mitigate confirmation bias may be particularly important if, as our results suggest, emotion responses moderate this bias. Yet, modern society facilitates and often values the delivery of political or other content that intends to provoke an emotional reaction. As a result, one confirmation bias mitigation strategy might be to limit one's use of social media or exposure to divisive content, in the same way it is likely beneficial to reduce exposure to any negative influence. This suggestion follows from an understanding that social media facilitates polarization (Van Bavel et al., 2021), and online media partly shapes opinion by asymmetrically using negative over positive words in headlines to increase click-through rates (Robertson et al., 2023). Alternatively, a cognitive therapy intervention may prove beneficial. The benefits of even brief doses of mindfulness meditation have been documented as an effective way to reduce induced negative emotions and rumination (Barcaccia et al., 2024; Conley et al., 2018). Still others have suggested behavioral (informational) nudges that can improve prosociality toward those belonging to partisan out-group, even if measures of polarization are less affected (Dimant, 2024). Thus, while our results highlight how negative emotion response may telegraph a stronger confirmation bias response to dissonant arguments, an improved understanding of this mechanism can help guide us toward effective mitigation strategies.

As always, this research points to areas where one may wish to focus future efforts to improve our understanding of the mechanisms that cause and moderate confirmation bias. At the very least, these findings help further clarify that confirmation bias, perhaps unlike other decision biases, involves input from both deliberative and emotional neural systems reinforce rather than counterbalance each other.

ACKNOWLEDGMENTS

The author is grateful for useful comment provided by seminar participants at GATE (Lyon, France), Appalachian State University, the University of Rennes 1, and participants at the Society for the Advancement of Behavioral Economics and the Economic Science Association conferences. An earlier version of this paper is posted as IZA Discussion Paper #15428. Institutional Review Board approval for this research was provided by Appalachian State University. The author wishes to thank the Walker College of Business and Appalachian State University for funding this work.

CONFLICT OF INTEREST STATEMENT

The author has no competing financial or non-financial interests that are directly or indirectly related to the work in this paper. No party had any rights to review this paper prior to its circulation.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in OPENICPSR at https://doi.org/10.3886/E207846V5 (Dickinson, 2024).

ORCID

David L. Dickinson https://orcid.org/0000-0003-3899-0276

ENDNOTES

[1] For example, views over whether climate changes is contributing to weather-related disasters (https://www.washingtonpost.com/climate-environment/2023/08/23/extreme-weather-climate-change-poll/) or beliefs regarding health misinformation (https://www.kff.org/coronavirus-covid-19/poll-finding/kff-health-misinformation-tracking-poll-pilot/) often diverge along party lines.

[2] See Knight (2006) for a review of the distinct conceptualizations of political ideology and how this has evolved over time. This present paper is most in-line with the view that ideology can be assessed along a liberal-conservative continuum, though it is true that binary group-identity itself may contribute to belief formation bias.

[3] See, for example, the distinction highlighted in Plato's *Protagoras* (https://www.gutenberg.org/files/1591/1591-h/1591-h.htm) between shipbuilding and city administration, where expert advice may be valued in the former than the latter case. I thank an anonymous reviewer for highlighting this example to me.

[4] See also Albertson et al. (2020) on the emotional component to political persuasion.

[5] For example, ideology likely drives one's perceptions of how economic systems function (Austin & Wilcox, 2007). And, another recent study documented reduced holiday time together among individuals visiting from opposite-voting precincts (Chen & Rohla, 2018).

[6] While emotion and mood are used somewhat interchangeably at times, most would consider that mood is a more long-lived state that may not have an identifiable trigger or start point, while emotion or emotion states are short-lived and can be linked to a particular object of trigger. That said, an accepted American Psychology Association definition of "mood" includes "any short-lived emotion state…." in addition to an alternative definition focused on mood being "a disposition to respond emotionally in a particular way that may last for hours, days, or even weeks…." (see https://dictionary.apa.org/mood). In this paper, we try to maintain consistency and refer to what participants report as emotion states, except for variables names where "mood" is used for simplicity and exposition. Though any longer-lived mood-disposition may impact self-reported emotional states, our study design that varies cognitive versus dissonant elements of political arguments or information in a within-subject fashion will still be able to test our key hypotheses.

