Bayesian or biased? Analytic thinking and political belief updating

Ben M. Tappin [1], Gordon Pennycook [2], David G. Rand [1,3]

[1] Sloan School of Management, Massachusetts Institute of Technology

[2] Hill/Levene School of Business, University of Regina

[3] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

This manuscript is forthcoming in *Cognition*.

Author note

We acknowledge funding from the Economic and Social Research Council and the Templeton World Charity Foundation. This article has supplementary information available online at https://osf.io/zwjkd/.

Correspondence concerning this article should be addressed to Ben Tappin, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02142, United States of America. Email: benmtappin@googlemail.com

Abstract

A surprising finding from U.S. opinion surveys is that political disagreements tend to be greatest among the most cognitively sophisticated opposing partisans. Recent experiments suggest a hypothesis that could explain this pattern: cognitive sophistication magnifies politically biased processing of new information. However, the designs of these experiments tend to contain several limitations that complicate their support for this hypothesis. In particular, they tend to (i) focus on people's worldviews and political identities, at the expense of their other, more specific prior beliefs, (ii) lack direct comparison with a politically unbiased benchmark, and (iii) focus on people's judgments of new information, rather than on their posterior beliefs following exposure to the information. We report two studies designed to address these limitations. In our design, U.S. subjects received noisy but informative signals about the truth or falsity of partisan political questions, and we measured their prior and posterior beliefs, and cognitive sophistication, operationalized as analytic thinking inferred via performance on the Cognitive Reflection Test. We compared subjects' posterior beliefs to an unbiased Bayesian benchmark. We found little evidence that analytic thinking magnified politically biased deviations from the benchmark. In contrast, we found consistent evidence that greater analytic thinking was associated with posterior beliefs closer to the benchmark. Together, these results are inconsistent with the hypothesis that cognitive sophistication magnifies politically biased processing. We discuss differences between our design and prior work that can inform future tests of this hypothesis.

Keywords: Political beliefs, Bayesian updating, cognitive sophistication, polarization

Bayesian or biased? Analytic thinking and political belief updating

Partisan disagreement is a salient feature of contemporary American politics (Pew Research Center, 2019). This disagreement is observed not only in people's political preferences and values, but also in their "factual" beliefs—that is, in beliefs one might reasonably expect to converge on publicly available and relevant empirical evidence. Canonical examples include people's beliefs about the performance of the US economy (Bartels, 2002; Dunn & Oliphont, 2018), the existence of weapons of mass destruction at the time of the 2003 American invasion of Iraq (Bullock, 2009; Nyhan & Reifler, 2010), and the danger to society posed by global warming, private gun ownership, and fracking (Kahan, 2015).

Perhaps surprisingly, a large body of evidence indicates that numerous such disagreements tend to widen in conjunction with partisans' level of cognitive sophistication. This has the effect that the most cognitively sophisticated opposing partisans tend to disagree most strongly over a variety of "factual" political questions like those described above. For example, indicators of cognitive sophistication such as educational attainment, science literacy, numeracy, specific topic knowledge, and a propensity for analytic thinking have all been associated with greater partisan disagreement in the U.S., over questions as diverse as the causes and dangers of climate change (Drummond & Fischhoff, 2017; Kahan et al., 2012; Kahan, Landrum, et al., 2017; Kahan & Corbin, 2016; Malka et al., 2009; McCright & Dunlap, 2011; van der Linden et al., 2018), the safety of vaccination (Hamilton et al., 2015; Joslyn & Sylvester, 2019), the risks associated with fracking (Kahan, 2015; Kahan, Landrum, et al., 2017), whether the human species is a product of evolution by natural selection (Drummond & Fischhoff, 2017; Hamilton & Saito, 2015), and whether Iraq possessed weapons of mass destruction in 2003 (Joslyn & Haider-Markel, 2014).

Cognitive Sophistication Magnifies Politically Biased Processing

An influential hypothesis for these patterns is that cognitive sophistication magnifies politically biased cognitive processing. This hypothesis begins with the assumption that all partisans—irrespective of their cognitive sophistication—use their cognitive capacities to process new information in a politically biased manner. The key prediction of the hypothesis is that, because cognitively sophisticated partisans enjoy the strongest of these capacities, their processing tends to be the *most* politically biased (Kahan, 2013; Kahan et al., 2012; Taber et al., 2009; Taber & Lodge, 2006); that is, they are "particularly susceptible" (p. 139) to the bias (Taber et al., 2009). Thus, the most cognitively sophisticated opposing partisans wind up being the most polarized in their beliefs about political issues, on average.

Most tests of this hypothesis have used a particular type of study design, in which subjects are asked to interpret and evaluate new information provided by the experimenter (Kahan, 2013; Kahan, Peters, et al., 2017; Kuru et al., 2017; Lind et al., 2018; Nurse & Grant, 2019; Sumner et al., 2018; Taber et al., 2009; Taber & Lodge, 2006). A key feature of this design is that the new information is manipulated to either *favor* or *impugn* subjects' political identities or general worldview, but is otherwise held the same across treatment groups.

For example, in one study (Kahan, 2013) U.S. partisans evaluated the validity of the Cognitive Reflection Test (CRT), a behavioral measure of the propensity to think analytically (Frederick, 2005). The CRT was described to partisans as a test of "open-minded and reflective" thinking. Before giving their evaluations, partisans completed the test themselves and were randomly assigned to one of two treatments (or control) in which they were provided information about the test. In treatment A, partisans were told that people who believe that climate change is happening tend to score higher on the test than people who are skeptical that climate change is happening; implying the former are more open-minded. In treatment B, partisans were told the reverse: that people who are skeptical that climate change is happening tend to score higher in the test, implying *they* are more open-minded.

The results showed that partisans who identified on the political left tended to evaluate the test as more valid in treatment A than B; and vice versa for partisans who identified on the political right. However, the key result was that these conditional evaluations tended to be strongest among those partisans who scored highest on the CRT. In other words, the most cognitively sophisticated opposing partisans tended to disagree most strongly in their evaluations of the information, consistent with the hypothesis that cognitive sophistication magnifies politically biased processing. Other studies using the same type of design have recorded a similar result: cognitive sophistication appears to magnify the correlation between subjects' political identities or worldviews and their evaluations of the validity of new information, where the information is manipulated to either favor or impugn their political identities or worldviews (Kahan, Peters, et al., 2017; Kuru et al., 2017; Nurse & Grant, 2019; Sumner et al., 2018; Taber et al., 2009; Taber & Lodge, 2006) (but see Lind et al., 2018).

While such results are frequently regarded as evidence that cognitive sophistication magnifies politically biased processing, a shortcoming of these study designs is that people's political identities and worldviews are often correlated with their pretreatment information exposure and issue-specific prior beliefs (Baron & Jost, 2019; Friedman, 2012; Tappin et al., 2020; Tappin & Gadsby, 2019). Consequently, in such designs, "switching" the information so that it impugns (versus favors) people's political identities or worldviews simultaneously "switches" the information so that it is incoherent (versus coherent) with their issue-specific prior beliefs and experiences. This introduces an ambiguity into the design, because evidence suggests that the coherence between new information and specific prior beliefs can affect people's reasoning about the new information *per se*—that is, in contexts where political and worldview considerations are absent (Evans et al., 1983; Klauer et al., 2000; Koehler, 1993; Markovits & Nantel, 1989; Mercier, 2012, 2017). Consequently, it remains somewhat unclear to what extent cognitive sophistication magnifies (a) politically biased processing of new

information versus (b) the more general influence on cognitive processing of the incoherence between specific prior beliefs and new evidence (Tappin et al., 2018).

Given the focus and designs of previous work, and the ambiguities that arise in interpretation of the evidence, in this paper we use an alternative design to study whether cognitive sophistication magnifies politically biased cognitive processing. In particular, we focus on the impact of the new information on people's posterior beliefs about the issue addressed by the information, rather than on how people evaluate the new information itself. In the following section, we outline our design and describe some of the benefits it possesses compared with designs that focus on people's evaluations of new information per se.

Our Design

We adapt a design in which people receive noisy but informative information about the truth or falsity of factual political questions (Hill, 2017). In this design, we measure people's prior beliefs about the questions, and we define (Study 1) or measure (Study 2) their perception about the informativeness of the information. Based on these data, we calculate the posterior beliefs that are expected according to Bayes' rule. We then compare individuals observed posterior beliefs to this Bayesian benchmark; evaluating the direction and extent to which their posterior beliefs diverge from the benchmark as a function of the political favorability of the new information (i.e., whether it is favorable or unfavorable for their stated political affiliation). For example, unambiguous evidence of politically biased deviations would be posterior beliefs that are "too far" in the direction of politically favorable information, and/or "not far enough" in the direction of politically *un*favorable information. Using this design, we ask to what extent cognitive sophistication magnifies politically biased deviations from the Bayesian benchmark.

This study design offers several benefits over designs that focus on people's evaluations of new information per se to test the hypothesis that cognitive sophistication magnifies politically biased processing. First, we define (Study 1) and measure (Study 2) how informative

people consider the new information to be. Without this piece of the belief updating equation, it is often difficult to make distinct predictions about how people's updated beliefs ought to look according to different theoretical processes (Hill, 2017). In particular, it is difficult to distinguish politically biased beliefs from beliefs that are politically *un*biased, or otherwise normative (Druckman & McGrath, 2019; Kahan, 2016; Sunstein et al., 2016). Distinguishing these beliefs is desirable, however, insofar as politically biased cognitive processing implies deviation from accuracy or optimality (Hahn & Harris, 2014).