[7] In their model, one subconsciously chooses the informativeness of an information signal in an ego-relevant task.

[8] We preregistered a larger set of hypotheses for Study 1, not all of which are discussed here (see Supporting Information S2: Appendix A for the full set of hypotheses). The focus of this paper is on those hypotheses that involve emotion effects of dissonant statements and their effect on confirmation bias in either the classic task (Hypotheses 1, 2a, and 2b) or the novel policy issue task (Hypotheses 3, 4a, 4b, 4c). We present full results testing additional preregistered hypotheses in Supporting Information S3: Appendix B, and results generally supported these additional hypotheses.

[9] The sources from which one could view statements were: The Democratic party, The Republican party, the Brady Campaign to Prevent Gun Violence, the National Rifle Association.

10 As an example, for the issue of *CRIME*, the *Amplify = 1* treatment statement was: "In a recent 2018 survey, the Bureau of Justice Statistics found that less than half of all violent crimes they track were reported to police, and an even lower percentage of property crimes were reported. Also, most crimes reported to the police were not 'solved' (did not result in arrest, charging, or a referral for prosecution). *Thus, it is clear that crime is a bigger problem in this country than we may have thought*." The information statement for the *Amplify = 0* treatment for the same policy issue was: "Recent data have shown consistent decreases in the violent crime rate in the U.S. And, recent data published by the Federal Bureau of Investigation from 2018 has shown an even larger drop in property crime. These drops in crime rates have been part of a 25-year trend now, and so *the magnitude of the crime problem in the U.S. is not as bad as one might think*." (highlights were included in what participants viewed for these issues and the information treatment statements). See Supporting Information S4: Appendix C for all statements used in the study.

11 After Task 2, participants were asked if they were registered to vote, and they were also asked if they intended to vote in the (upcoming, at the time) 2020 U.S. presidential election. The survey then asked for two final emotion states reports administered after viewing an image of Donald Trump and Joe Biden (presented in random order). A 6-item cognitive reflection task (Primi et al., 2016) then completed the survey for Study 1.

12 Though the alternative of subtracting negative emotion state measures from the positive ones would produce a measure that is negative in value for worse emotional states, we chose to construct *Relative Negative Mood* as a measure that is greater than zero for a *worse* (i.e., more negative) emotion states. We will be clear to describe the results considering our approach to constructing *Relative Negative Mood*, which was done to more closely align with our preregistered hypotheses that predicted effects on the negativity of one's state-level emotion.

13 We also preregistered a hypothesis to test whether emotion moderation depends on one's explicit approach to emotion regulation via cognitive reappraisal, which is considered a healthy emotion regulation strategy (e.g., see Heilman et al., 2010, for an example of cognitive reappraisal effects on decision making). See Supporting Information S3: Figure B4 and Table B7 for these results.

14 The construction of *Dissonance Degree* in this way creates a perfect collinearity misspecification in the model, which we avoid by using log (*Rating*) for preference measures. Alternatively, one can estimate the model without the $Rating_{pre}$ explanatory variable to avoid the issue of collinearity of *Dissonance Degree* with one's initial preference rating. Doing so yields similar results to those reported in that the combined coefficients on *Dissonance Degree* and the *Dissonance Degree* × *Amplify* interaction term are negative and statistically significant ($p < .001$). In other words, information that amplifies the importance of the policy issue, when it is more dissonant with one's preference rating, updates one's preference rating to a lesser degree.

15 Key results not shown in the main text include the predicted positive and significant effect of one's baseline preferences rating on post-information ratings ($\beta_1 > 0$ in all instances, $p < .01$), and a significant and positive main effect of *Amplify*, as expected if one believes preference ratings respond in the direction of the information statement ($\beta_2 > 0$, $p < .01$ in all instances). These effects can be seen in Supporting Information S2: Table A6.

16 Re-estimation using a *Minimize* = 1 information treatment indicator variable (equivalent to *Amplify* = 0), which merely redefines *Amplify* = 1 as the omitted reference category, yields results that would be interpreted similarly—this merits mention because the effect of *Amplify* = 0 treatment is difficult to conclude from the specification reported here, though easily seen using the opposite treatment indicator.

17 For example, assume one thinks a policy issue is relatively unimportant, and so the perceived problem $Rating_{pre} = 30$. In this case, *Amplify* = 1 implies that one's *Dissonance Degree* = 70. Compare this to a statement for which $Rating_{pre} = 60$ such that *Dissonance Degree* = 40 when *Amplify* = 1. Of course, the model estimates imply that one's pre-statement rating is the most significant determinant of one's post-statement rating. However, the combined coefficients ($\beta_2 + \beta_3 + \beta_4$) for *Dissonance Degree* = 70 predict an approximate 29% increase in one's perceived problem $Rating_{post}$ when *Amplify* = 1, but an approximate 39% increase for *Dissonance Degree* = 30 (i.e., after exponentiating the combined coefficients). One could also interpret this as the predicted between-participant difference in the information statement's impact between a typical politically liberal versus conservative participant. Recall that *Dissonance Degree* is defined not by political ideology measure, but rather direction by how one's issue-specific preference rating differed from the information statement treatment received. For some of the more partisan issues used in the task the average preference ratings differed by more than 30 points, while they differed by less on other less-partisan issues.

18 For example, recall that in Task 2 *Degree of Dissonance* is defined based on the participant's preference ratings, and so if the rating is not trustworthy or sincere, then we would not expect to find the hypothesized impact of an insincere preference ratings (which sets the level of the constructed dissonance degree variable) on emotion. Of course, emotion state is given as a self-report also, but a bias in our estimated effects highlighted in Figure 3 would seem to require that participants deliberately report their emotion state in conjunction with an insincere preference rating in the specific way required to support our hypothesized effect.

19 For Study 2, all participants "passed" attention checks because these were automated within Study 2 such that those failing the attention check would not be allowed to continue. This allowed the study to be returned to Prolific to open up the slot for another participant. This approach was not used for Study 1, which is why some in Study 1 failed attention checks and were not included in the analysis (as per our preregistration plans for that study).

20 This distinction is due to the use of a political "ideology" identification question from the Prolific screeners for Study 1, but we used a related but distinct "political party identification" question in Study 2.

21 Age and sex on Study 2 participants were obtained from available data within Prolific. We have control measures for those Study 2 participants who also completed Study 1, but this ended up being less than half the final Study 2 sample size. While more characteristics of

the Study 2 participants were available from Prolific, none of these corresponded to the other measures used in our set of Study 1 control variables.

[22] As noted by an anonymous reviewer, Statement 2 regarding COVID-19 is difficult to evaluate without CDC data, and is probably a true rather than false statement as indicated in Table 1. The source article used the first 9 months of 2021 when speaking of COVID-19 deaths under President Biden, but I changed this for the study instrument to be more comparable and focus only on the first 9 months of the Biden presidency (which then means swapping some high COVID-19 deaths weeks in January 2021 before Biden's inauguration with lower COVID-19 deaths weeks in October 2021). That said, the point of the Statement was that it *not* be completely transparent whether the statement was true or false, in order to allow for a meaningful analysis of Bayesian updating to the noisy signal. Recall also that the statement's *Dissonance Level* was defined with respect to whether the statement as phrased (i.e., if true) was dissonant with the individuals *Liberal Score* ideology measure. As such, this concern would affect the intended balance across design cells in Table 1, but the key *Dissonance Level* measure remains valid even if a statement intended to be false were actually true (or vice versa).

[23] This procedure is a variation of the BDM mechanism first used as an incentive compatible way to elicit willingness-to-pay (Becker et al., 1964) and is invariant to risk attitude (see Holt & Smith, 2016, for a comparison of several belief elicitation procedures).

[24] A singular (pooled sample) model only shows *negative* significance of the *Rel Neg Mood* $\times$ *Consonant Signal* $\times$ *lnlr* triple interaction ($p < .05$), but it is hard to interpret whether that results from an increase in negative emotion state decreasing one's updating emphasis of consonant signals, or whether it results from an increase in positive emotion state (negative change in *Neg Mood*) increasing one's weight on consonant signals. The emotion-split sample estimations above clear this up. Note that the Dissonant signal interaction becomes negative and marginally significant if considering those with *Rel Neg Mood* > 1, which means marginal significance of the result that relatively more negative emotion reduces one's emphasis on dissonant signals (even with samples above, the results is $p < .06$).