Second, in our design people's specific prior beliefs are explicitly accounted for when computing their expected posterior beliefs according to Bayes' rule. That is, the Bayesian-expected posterior belief is a function of the specific prior belief of the individual (plus the diagnosticity of the new information provided). Because we evaluate politically biased processing with respect to these expectations, the prior belief confound often present in study designs that focus on information evaluations as the outcome variable (described previously) is minimized.

Third, it is somewhat unclear how researchers should interpret people's self-reported evaluations of new information when inferring their capacity for politically biased cognitive processing. In particular, the two outcome variables of (i) information evaluations versus (ii) posterior beliefs following exposure to the information can yield divergent results (Anglin, 2019), and sometimes dramatically so (Kunda, 1987). For example, Kunda (1987) studied how heavy-coffee-drinking women evaluated a piece of research that linked heavy coffee drinking among women to a specific type of disease. The heavy-coffee-drinking women evaluated the research more negatively than women who drank little coffee, and men. This is conceptually akin to the biased-evaluations-of-information result described previously, but in an apolitical domain, and implies that the heavy-coffee-drinking women were resistant to the new and undesirable information. However, Kunda also measured subjects' posterior beliefs about their likelihood of contracting the disease. She found evidence of substantial belief updating towards the information on the part of the heavy-coffee-drinking women, suggesting they incorporated the

information into their prior beliefs, despite their negative evaluations. This result was not discernible from the information evaluations alone, however, and illustrates that the two outcomes can privilege different interpretations.

This is important because the hypothesis that we consider here—that cognitive sophistication magnifies politically biased processing—fundamentally seeks to explain why the most cognitively sophisticated opposing partisans show the widest disagreement in their political beliefs. In other words, the hypothesis is first and foremost concerned with how information affects people's beliefs, not with their self-reported evaluations of new information itself.

Fourth, because our design explicitly identifies a normative and unbiased (Bayesian) benchmark for the posterior beliefs, in addition to the magnified-political-bias hypothesis, we can also test the distinct hypothesis that cognitive sophistication is correlated with more normative posterior beliefs overall. This hypothesis is loosely implied by previous work that finds, for example, that performance on indicators of cognitive sophistication is associated with more accurate beliefs about political news headlines (Pennycook & Rand, 2019) and with epistemic rationality more generally (Pennycook et al., 2015). Notably, this hypothesis is not mutually exclusive with the magnified-political-bias hypothesis. That is, we could in principle find evidence consistent with both (or neither) hypotheses. For example, cognitively sophisticated individuals could be closer to the Bayesian benchmark overall, but nevertheless demonstrate greater political bias than their less sophisticated counterparts—in the form of a larger asymmetry in posterior beliefs for politically favorable versus unfavorable information.

In order to gain the aforementioned benefits, it is necessary for our design to depart in a number of ways from previous work that has studied the magnified-political-bias hypothesis. As already mentioned, one way in which our design departs from previous work is that we focus on how new information influences people's subsequent beliefs, rather than on how people evaluate the information itself—an intentional and beneficial departure, for reasons we have explained. However, another important difference is the type of information to which people are exposed.

Whereas previous work (Taber et al., 2009; Taber & Lodge, 2006) provided people with verbal arguments for and against public policy issues, for example, such information is not so easily translatable into the framework of our design. Specifically, in order to define the Bayesian-expected posterior beliefs, it is necessary for us to simplify the type of information to which people are exposed. These differences are important to bear in mind in the interpretation of our results, and we return to them in the Discussion of the paper.

Study 1

Methods

Study 1 was preregistered at https://osf.io/e39kq. The project hub for this manuscript, which contains supplementary information, data, and analysis code for both Studies 1 and 2 is available on the Open Science Framework at https://osf.io/zwjkd/.

Sample

We sought to recruit N=500 subjects, providing 90% power to detect a small association ($r \approx .15$) between CRT performance and deviation from Bayesian posterior beliefs (a power analysis is reported in the preregistered protocol). Subjects were adults from the U.S. recruited via Amazon's Mechanical Turk (MTurk). A total of N=501 subjects completed the study. Samples recruited via MTurk are not representative of the general U.S. population. However, recent work suggests that this is not such a worry for our studies, which focus on political psychological questions and use random assignment: Clifford and colleagues (2015) find that correlations between various psychological variables (e.g., personality, values) and political identification are of comparable size and direction on MTurk as in benchmark national samples of U.S. adults. This suggests that MTurk samples are not wholly politically dissimilar from the general population. In addition, several recent studies report that average treatment effects

estimated in experiments on convenience samples (such as MTurk) tend to align well with those obtained from general population samples (Coppock et al., 2018; Mullinix et al., 2015)

Belief Update Task

The task has two phases. In phase one (P1), subjects judged the truth of 16 political statements that were presented sequentially. We refer to these judgments as their *prior beliefs*. Subjects were informed before P1 began that the statements are either true or false, but not told whether any given statement is true versus false. Prior beliefs were elicited as percentages on a 0-100 sliding scale (whole integers only) anchored at "certainly false" (0) and "certainly true" (100). The 16 political statements corresponded to 16 trials in P1. Of these 16 statements, 8 were pro-Democrat (i.e., favorable to the Democratic Party if true); 4 of which were true, and 4 of which were false. The remaining 8 statements were pro-Republican (favorable to the Republican Party if true); again, 4 were in fact true and 4 were false. The political statements are real statements taken from fact-checking websites such as www.politifact.com, and were selected via pre-testing according to several criteria (reported in the SI). For example, the statements were selected such that supporters of one party would not receive statements that were much more (or less) likely to be true than supporters of the other party, on average. Four of the statements are shown in Table 1. The statements are unlikely to be perfectly balanced on all attributes—for example, a higher proportion of the "pro-Republican" statements could be considered more anti-Democrat than the "pro-Democrat" statements considered anti-Republican. Nevertheless, the statements are closely balanced on the attributes that we consider to be most important for the present investigation (see SI for further details).

Table 1. Four Example Political Statements used in Studies 1 and 2.

	Pro-Democrat	**Pro-Republican**
True	During former President Barack Obama's final 4 years in office, wages of the average American worker went up.	Under President Donald Trump's administration, unemployment has fallen to a 17-year low.
False	Within two years of "Obamacare" being signed into law, health care premiums were going up more slowly than at any time in the previous 50 years.	Only 10 cents on every dollar from the Clinton Foundation goes to charitable causes.

Note. Statements were true or false at the time the studies were conducted.

Immediately after rating each political statement in P1, subjects received a signal about whether that statement is true or false. These signals simply reported the word TRUE or the word FALSE. Signals were accurate with probability 2/3. Thus, signals provided noisy but on average informative evidence about the truth of the political statements. Subjects were told the 2/3 probability of receiving an accurate signal and answer comprehension questions before P1 to ensure their understanding (verbatim task instructions are available on the OSF) (subjects were required to provide correct responses to the comprehension questions before they could proceed with the study). Subjects were reminded of the 2/3 probability of receiving an accurate signal after receiving each signal in P1.

Immediately after P1 had ended, subjects moved onto P2. In P2, subjects saw each political statement from P1 again (sequentially), and provided another judgment about the truth of each statement; their *posterior beliefs*. They were not reminded of their prior belief nor the signal they received in P1. Upon completing P2, the task was over. The presentation order of the

political statements in P1 and P2 was randomized (separately). Before P1, all subjects completed a practice P1 and P2 trial comprising a statement unrelated to politics (not included in analysis).

Bayesian Benchmark

We compute the posterior beliefs of a normative agent by conditioning on subjects' prior belief and the signal they received on each trial using Bayes' rule,

$$P(T|S) = \frac{P(T)P(S|T)}{\left((P(T)P(S|T) + P(\neg T)P(S|\neg T)\right)}$$

where P(T|S) is the posterior probability that the statement is true, given the signal received on that trial; P(T) is the prior probability that the statement is true, provided by the subject on each trial of P1; P(¬T) is the prior probability that the statement is not true (i.e., false; in this case, equal to 1 - P(T)); P(S|T) is the probability of receiving the signal assuming the statement is true; and P(S|¬T) is the probability of receiving the signal assuming the statement is not true (i.e., false; in this case, equal to 1 - P(S|T)). P(S|T) is equal to 2/3 when the signal reports TRUE and 1/3 when the signal reports FALSE. These computations are conducted on the 0-1 scale but are converted back to 0-100 for analysis. We compute two outcome variables to quantify subjects' deviation from the Bayesian posterior belief on each trial:

(1) The **difference** between the subject's posterior and Bayesian posterior;

(2) The **ratio** of the subject's posterior to the Bayesian posterior

To illustrate each outcome variable, suppose that Jane reports a prior belief of 60% and receives a signal of TRUE regarding political statement *i*. She subsequently reports a posterior belief of 80% that the statement is true. The Bayesian posterior on this trial is calculated as 75%,

$$P(T_i|S_{TRUE}) = \frac{0.6 \times \frac{2}{3}}{\left(0.6 \times \frac{2}{3} + 0.4 \times \frac{1}{3}\right)} = 0.75$$

Thus, for this trial the value of the difference outcome variable is equal to 80 – 75 = 5; and the ratio outcome variable is equal to 80/75 ≈ 1.07. We take the natural logarithm of the latter outcome variable for all analyses due to asymmetry in the untransformed variable (e.g., belief updating 10x less than Bayesian = 0.1, but belief updating 10x more than Bayesian = 10). This transformation was preregistered. The two outcome variables are computed such that, relative to the Bayesian benchmark, values < 0 imply that the subject's posterior belief fell short of the benchmark, and values > 0 imply their posterior belief went beyond what the benchmark required. Accordingly, values of zero imply that the subject's posterior belief approximated the Bayesian expected posterior belief.