[25] Because the sample of participants for Study 2 included both return, or *Original Wave*, participants from Study 1 as well as *New* participants, we also conducted sensitivity analysis on this key Hypothesis 5 (see Supporting Information S2: Tables A14–A16). In Supporting Information S2: Tables A14 and A15, we estimated the model (5) specification on the subsample of 193 participants in Study 2 who also participated in Study 1 (i.e., original wave [return] participants): Supporting Information S2: Table A14 shows the results for the relatively negative emotion observations, and Supporting Information S2: Table A15 shows results for the relatively positive emotion observations. We estimated models with and without controls, and also with and without an inverse-probability weight correction for potential sample selection (Supporting Information S2: Table A13 gives the return-participant selection equation). These estimations fail to find robust support for consonant signals being weighted more strongly for those with a positive emotion state (Supporting Information S2: Table A15)—the results are qualitatively similar among *Original Wave* participants compared to those found when pooling all Study 2 participants, but statistical significance is lost when controlling for sample selection (see Supporting Information S2: Table A15). However, we find robust support in Supporting Information S2: Table A14 that those with a relatively negative post-statement emotion state weight dissonant signals less. A final set of estimations (Supporting Information S2: Table A16) confirmed that the initial result reported in Table 8 is driven mostly by the *New* participants in Study 2. The consistent finding is that a *relatively* more positive (negative) emotion response to the political statement leads to an estimated higher *relative* weight placed on consonant (dissonant) signals in one's belief updating.

[26] However, separate exploratory estimations did not find a significant interaction effect between sex and the extent to which a statement's dissonance promoted negative emotion.

REFERENCES

Albertson, B., Dun, L. & Gadarian, S.K. (2020) The emotional aspects of political persuasion. In: Suhay, E., Grofman, B. & Trechsel, A.H. (Eds.) *The Oxford handbook of electoral persuasion*. New York: Oxford University Press, pp. 169–183.

Allcott, H. & Gentzkow, M. (2017) Social media and fake news in the 2016 election. *The Journal of Economic Perspectives*, 31(2), 211–236. Available from: https://doi.org/10.1257/jep.31.2.211

Austin, D.A. & Wilcox, N.T. (2007) Believing in economic theories: Sex, lies, evidence, trust, and ideology. *Economic Inquiry*, 45(3), 502–518. Available from: https://doi.org/10.1111/j.1465-7295.2006.00042.x

Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Hunzaker, M.F., et al. (2018) Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216–9221. Available from: https://doi.org/10.1073/pnas.1804840115

Bakshy, E., Messing, S. & Adamic, L.A. (2015) Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132. Available from: https://doi.org/10.1126/science.aaa1160

Barcaccia, B., Medvedev, O.N., Pallini, S., Mastandrea, S. & Fagioli, S. (2024) Examining Mental Health Benefits of a Brief Online Mindfulness Intervention: A Randomised Controlled Trial. *Mindfulness*, 15(4), 1–9. Available from: https://doi.org/10.1007/s12671-024-02331-8

Barron, K. (2021) Belief updating: does the 'good-news, bad-news' asymmetry extend to purely financial domains? *Experimental Economics*, 24(1), 31–58. Available from: https://doi.org/10.1007/s10683-020-09653-z

Bauer, K., Chen, Y., Hett, F. & Kosfeld, M. (2022) Partisan group identity predicts polarization. *Working Paper*.

Becker, G.M., DeGroot, M.H. & Marschak, J. (1964) Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3), 226–232. Available from: https://doi.org/10.1002/bs.3830090304

Charpentier, C.J., Bromberg-Martin, E.S. & Sharot, T. (2018) Valuation of knowledge and ignorance in mesolimbic reward circuitry. *Proceedings of the National Academy of Sciences*, 115(31), E7255–E7264. Available from: https://doi.org/10.1073/pnas.1800547115

Chen, M.K. & Rohla, R. (2018) The effect of partisanship and political advertising on close family ties. *Science*, 360(6392), 1020–1024. Available from: https://doi.org/10.1126/science.aaq1433

Conley, S.L., Faleer, H.E., Raza, G.T., Bailey, B.E. & Wu, K.D. (2018) The moderating effects of rumination facets on the relationship between mindfulness and distress reduction. *Cognitive Therapy and Research*, 42(4), 436–446. Available from: https://doi.org/10.1007/s10608-018-9896-7

Cotton, J.L. (1985) Cognitive dissonance in selective exposure. In: Zillmann, D. & Bryant, J. (Eds.) *Selective Exposure to Communication*. New York and London: Routledge, Taylor and Francis Group, pp. 11–33.

Coutts, A. (2019) Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, 22(2), 369–395. Available from: https://doi.org/10.1007/s10683-018-9572-5

Dickinson, D. (2024) ECIN Replication Package for "Political ideology, emotion response, and confirmation bias." Ann Arbor: Inter-university Consortium for Political and Social Research [distributor]. Available from: https://doi.org/10.3886/E207846V5

Dickinson, D.L. (2020) Deliberation enhances the confirmation bias in politics. *Games*, 11(4), 57. Available from: https://doi.org/10.3390/g11040057

Dimant, E. (2024) Hate trumps love: The impact of political polarization on social preferences. *Management Science*, 70(1), 1–31. Available from: https://doi.org/10.1287/mnsc.2023.4701

Drobner, C. & Goerg, S.J. (2024) Motivated belief updating and rationalization of information. *Management Science*, 70(7), 4583–4592. in press. Available from: https://doi.org/10.1287/mnsc.2023.02537

Druckman, J.N. (2022) A framework for the study of persuasion. *Annual Review of Political Science*, 25(1), 65–88. Available from: https://doi.org/10.1146/annurev-polisci-051120-110428

Engelmann, J.B., Lebreton, M., Salem-Garcia, N.A., Schwardmann, P. & van der Weele, J.J. (2024) Anticipatory anxiety and wishful thinking. *The American Economic Review*, 114(4), 926–960. Available from: https://doi.org/10.1257/aer.20191068

Festinger, L. (1957) *A Theory of Cognitive Dissonance*, vol. 2. Stanford University Press. Available from: https://doi.org/10.1515/9781503620766

Flynn, D.J., Nyhan, B. & Reifler, J. (2017) The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, 38(S1), 127–150. Available from: https://doi.org/10.1111/pops.12394

Frey, D. (1986) Recent research on selective exposure to information. In Berkowitz, L. (Ed.) *Advances in Experimental Social Psychology*. New York: Academic Press, pp. 41–80.

Frimer, J.A., Skitka, L.J. & Motyl, M. (2017) Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions. *Journal of Experimental Social Psychology*, 72, 1–12. Available from: https://doi.org/10.1016/j.jesp.2017.04.003

Gentzkow, M., Shapiro, J.M. & Stone, D.F. (2015) Media bias in the marketplace: Theory. In: *Handbook of Media Economics*, vol. 1. North-Holland, pp. 623–645.

Gross, J.J. & John, O.P. (2003) Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, 85(2), 348–362. Available from: https://doi.org/10.1037/0022-3514.85.2.348

Hart, W., Albarracín, D., Eagly, A.H., Brechan, I., Lindberg, M.J. & Merrill, L. (2009) Feeling validated versus being correct: a meta-analysis of selective exposure to information. *Psychological Bulletin*, 135(4), 555–588. Available from: https://doi.org/10.1037/a0015701

Heilman, R.M., Crişan, L.G., Houser, D., Miclea, M. & Miu, A.C. (2010) Emotion regulation and decision makingunder risk and uncertainty. *Emotion*, 10(2), 257.