These two outcome variables provide advantages over an operationalization of belief updating that focuses merely on the discrepancy between the subject's prior and posterior belief. Most clearly, they afford comparison against a Bayesian benchmark and thus better insight into whether and where there is bias in belief updating (Eil & Rao, 2011; Shah et al., 2016). However, there are other ways of evaluating deviation from the Bayesian benchmark. For example, one could compare the direction and magnitude by which the subject's belief changed vs. the direction and magnitude of belief change expected on a Bayesian evaluation (Peterson et al., 1965; Peterson & Miller, 1965; Phillips & Edwards, 1966)[1]. These different operationalizations possess different strengths and weaknesses which must be borne in mind.

For example, directly comparing the subject and Bayesian posterior beliefs means that deviations from the benchmark can receive different classifications according to the difference

[1] We are grateful to an anonymous reviewer for directing us to this operationalization.

and ratio outcome variables. To illustrate, suppose that on one trial Jane gives a posterior belief of 46% and the Bayesian posterior is 47%. On a second trial, she gives a posterior of 86% and the Bayesian posterior is 87%. The difference outcome variable classifies these two trials as identical deviations (i.e., a deviation of 1). However, the ratio outcome variable classifies the first trial as a greater deviation than the second trial because 46/47 is more distant from 1 than is 86/87. This example illustrates (i) the importance of analyzing both outcome variables and (ii) the limits inherent to scale-reported beliefs in general (Shah et al., 2016).

Indeed, one could go further still by considering a belief change operationalization that computes the ratio of subjects' belief updating to that expected by Bayesian updating; a so-called "accuracy ratio" score (Peterson et al., 1965; Peterson & Miller, 1965; Phillips & Edwards, 1966). On this operationalization, the *second* trial from above would be classified as a greater deviation, because updating from a subjective probability of .86 to .87 calls for stronger evidence than updating from .46 to .47. In this case, the decisive factor determining classification is the distance from the scale midpoint (50).

Ultimately, we chose to focus on the current operationalization—a comparison of subject posterior beliefs to Bayesian posterior beliefs using the difference and ratio outcome variables above—because (i) we consider it the most readily interpretable, and (ii) to be consistent with closely-related previous work (Eil & Rao, 2011; Shah et al., 2016). Nevertheless, in the SI we also report exploratory analyses using a belief change operationalization (under heading "accuracy ratio outcome variable"). The key results from that analysis are substantively similar to the results reported below. For ease of exposition, we thus use "deviation from Bayesian posteriors" and "belief updating" largely interchangeably during our results section.

Additional Variables

Following the belief update task, subjects complete a memory test (for exploratory purposes); they are shown each political statement again, and must indicate whether the signal they received reported TRUE or FALSE for that statement.

Subjects then complete a 7-item version of the Cognitive Reflection Test (CRT), comprised of a reworded version of the original 3-item CRT (Shenhav et al., 2012)—which consists of numeric problems—and a 4-item non-numeric CRT (Thomson & Oppenheimer, 2016) (items are reported in the SI). The CRT is a behavioral task considered to indicate the propensity to engage in "analytic thinking," which is the ability to override intuitive but incorrect responses (Frederick, 2005; Pennycook et al., 2016). However, the test also indicates cognitive ability per se: CRT scores are positively correlated with performance on a variety of cognitive ability and "rational thinking" tests (Toplak et al., 2011), and with latent general mental ability (*g*) (Blacksmith et al., 2019). Test-retest reliability estimates for the original 3-item measure range from 0.75 to 0.81 (Stagnaro et al., 2018). Correct responses are summed to create a 0-7 score for each subject (α = .79, 95% CI [.77, .82]).

Finally, subjects complete a self-report scale that asks about their responsiveness to new evidence (for exploratory purposes), and they provide basic demographic information including which U.S. political party they prefer: Democratic or Republican.

Results

Descriptive Statistics

Figure 1 displays the posterior beliefs provided by subjects (on the raw data scale) as a function of the expected posterior belief according to Bayes' rule, subjects' CRT performance, and each of the 16 political statements in Studies 1 and 2. Further descriptive statistics for both Studies 1 and 2 are reported in the SI, including distributions of prior beliefs for each political statement, raw magnitudes of belief change, CRT sum scores, and political identities.

The raw data in Figure 1 suggest an association between CRT performance and deviation from Bayesian posterior beliefs: Moving from left to right across the panels (low to high CRT performance) corresponds with closer adherence to the expected Bayesian posterior belief (dashed black line on the diagonal). In particular, the solid colored slopes are steeper towards the diagonal at high CRT, suggesting greater sensitivity to the signals (Eil & Rao, 2011), and there also appears to be reduced scatter about the diagonal at the highest levels of CRT. In the next section, we examine these trends statistically, and we also examine the link between the political favorability of the signals and subjects' deviation from the Bayesian benchmark.

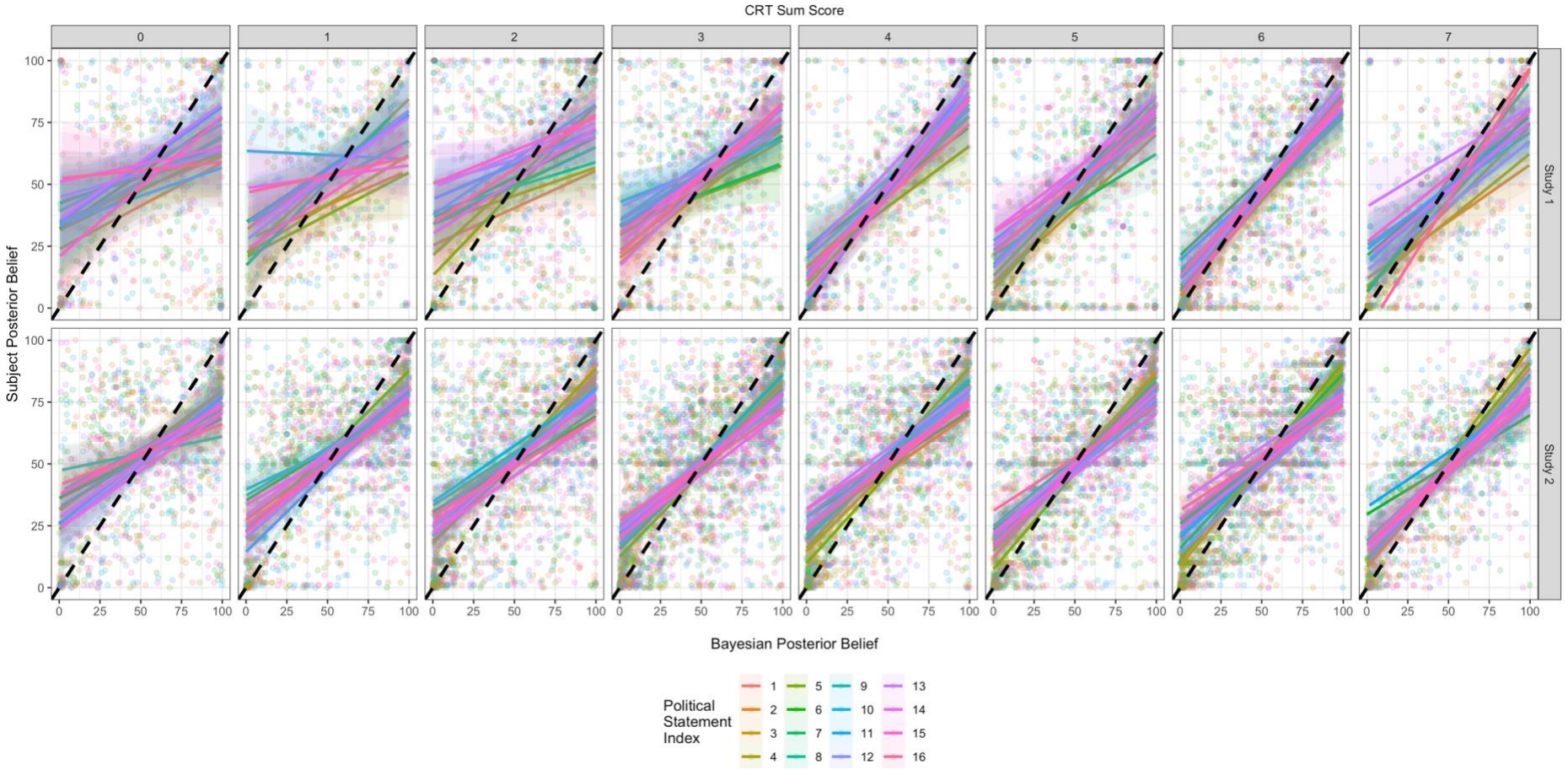

Figure 1. Subject posterior beliefs as a function of expected Bayesian posterior beliefs, CRT performance, and political statements in Studies 1 and 2. The top line panels (indexed 0-7) correspond to sum score on the CRT. The dashed black line on the diagonal indicates a perfect fit between expected Bayesian posterior beliefs and subject posterior beliefs. The solid colored lines are linear fits to the data. Shaded bands are 95% CI.

Preregistered Data Exclusions

For both hypothesis tests, we exclude duplicate subjects (according to IP address, N = 9, 1.80%). We also exclude trials on which subjects' prior belief was given as 0 or 100 and the signal reported FALSE or TRUE, respectively (N_{trials} = 254, 3.23%). This is because updating towards the signal is not possible on these trials. (In both Studies 1 and 2, prior beliefs of 0 and 100 were treated as 0.5 and 99.5, respectively, for the purposes of computing Bayesian posterior beliefs. This was preregistered.) Subjects who did not report a U.S. political party preference were excluded from the test of Hypothesis 2 (N = 2, 0.41%).

Hypothesis 1: Magnified Normative Posterior Beliefs

After data exclusions, we retain N=492 subjects for the preregistered test of Hypothesis 1. H1 is that people who score higher on the CRT deviate less from the Bayesian benchmark in their posterior beliefs. To test H1, we first compute absolute values of both outcome variables (i.e., |difference|, |log(ratio)|). These absolute values thus represent deviation from Bayesian posteriors collapsing over the particular *direction* of the deviation; that is, collapsing over whether the posteriors fell-short-of or went further than that prescribed by Bayes' rule (the specific direction of the deviation is the focus of H2). We then compute the mean of these two outcomes over the 16 trials for each subject. These mean absolute deviation values are displayed in Figure 2A as a function of subjects' performance on the CRT.

We conduct nonparametric Kendall's tau (τ) correlation tests between these outcome values and CRT performance. CRT performance is negatively correlated with mean absolute deviation from Bayesian posteriors, though the association is larger for the difference outcome variable [τ = -.20, Z = -6.27, p <.001] than the log(ratio) outcome variable [τ = -.06, Z = -1.72, p =.086]. In the SI, we conduct additional analyses that test the robustness of this main result. These analyses indicate that the result is robust to different analytic specifications and potential confounds. Therefore, broadly consistent with H1, subjects who scored higher on the CRT

deviated less from the Bayesian posteriors overall, indicating that they combined their prior beliefs with the new information in a more normative manner.

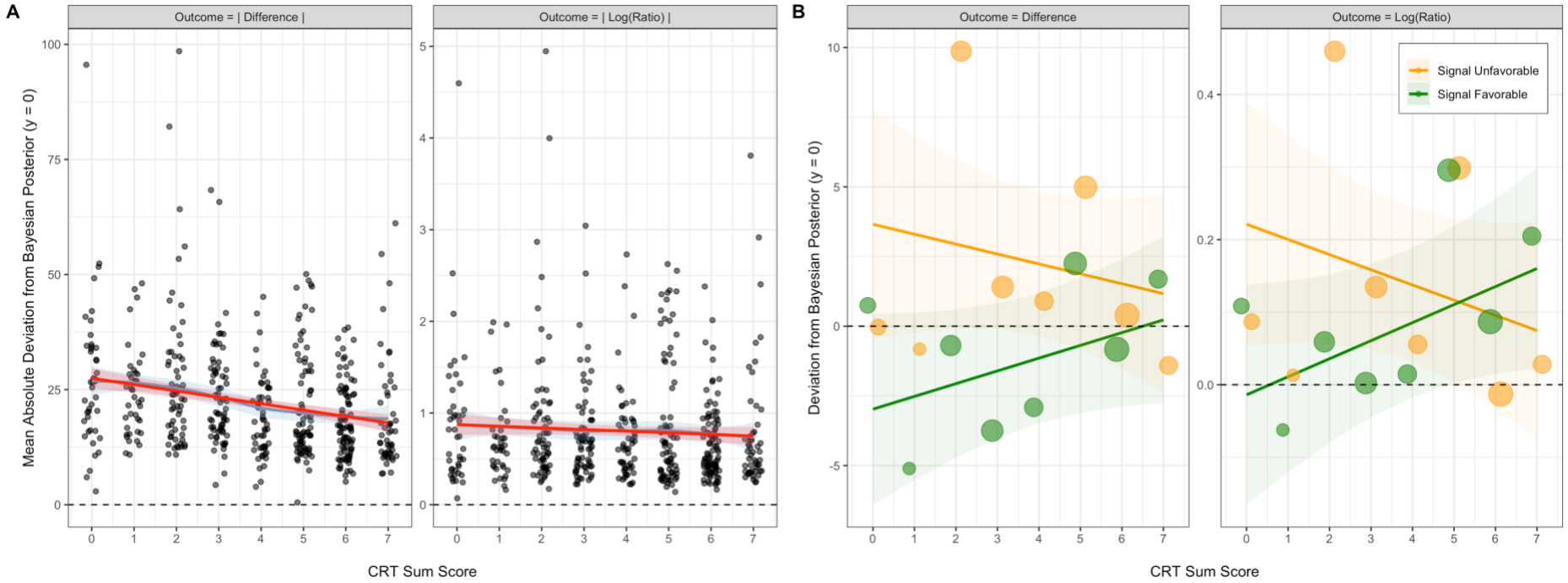

Figure 2. Deviation from Bayesian posterior beliefs as a function of CRT performance and the political favorability of signals in Study 1. **A**, mean absolute deviation from the Bayesian posterior (represented at y = 0). One point corresponds to the mean absolute deviation of one subject. Data points have slight horizontal jitter to aid visibility. Solid lines are linear regressions (red) and locally-weighted regressions (blue) fit to the data. N=492. **B**, model-predicted magnitude and direction of deviation from the Bayesian posterior (y = 0). Data points represent the means computed on the raw data. To compute these means, the mean is first taken for each subject over their 16 trials, and then the mean of these values is taken for each CRT sum score over all subjects. The size of the data points corresponds to the number of subjects with that CRT score (N range = [39, 92]). **A**, **B**, shaded regions are 95% CI. CRT = Cognitive Reflection Test.

Hypothesis 2: Magnified Political Bias in Posterior Beliefs

After data exclusions, we retain N=490 for the preregistered test of H2. H2 is that people who score higher on the CRT are more politically biased in their belief updating, where bias is defined as posterior beliefs that go "too far" on politically favorable signals (deviations > 0) and posterior beliefs that "fall short" on politically *un*favorable signals (deviations < 0). Therefore, the critical test of H2 is on the interaction between CRT performance and the political favorability of signals in predicting deviation from Bayesian posterior beliefs. To classify whether signals are politically favorable or unfavorable, we consult subjects' political party preference (Democratic or Republican), the pre-tested partisanship of the political statement (pro-Democrat or pro-Republican), and the signal report they received (TRUE or FALSE). This classification is displayed in Table 2.

Table 2. Classifying the Political Favorability of Signals.

Subject Party Preference	**Partisanship of Political Statement**	**Signal Report**	**Political Favorability**
Democratic Party	Pro-Democrat	True	Favorable
		False	Unfavorable
	Pro-Republican	True	Unfavorable
		False	Favorable
Republican Party	Pro-Democrat	True	Unfavorable
		False	Favorable
	Pro-Republican	True	Favorable
		False	Unfavorable

Note. The partisanship of the political statements was determined in a pre-test study (see SI).

Analysis strategy. We fit two linear mixed effects models to the trial-level data, one model for each outcome variable [difference, log(ratio)]. Recall that these outcome variables are no longer the absolute transformations used in the test of H1. Therefore, here the values can be above zero and below zero, corresponding to over- and under-updating relative to the Bayesian benchmark, respectively.

Linear mixed effects models offer several advantages over classic ANOVA for analyzing the type of data produced by our design (Barr et al., 2013; Brauer & Curtin, 2018; Judd et al., 2012). Most notably, they allow us to model the non-independence of the data across subjects and statements while maintaining the nominal Type I error rate of 5%. Concretely, the effect of our predictors of interest—CRT performance, the favorability of signals, and their interaction—plausibly vary between subjects *and* between political statements, known as "random" effects. Modelling this variance maintains the Type I error rate (5%) on the fixed effect estimate of interest—in our case, the interaction between CRT performance and signal favorability.

In both outcome models, we first attempt to fit a "maximal" random effects structure, allowing all relevant predictor effects to vary between subjects and between political statements, as well as estimating the full set of covariances. This provides a conservative test of the fixed effect of interest (Barr et al., 2013). However, these models can sometimes be too complex for the data, and do not converge in estimation of the parameters. When this happens, our strategy is to incrementally simplify the random effects structure until model convergence is achieved (Brauer & Curtin, 2018). After model convergence is achieved, we use a Likelihood Ratio Test (LRT) to statistically evaluate whether the fixed effect interaction between CRT and signal favorability improves model fit. This is the key test of H2. To aid understanding of the fitted models, we report the specification of the converged model (in *R* syntax) in text.

Model results.