Hill, S.J. (2017) Learning together slowly: Bayesian learning about political facts. *The Journal of Politics*, 79(4), 1403–1418. Available from: https://doi.org/10.1086/692739

Holt, C.A. & Smith, A.M. (2009) An update on Bayesian updating. *Journal of Economic Behavior & Organization*, 69(2), 125–134. Available from: https://doi.org/10.1016/j.jebo.2007.08.013

Holt, C.A. & Smith, A.M. (2016) Belief elicitation with a synchronized lottery choice menu that is invariant to risk attitudes. *American Economic Journal: Microeconomics*, 8(1), 110–139. Available from: https://doi.org/10.1257/mic.20130274

Jerit, J. & Zhao, Y. (2020) Political misinformation. *Annual Review of Political Science*, 23(1), 77–94. Available from: https://doi.org/10.1146/annurev-polisci-050718-032814

Jones, M. & Sugden, R. (2001) Positive confirmation bias in the acquisition of information. *Theory and Decision*, 50(1), 59–99. Available from: https://doi.org/10.1023/a:1005296023424

Kahan, D.M. (2012) Ideology, motivated reasoning, and cognitive reflection: An experimental study. *Judgment and Decision Making*, 8(4), 407–424. Available from: https://doi.org/10.1017/s1930297500005271

Knight, K. (2006) Transformations of the Concept of Ideology in the Twentieth Century. *American Political Science Review*, 100(4), 619–626. Available from: https://doi.org/10.1017/s0003055406062502

Knobloch-Westerwick, S., Mothes, C. & Polavin, N. (2020) Confirmation bias, in-group bias, and negativity bias in selective exposure to political information. *Communication Research*, 47(1), 104–124. Available from: https://doi.org/10.1177/0093650217719596

Kobayashi, K. & Hsu, M. (2019) Common neural code for reward and information value. *Proceedings of the National Academy of Sciences*, 116(26), 13061–13066. Available from: https://doi.org/10.1073/pnas.1820145116

Lodge, M. & Taber, C.S. (2013) *The rationalizing voter*. New York: Cambridge University Press.

Lord, C.G., Ross, L. & Lepper, M.R. (1979) Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098–2109. Available from: https://doi.org/10.1037//0022-3514.37.11.2098

Nicolle, A., Fleming, S.M., Bach, D.R., Driver, J. & Dolan, R.J. (2011) A regret-induced status quo bias. *Journal of Neuroscience*, 31(9), 3320–3327. Available from: https://doi.org/10.1523/jneurosci.5615-10.2011

Palan, S. & Schitter, C. (2018) Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27. Available from: https://doi.org/10.1016/j.jbef.2017.12.004

Peer, E., Brandimarte, L., Samat, S. & Acquisti, A. (2017) Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 153–163. Available from: https://doi.org/10.1016/j.jesp.2017.01.006

Primi, C., Morsanyi, K., Chiesi, F., Donati, M.A. & Hamilton, J. (2016) The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making*, 29(5), 453–469. Available from: https://doi.org/10.1002/bdm.1883

Robertson, C.E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J.J. & Feuerriegel, S. (2023) Negativity drives online news consumption. *Nature Human Behaviour*, 7(5), 812–822. Available from: https://doi.org/10.1038/s41562-023-01538-4

Taber, C.S. & Lodge, M. (2006) Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755–769. Available from: https://doi.org/10.1111/j.1540-5907.2006.00214.x

Thaler, M. (2021) Gender differences in motivated reasoning. *Journal of Economic Behavior & Organization*, 191, 501–518. Available from: https://doi.org/10.1016/j.jebo.2021.09.016

Van Bavel, J.J., Rathje, S., Harris, E., Robertson, C. & Sternisko, A. (2021) How social media shapes polarization. *Trends in Cognitive Sciences*, 25(11), 913–916. Available from: https://doi.org/10.1016/j.tics.2021.07.013

Van Veen, V., Krug, M.K., Schooler, J.W. & Carter, C.S. (2009) Neural activity predicts attitude change in cognitive dissonance. *Nature Neuroscience*, 12(11), 1469–1474. Available from: https://doi.org/10.1038/nn.2413

Westen, D., Blagov, P.S., Harenski, K., Kilts, C. & Hamann, S. (2006) Neural bases of motivated reasoning: An fMRI study of emotional constraints on partisan political judgment in the 2004 US presidential election. *Journal of Cognitive Neuroscience*, 18(11), 1947–1958. Available from: https://doi.org/10.1162/jocn.2006.18.11.1947

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Dickinson, D.L. (2024) Political ideology, emotion response, and confirmation bias. *Economic Inquiry*, 1–25. Available from: https://doi.org/10.1111/ecin.13253