Difference outcome variable. The maximal model for the difference outcome variable did not converge. Thus, we incrementally simplify the random effects structure to achieve convergence. The specification of the converged model is,

$$
\begin{aligned}
&difference \sim \\
&CRT + favorable\ signal + CRT{:}\ favorable\ signal + \\
&(1 + favorable\ signal \mid Subject) + \\
&(1 + CRT \mid Statement)
\end{aligned}
$$

The first line specifies the outcome variable; the second line specifies the fixed effects; and the third and fourth lines specify the random effects on subjects and statements, respectively. This model specification allows the intercept parameter to vary between subjects and between statements; and the effect of signal favorability and CRT to vary between subjects and between statements, respectively. In this model, the fixed effect interaction between CRT performance and signal favorability is statistically significant (Table 3 displays LRT results for the key interaction in both Studies 1 and 2). Full model estimates are reported in the SI.

Table 3. Likelihood Ratio Tests on the Fixed Effect Interaction in Studies 1 and 2.

Study	Outcome in Model	χ^2	df	p
1	Difference	3.97	1	.046
	Log(Ratio)	6.81	1	.009
2	Difference	6.56	1	.010
	Log(Ratio)	4.81	1	.028

To interpret the interaction, we generate predicted deviations from the Bayesian posterior according to the model. These are shown in the slopes in Figure 2B. The slopes show that the average marginal effect of CRT performance on deviation from the Bayesian posterior is negatively signed for unfavorable signals, and positively signed for favorable signals (Table 4 displays statistical evaluation of the average marginal effects of CRT in Studies 1 and 2). In other words, this pattern suggests that high (versus low) CRT scorers tended to update less on unfavorable signals and more on favorable signals. This is seemingly consistent with H2.

However, evaluated against the Bayesian benchmark (y = 0), high CRT scorers in fact tend toward the Bayesian posterior in both cases. This is inconsistent with a straightforward interpretation of magnified political bias (H2)[2]. Furthermore, examining high CRT scorers specifically, deviation from Bayesian on favorable signals is not systematically greater than deviation on unfavorable signals. The direction of the slopes instead reflects the fact that the model expects low CRT scorers to *over*-update on unfavorable signals and *under*-update on favorable signals (relative to Bayesian); that is, seeming *anti*-politically biased updating. Taken together, these patterns are inconsistent with the expectation that people who score higher on the CRT are more politically biased in their belief updating. The clearer result is that they are more normative overall (H1).

Log(ratio) outcome variable. The maximal model for the log(ratio) outcome also did not converge. Model convergence is achieved with the specification,

2 A pattern more indicative of magnified political bias in these data would be if the posterior beliefs of all subjects fell short of the Bayesian posterior on average, but the posteriors of those subjects scoring high (versus low) on the CRT fell *even shorter* when provided unfavorable signals, but showed no such difference when the signals are favorable. With respect to Figure 2B, this would be represented by a steep downward yellow slope that has an intercept <0; and a flat green slope with an intercept <0. In Study 1, the intercepts of both slopes are relatively close to zero, meaning such a pattern is not easily observed. However, in Study 2, all intercepts are <0, but we still do not see the aforementioned pattern.

$$\log(ratio) \sim$$
$$CRT + favorable\ signal + CRT{:}\ favorable\ signal +$$
$$(1 + favorable\ signal \mid Subject) +$$
$$(1 + favorable\ signal \mid Statement)$$

In this model, the fixed effect interaction between CRT and signal favorability is statistically significant (Table 3). As before, to interpret the interaction effect we generate and plot deviations from the Bayesian posterior as predicted by the model (Figure 2B). The slopes are substantively similar in their implication to the difference outcome model. That is, the patterns suggest high (versus low) CRT scorers tended to update less on unfavorable signals and more on favorable signals. However, evaluated against the Bayesian benchmark, it is difficult to interpret the posteriors as more politically biased. Rather, they appear to be more normative overall, albeit in a less compelling fashion than is implied by the difference outcome model.

Table 4. Average Marginal Effect of CRT in Studies 1 and 2.

Study	Outcome in Model	Signal	AME	SE	z	*p*	95% lower	95% upper
1	Difference	Unfavorable	-0.76	0.88	-0.87	.386	-2.48	0.96
		Favorable	0.98	0.68	1.44	.149	-0.35	2.31
	Log(Ratio)	Unfavorable	-0.04	0.04	-1.22	.223	-0.12	0.03
		Favorable	0.05	0.03	1.68	.093	-0.01	0.12
2	Difference	Unfavorable	0.47	0.40	1.17	.244	-0.32	1.26
		Favorable	1.50	0.37	4.10	< .001	0.78	2.22
	Log(Ratio)	Unfavorable	-0.01	0.01	-0.69	.492	-0.04	0.02
		Favorable	0.02	0.01	1.78	.075	-0.00	0.05

Note. Average marginal effects (AMEs) are evaluated using the *margins* package (Leeper et al., 2018) available in *R*. CRT = Cognitive Reflection Test. SE = Standard Error.

Study 1 Summary

In Study 1, we estimate that people who score higher on the CRT deviate less overall from the expected Bayesian posteriors, suggesting they combined their prior beliefs with the new information in a more normative manner. This is consistent with H1. We find some evidence to suggest that high CRT scorers updated their beliefs more (less) than low CRT scorers on favorable (unfavorable) signals, seemingly consistent with greater political bias, as per H2. However, evaluating against the Bayesian benchmark suggests this pattern is driven mostly by low CRT scorers engaging in seeming *anti*-politically biased updating. Overall, the clearer result from Study 1 is that high CRT scorers' posterior beliefs were closer to the Bayesian benchmark.

Nevertheless, a potential confound in the design of Study 1 is the lack of a control group: we did not have any subjects who received no new information and were thus not expected to update their beliefs. This is a weakness because we take repeated measurements of people's beliefs, and repeated measurements of unstable variables like beliefs are prone to regression to the mean (RTM) (Yu & Chen, 2015). Without a control group who receive no new information, changes in beliefs between the prior and posterior could be due to the new information, as we assumed—but they could also be due to RTM. Importantly, if RTM is negatively correlated with CRT performance, the association between CRT performance and deviation from the Bayesian benchmark that we observe could be an artefact of the design.

To exclude this possibility, we conducted an identical version of Study 1, but, this time, we did not provide subjects new information about the truth or falsity of the political statements. More specifically, the signals were still randomly assigned, but were hidden from view. Therefore, any "updating" that occurs between the time 1 and 2 belief measurements is attributable to random variation and RTM only. If CRT performance predicts lesser deviation from the Bayesian benchmark in this case, it suggests that differences in RTM rather than differences in belief updating *per se* drive the estimates in Study 1 (Yu & Chen, 2015). The analysis protocol for this control study was identical to Study 1 and the results are reported in the

SI. The key result is that the negative correlation between CRT performance and deviation from the Bayesian benchmark is not observed in the data from the control study, on either outcome variable. Thus, we conclude that the Study 1 estimates are not due to a design artefact and RTM.

Study 2

We conducted a second study to address a more substantive limitation of Study 1. That is, in Study 1 we imposed on subjects the diagnosticity of the signals; the "likelihood ratio" in Bayesian terms. The likelihood ratio indicates how much more likely it is to observe the information when one hypothesis holds (e.g., the statement is true) versus when the hypothesis does not hold (the statement is false). Thus, it is a measure of how diagnostic the information is and how much updating "should" happen, according to Bayes' rule.

In Study 1, we informed subjects of the (true) probability that the signals were accurate (2/3)—giving a fixed likelihood ratio of 2 for signals that reported TRUE (i.e., 2/3 ÷ 1/3), and ½ for signals that reported FALSE (1/3 ÷ 2/3)—and we assumed that all subjects applied this knowledge similarly in their updating behavior. This is a strong assumption. As outlined in the Introduction, previous work suggests that CRT performance correlates with greater political bias in people's self-reported evaluations of new information; and, importantly, self-reported evaluations of information have been interpreted by some researchers as a proxy for the likelihood ratio that people assign the information (Kahan, 2013, 2016). Insofar as this is the case, then, the association we observed between CRT performance and deviation from Bayesian posteriors in Study 1 may have been due to an association between CRT performance and reasoning about the likelihood ratio, rather than belief updating per se. That is, in principle all subjects could have updated their beliefs identically, with the observed differences being driven by different assessments of the likelihood ratio alone.

To mitigate this limitation, in Study 2 we ask people for their subjective judgments regarding the overall diagnosticity of the signals in the task. We then use these idiosyncratic judgments in conjunction with their prior beliefs to compute the Bayesian posteriors. This isolates the process of belief updating as distinct from the process of reasoning about the likelihood ratio. Thus, removing a potential confound in the link between CRT performance and deviation from Bayesian posteriors per se. In short, because each person now provides their own subjective judgment about the diagnosticity of the signals, any differences in perceptions of the likelihood ratio due to CRT performance are "built in" to our estimates of their posterior beliefs.

Finally, in Study 2 we also ask subjects to provide an accuracy rating of each individual signal. This allowed us to replicate the previously described finding (Kahan, 2013) that high CRT scorers are the most politically polarized in their evaluations of new information (distinct from their belief updating on the basis of that information), but in the context of our novel design. This information-evaluation analysis was preregistered, but, for brevity and because it was not the primary focus of the two studies, we report it in the SI (under heading "Hypothesis 3").

Methods

Study 2 was preregistered at https://osf.io/9yj57.

Sample

We sought to double the sample size from Study 1 and recruit N=1000 subjects. A total of N=1004 subjects completed the study (recruited via MTurk). Subjects who completed Study 1 were unable to take part in Study 2.

Belief Update Task

The task is identical to Study 1, with two adjustments. First, subjects were not told the exact probability that signals were accurate. Instead, they were simply informed that signals were

accurate with some fixed probability, but that signals were, on average, accurate (i.e., P(accurate) > 0.5). Second, upon receipt of each signal in P1, subjects reported whether they believed that that *particular* signal was accurate (on a 1-5 scale, 1=Definitely NOT, 5 = Definitely YES). We refer to this as the "signal accuracy" rating. Following each of these ratings, subjects were asked to consider all the signals they had seen thus far in the task, and to provide a judgment about the *overall* likelihood of receiving an accurate signal in the task (i.e., what they perceived to be the fixed probability of receiving an accurate signal in the task). We refer to these as "likelihood judgments." The likelihood judgments were provided on a sliding scale in whole integers, from 51% (labelled "signals are almost random/uninformative") through 75% ("signals are mostly accurate/quite informative") to 99% ("signals are almost perfectly accurate/very informative").

Bayesian Benchmark

Bayesian posterior beliefs are computed as in Study 1. However, because we obtain people's subjective judgments about the diagnosticity of the signals in the task (their "likelihood judgments") we use this information to compute Bayesian posterior beliefs that are unique to each subject's prior belief *and* implied likelihood ratio, P(S|T) / P(S|¬T). Concretely, to compute the Bayesian posterior for a given political statement, we consult the signal received by the subject for that statement (i.e., "TRUE" or "FALSE"), and their likelihood judgment, *LHJ*, provided immediately after the focal statement-signal pairing (the LHJ is scaled to lie between 0-1 for the computations). Formally, when the signal reports TRUE,

$$P(T|S_{TRUE}) = \frac{P(T)LHJ}{(P(T)LHJ + P(\neg T)(1 - LHJ))}$$

When the signal reports FALSE,

$$P(T|S_{FALSE}) = \frac{P(T)(1-LHJ)}{(P(T)(1-LHJ) + P(\neg T)LHJ)}$$

We chose to consult the likelihood judgment subjects provided after the focal statement-signal pairing (rather than before) to allow for the possibility that subject's impression of the informativeness of the signals may immediately change upon receipt of the focal signal, affecting their updating on the focal statement (Cheng & Hsiaw, 2019). This choice was preregistered. To illustrate, assume Jane reports a prior belief of 70% that statement *i* is true, and she receives a signal of TRUE for that statement. On this trial, she judges that most signals in the task are likely to be accurate—say, her likelihood judgment is 80%. The Bayesian posterior for statement *i* is thus approximately 90%,

$$P(T_i|S_{TRUE}) = \frac{0.7 \times 0.8}{(0.7 \times 0.8 + 0.3 \times 0.2)} \approx 0.90$$

If Jane's subjective likelihood judgment on this trial had instead been 55% – indicating that she considered signals only weakly informative at this point during the task – her Bayesian-predicted posterior belief would have equalled approximately 74%. After computing by-trial Bayesian posterior beliefs for each subject according to this method, we compute the two outcome variables [difference, log(ratio)] exactly as in Study 1.

Additional Variables

Following the belief update task, subjects complete the 7-item CRT (α = .76, 95% CI [.74, .79]) as in Study 1, and provide basic demographic information including political party preference (Democratic or Republican). There was no memory test or other exploratory measures in Study 2.

Results

Preregistered Data Exclusions

As in Study 1, for all hypothesis tests we exclude duplicate subjects according to IP address (N = 12, 1.20%). For the tests of H1 and H2, we also exclude trials on which subjects' prior belief was given as 0 or 100 and the signal reported FALSE or TRUE, respectively (N_{trials} = 346, 2.18%). There were no missing values for political party preference.

Hypothesis 1: Magnified Normative Posterior Beliefs

After data exclusions, we retain N=992 for the preregistered test of H1. The analysis plan is the same as in Study 1. Subjects' mean absolute deviation from the Bayesian posterior as a function of their CRT performance is displayed in Figure 3A. Nonparametric Kendall's tau (τ) correlation tests show that, consistent with H1, CRT performance is negatively correlated with mean absolute deviation from Bayesian posteriors: As in Study 1, the association is larger for the difference outcome variable [τ = -.18, Z = -8.04, p <.001] than the log(ratio) outcome variable [τ = -.07, Z = -3.14, p =.002]. Robustness checks (reported in the SI) suggest that this result is robust to alternative analytic specifications. Replicating Study 1 and consistent with H1, subjects who scored higher on the CRT deviated less from Bayesian posteriors overall, suggesting that they combined their prior beliefs with the new information in a more normative manner.

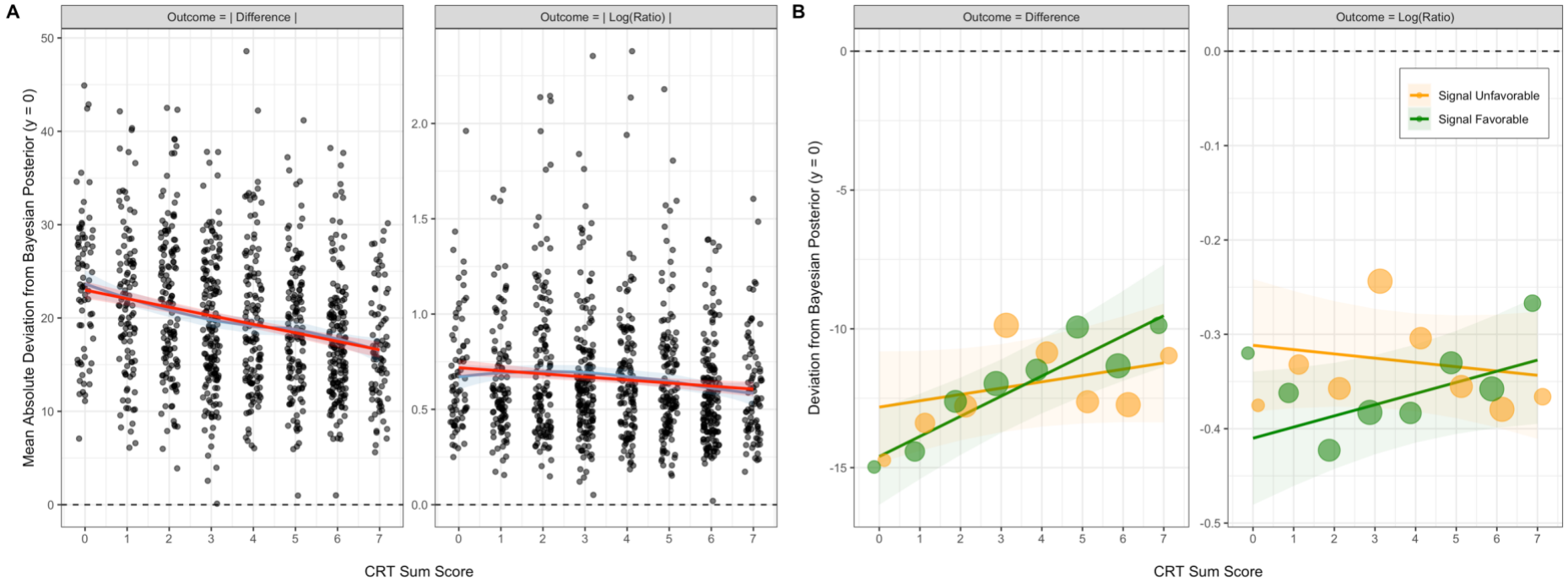

Figure 3. Deviation from Bayesian posterior beliefs as a function of CRT performance and the political favorability of signals in Study 2. **A**, mean absolute deviation from the Bayesian posterior (represented at y = 0). One point corresponds to the mean absolute deviation of one subject. Data points have slight horizontal jitter to aid visibility. Solid lines are linear regressions (red) and locally-weighted regressions (blue) fit to the data. N=992. **B**, model-predicted magnitude and direction of deviation from the Bayesian posterior (y = 0). Data points represent the mean computed on the raw data. To compute these means, the mean is first taken for each subject over their 16 trials, and then the mean of these values is taken for each CRT sum score over all subjects. The size of the data points corresponds to the number of subjects with that CRT score (N range = [73, 167]). **A**, **B**, shaded regions are 95% CI. CRT = Cognitive Reflection Test.

Hypothesis 2: Magnified Political Bias in Posterior Beliefs

After data exclusions, we retain N=992 for the preregistered test of H2. The analysis strategy is identical to Study 1.

Model results.

Difference outcome variable. The maximal model for the difference outcome variable did not converge. Incrementally simplifying the random effects structure to achieve convergence resulted in model specification,

$$
\begin{aligned}
&difference \sim \\
&CRT + favorable\ signal + CRT{:}\ favorable\ signal + \\
&(1 + favorable\ signal \mid Subject) + \\
&(1 + CRT + favorable\ signal \mid Statement)
\end{aligned}
$$

In this model, the fixed effect interaction between CRT and signal favorability is statistically significant (Table 3). As in Study 1, to interpret the interaction we generate and plot deviations from the Bayesian posterior as predicted by the model (Figure 3B). Examining the slopes in Figure 3B with respect to the Bayesian benchmark, the impact of incorporating people's subjective likelihood judgments in Study 2—rather than imposing 2/3 as in Study 1—is abundantly clear. Specifically, subjects' posterior beliefs are substantially *less* than that prescribed by Bayes' rule (i.e., the average implied deviation from Bayesian is well below zero). This is consistent with past studies of human belief updating, where "conservatism" in updating is regularly observed (Hahn & Harris, 2014), and, taken together with those studies, suggests that subjects did not uniformly apply the likelihood of 2/3 imposed in Study 1.

The slopes in Figure 3B show that the marginal effect of CRT performance on deviation from the Bayesian posterior is positively signed for both favorable signals and unfavorable signals, but is stronger for the former (Table 4 contains statistical evaluation of the slopes

themselves). In other words, this pattern suggests that high (versus low) CRT scorers tended to update more on both types of signal, but this updating was relatively greater on signals that were favorable. This is broadly consistent with magnified political bias, as per H2. At the same time however, as in Study 1, the general pattern is also consistent with the conjecture that high CRT scorers are more normative in their belief updating overall, because both slopes tend toward the Bayesian posterior. Moreover, looking at high CRT scorers specifically, there is minimal evidence to suggest that favorable signals were met with greater belief updating than unfavorable signals; instead, the two slopes are largely overlapping.

Log(ratio) outcome variable. The maximal model for the log(ratio) outcome did not converge. Model convergence was achieved with the specification,

$$\log(ratio) \sim$$
$$CRT + favorable\ signal + CRT{:}\ favorable\ signal +$$
$$(1 \mid Subject) +$$
$$(1 \mid Statement)$$

In this model, the fixed effect interaction between CRT and signal favorability is statistically significant (Table 3). The model predicted deviations from the Bayesian posterior are displayed in Figure 3B. The slopes are similar to those observed in Study 1: Negatively signed for unfavorable signals and positively signed for favorable signals. The former negative slope would be most consistent with H2, because it suggests that high CRT scorers under-updated (relative to Bayesian) to a greater extent than low CRT scorers when receiving unfavorable signals. However, this slope is somewhat flat and not significantly different from zero (p =.492, Table 4).

Discussion

In this paper, we reported two experiments whose primary aim was to test the hypothesis that cognitive sophistication is associated with magnified political bias in belief updating. We also tested the distinct hypothesis that cognitive sophistication is associated with more normative (i.e., less biased) political belief updating. Overall, we found little direct evidence for the former hypothesis, but somewhat clearer evidence for the latter hypothesis.

Specifically, we observed consistent evidence to suggest that analytic thinking—as inferred via performance on the Cognitive Reflection Test—was associated with posterior beliefs closer to the Bayesian benchmark. The association was small, but was apparent in both studies and across outcome variables. In contrast, while we observed fairly consistent evidence to suggest that higher CRT scorers updated more (less) on politically favorable (unfavorable) signals than lower CRT scorers, the Bayesian benchmark implies that these patterns offer little evidence of magnified political bias. This is because subjects who scored lower on the CRT tended to report posterior beliefs that exceeded the benchmark on politically *un*favorable information, but fell short of the benchmark on politically favorable information; a seeming anti-political-bias (we discuss this result further below). The patterns of posterior beliefs among higher CRT scorers, by contrast, mostly resulted in a correction of this tendency observed among lower CRT scorers; thus, resulting in more normative, rather than biased, posterior beliefs.

Implications for Hypotheses and Existing Research

The overriding conclusion we draw from our results is that individuals who scored higher on the CRT tended to be more appropriately-responsive to the new information than were their lower scoring counterparts. In particular, the patterns of posterior beliefs suggested they tended to combine their prior beliefs with the new information in a more normative manner to arrive at updated beliefs about the truth or falsity of the political statements. In contrast, there was minimal evidence that they were more politically biased in their belief updating.

Consequently, our results are at odds with the hypothesis that cognitive sophistication magnifies politically biased cognitive processing. They align somewhat more naturally with evidence that cognitive sophistication correlates with greater accuracy in beliefs about the truth of political news headlines (Pennycook & Rand, 2019), as well as with epistemic rationality in a variety of other domains (Pennycook et al., 2015).

Of course, "normative" can be defined in numerous ways. In our study design, it was defined as closeness to an unbiased Bayesian benchmark. This benchmark was defined as a function of individuals' prior beliefs about the truth of each political statement, as well as the detail we provided (Study 1) or subjectively elicited from (Study 2) them about the informativeness of the new information. Bayesian benchmarks of this kind are a common feature of experimental studies of belief updating (Coutts, 2019; Eil & Rao, 2011; Hill, 2017; Hornikx et al., 2018; Shah et al., 2016). They are useful insofar as they allow researchers to more clearly understand and diagnose systematic biases in human learning (Bullock, 2009; Hahn & Harris, 2014). However, it is important to emphasize that the Bayesian benchmark in general, and ours in particular, are not the only standards for belief updating with a claim to normativity. Furthermore, there is no *de facto* Bayesian benchmark, because the specific benchmark depends upon assumptions about the prior and the new information, both of which can be underspecified or biased by measurement, debated over, and flexibly modelled (Bowers & Davis, 2012; Tappin & Gadsby, 2019; Williams, 2018).

Nevertheless, to our knowledge, our studies are the first to investigate whether politically biased belief updating is both (a) magnified by cognitive sophistication and (b) represented in systematic deviation from an unbiased Bayesian benchmark. Previous studies have investigated either (a) or (b), but not both.

Most notably, in two highly influential studies, Taber and Lodge (2006) studied how U.S. subjects updated their beliefs about gun control and affirmative action after receiving information on these topics, and how cognitive sophistication moderated their belief updating.

The authors' design entailed exposing subjects to an "information board" where they were free to select into reading pro or con arguments about one of the topics under study, and, later in the design, all subjects received pro and con arguments regarding the other topic. The measure of cognitive sophistication was performance on a political knowledge quiz. The authors reported that subjects who scored highest on the quiz tended to show more extreme time 2 (versus time 1) political beliefs than those who scored lower on the quiz, consistent with the hypothesis that cognitive sophistication magnifies politically biased belief updating.

An important drawback of their design is that it is difficult to assess how subjects might have updated their beliefs were they politically *un*biased, because no unbiased benchmark was specified. Indeed, the authors wrestle with this problem in their interpretation of their results. While perfect knowledge of the politically-unbiased counterfactual is unattainable, our design reduces the uncertainty by explicitly defining an unbiased (albeit imperfect) benchmark. That is, the expected Bayesian posterior belief. In doing so, we do not find meaningful evidence that cognitively sophisticated subjects (i) deviated from this benchmark in a politically biased manner and (ii) deviated to a greater extent than their less sophisticated counterparts. On the contrary, we find a distinct lack of politically biased deviations from the benchmark (across all subjects); and, in addition, subjects who scored higher on the CRT deviated less from the benchmark overall. The discrepancy between our results and Taber and Lodge illustrates the importance of explicitly considering what belief updating should (or could) look like in the absence of political bias, and, furthermore, suggests that more work is needed to determine whether and when cognitive sophistication magnifies politically biased belief updating (see also Taber et al., 2009).

For example, one important difference between our study design and theirs is the political stimuli. In particular, although we pre-tested and selected our political statement stimuli to be clearly partisan in flavor, none were strongly politicized or culturally entrenched U.S. issues like gun control or affirmative action (see also Sunstein et al., 2016). Given that these and other such issues tend to be associated with the strongest partisan disagreement in the real world

(Drummond & Fischhoff, 2017; Kahan, 2015), they may be more likely contenders to reveal magnified-political-bias in belief updating among the more cognitively sophisticated partisans. This highlights a condition that may be necessary for cognitive sophistication to magnify politically biased belief updating: the issues may have to be strongly politicized and entrenched. This could explain why we observed limited political bias in posterior beliefs *per se* (across all subjects). Future work can formally evaluate this possibility by extending the current design to include political stimuli that draw upon strongly politicized and culturally entrenched issues. In the meantime, it is important that conclusions about magnified politically biased processing refrain from generalization across these different classes of political stimuli.

Limitations

Previous studies investigating whether cognitive sophistication magnifies politically biased processing have used designs that differ from ours on a number of additional dimensions. For example, as described in the Introduction, a key difference is the type of outcome variable: whereas we focused on posterior beliefs as they were affected by new information, most previous studies have focused on how individuals interpret the new information itself, without measuring whether or how the information changed their beliefs about the issue in question. While this difference was explicitly motivated by the drawbacks of the information evaluation outcome variable—as we describe in the Introduction—there are other design differences, too; such as the fact that previous studies tend only to provide one piece of information to subjects in the experiment—whereas subjects in ours received 16 signals in total (one per statement), in a repeated measures design. Repeated measures designs have been found to cause higher ability participants to adhere more to normative principles (LeBoeuf & Shafir, 2003), and thus our design may have affected the behavior of subjects who scored highly on the CRT more than it did those who scored low—limiting the generalizability of our results to designs that are strictly between-subjects.

Furthermore, also as highlighted in the Introduction, our design differed in the *richness* of the information provided to individuals: a signal that reports "TRUE" or "FALSE" is very different from a 2x2 contingency table (Kahan, Peters, et al., 2017; Nurse & Grant, 2019), or verbal arguments for or against public policy issues (Taber et al., 2009; Taber & Lodge, 2006). These design differences may also be important for understanding whether and when cognitive sophistication magnifies politically biased processing.

We adopted a narrow operationalization of cognitive sophistication; propensity for analytic thinking, inferred via performance on the Cognitive Reflection Test. While this choice places some constraints on the generalizability of our results vis-à-vis other domains of cognitive sophistication, we expect these constraints are somewhat minimal—for two reasons. First, performance on the CRT shares a robust and sizable correlation with diverse indicators of cognitive aptitude (Toplak et al., 2011), and with latent general mental ability (Blacksmith et al., 2019). Second, recent work that investigated the hypothesis that cognitive sophistication magnifies politically biased information evaluations used a variety of cognitive indicators alongside CRT, including a test of verbal and numeric intelligence, science reasoning, and political knowledge (Tappin et al., 2018). The relevant results of that work were highly similar across these diverse indicators. Together, these two reasons lead us to expect that the results we observe here would generalize to other measures of cognitive sophistication.

Our analysis operationalized bias in belief updating as deviation from the expected Bayesian posterior belief, similar to prominent recent work on related research questions (Eil & Rao, 2011; Shah et al., 2016). However, an alternative operationalization is to compare the direction and magnitude of belief *change* that is observed among subjects with the direction and magnitude of change expected on a Bayesian evaluation (Peterson et al., 1965; Peterson & Miller, 1965; Phillips & Edwards, 1966). As mentioned, we conducted exploratory analyses using this operationalization, reported in the SI (under heading "accuracy ratio outcome variable").

The results offer additional context for understanding why individuals scoring lower on the CRT tended to deviate more from the Bayesian posterior: compared to their high scoring counterparts, they more often changed their beliefs in a direction *opposite* to that expected by Bayes' rule. It is unlikely that this pattern reflects a "backfire" effect as commonly understood (Nyhan & Reifler, 2010); more likely, it is attributable to some combination of a failure to integrate the signals with prior beliefs per se, plus regression-to-the-mean and natural variation in the reporting of beliefs. Importantly, the latter processes (RTM/natural variation) do not seem able to fully explain our key result; as per our "control" study (see Study 1 Summary), we did not find that CRT predicted closer adherence to the Bayesian posteriors when signals were not available. Thus, some differential integration of the signals between low and high CRT scorers appears necessary to explain our results. Future work could directly explore this idea by using a design in which signals are harder to ignore or miss—for example, by asking individuals to confirm what the signal says, and to report their posterior beliefs immediately after each signal.

The stronger tendency for opposite-updating among lower CRT scorers highlights a possible explanation for why they exhibited a seeming anti-political-bias in their posterior beliefs. Specifically, if we assume that (i) low CRT scorers' beliefs were prone to greater variation in reporting than were higher CRT scorers', and (ii) all individuals' prior beliefs tended to align with their political leanings—that is, politically-favorable statements tended to be rated as true, a prior belief of > 50, and politically-*un*favorable statements as false, a prior belief of < 50 (e.g., see Figure S1 in the SI). Assuming that (i) and (ii) are correct, there would be more "room" for the greater variation in low CRT scorers' belief-reporting to manifest in the politically-*un*favorable direction on the belief scale: observed as a decreasing belief in politically-favorable statements, and an increasing belief in politically-unfavorable statements. This explanation highlights the limitations that are inherent to the study of self-reported beliefs and belief change in general (Shah et al., 2016; Yu & Chen, 2015), and emphasizes an important goal for future work in this area: to develop alternative measures of beliefs that obviate self-reporting on bounded scales.

Conclusion

In summary, a robust observation of the last decade is that the most cognitively sophisticated opposing partisans tend to disagree most strongly over a variety of "factual" political questions. The causes of this phenomenon are not well understood. An influential hypothesis is that cognitive sophistication magnifies politically biased processing of new information, driving apart the beliefs of sophisticated opposing partisans. We investigated this hypothesis by giving U.S. partisans new information about various factual political questions, measuring their prior and posterior beliefs and analytic thinking, and comparing their posterior beliefs to a normative, politically unbiased (Bayesian) benchmark. Our findings offer minimal support for the aforementioned hypothesis of cognitive sophistication magnifying political bias. Rather, greater analytic thinking was associated with posterior beliefs closer to the benchmark. More work is needed to understand why the most cognitively sophisticated opposing partisans often disagree most strongly over various factual political questions.

References

Anglin, S. M. (2019). Do beliefs yield to evidence? Examining belief perseverance vs. change in response to congruent empirical findings. *Journal of Experimental Social Psychology*, *82*, 176–199. https://doi.org/10.1016/j.jesp.2019.02.004

Baron, J., & Jost, J. T. (2019). False Equivalence: Are Liberals and Conservatives in the United States Equally Biased? *Perspectives on Psychological Science*, *14*(2), 292–303. https://doi.org/10.1177/1745691618788876

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bartels, L. M. (2002). Beyond the Running Tally: Partisan Bias in Political Perceptions. *Political Behavior*, *24*(2), 117–150. https://doi.org/10.1023/A:1021226224601

Blacksmith, N., Yang, Y., Behrend, T. S., & Ruark, G. A. (2019). Assessing the validity of inferences from scores on the cognitive reflection test. *Journal of Behavioral Decision Making*, 1–14. https://doi.org/10.1002/bdm.2133

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389–414. https://doi.org/10.1037/a0026450

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389–411. https://doi.org/10.1037/met0000159

Bullock, J. G. (2009). Partisan Bias and the Bayesian Ideal in the Study of Public Opinion. *The Journal of Politics*, *71*(3), 1109–1124. https://doi.org/10.1017/S0022381609090914

Cheng, I.-H., & Hsiaw, A. (2019). *Trust in Signals and the Origins of Disagreement* (SSRN Scholarly Paper ID 2864563). Social Science Research Network. https://papers.ssrn.com/abstract=2864563

Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from Mechanical Turk valid for research on political ideology? *Research & Politics*, *2*(4), 2053168015622072. https://doi.org/10.1177/2053168015622072

Coppock, Leeper, & Mullinix. (2018). Generalizability of heterogeneous treatment effect estimates across samples. *Proceedings of the National Academy of Sciences*, *115*(49), 12441–12446. https://doi.org/10.1073/pnas.1808083115

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

Druckman, J. N., & McGrath, M. C. (2019). The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change*, *9*(2), 111. https://doi.org/10.1038/s41558-018-0360-1

Drummond, C., & Fischhoff, B. (2017). Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. *Proceedings of the National Academy of Sciences*, *114*(36), 9587–9592. https://doi.org/10.1073/pnas.1704882114

Dunn, A., & Oliphont, J. B. (2018). Americans divided on security of economic system. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2018/09/28/americans-are-divided-on-security-of-u-s-economic-system/

Eil, D., & Rao, J. M. (2011). The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself. *American Economic Journal: Microeconomics*, *3*(2), 114–138. https://doi.org/10.1257/mic.3.2.114

Evans, J. St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*(3), 295–306. https://doi.org/10.3758/BF03196976

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Friedman, J. (2012). Motivated Skepticism or Inevitable Conviction? Dogmatism and the Study of Politics. *Critical Review*, *24*(2), 131–155. https://doi.org/10.1080/08913811.2012.719663

Hahn, U., & Harris, A. J. L. (2014). Chapter Two - What Does It Mean to be Biased: Motivated Reasoning and Rationality. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 61, pp. 41–102). Academic Press. https://doi.org/10.1016/B978-0-12-800283-4.00002-2

Hamilton, L. C., Hartter, J., & Saito, K. (2015). Trust in Scientists on Climate Change and Vaccines. *SAGE Open*, *5*(3), 2158244015602752. https://doi.org/10.1177/2158244015602752

Hamilton, L. C., & Saito, K. (2015). A four-party view of US environmental concern. *Environmental Politics*, *24*(2), 212–227. https://doi.org/10.1080/09644016.2014.976485

Hill, S. J. (2017). Learning Together Slowly: Bayesian Learning about Political Facts. *The Journal of Politics*, *79*(4), 1403–1418. https://doi.org/10.1086/692739

Hornikx, J., Harris, A. J. L., & Boekema, J. (2018). How many laypeople holding a popular opinion are needed to counter an expert opinion? *Thinking & Reasoning*, *24*(1), 117–128. https://doi.org/10.1080/13546783.2017.1378721

Joslyn, M. R., & Haider-Markel, D. P. (2014). Who Knows Best? Education, Partisanship, and Contested Facts. *Politics & Policy*, *42*(6), 919–947. https://doi.org/10.1111/polp.12098

Joslyn, M. R., & Sylvester, S. M. (2019). The Determinants and Consequences of Accurate Beliefs About Childhood Vaccinations. *American Politics Research*, *47*(3), 628–649. https://doi.org/10.1177/1532673X17745342

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, *8*(4), 18.

Kahan, D. M. (2015). Climate-Science Communication and the Measurement Problem. *Political Psychology*, *36*(S1), 1–43. https://doi.org/10.1111/pops.12244

Kahan, D. M. (2016). The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning Is and How to Measure It. In *Emerging Trends in the Social and Behavioral Sciences* (pp. 1–16). https://doi.org/10.1002/9781118900772.etrds0417

Kahan, D. M., & Corbin, J. C. (2016). A note on the perverse effects of actively open-minded thinking on climate-change polarization. *Research & Politics*, *3*(4), 2053168016676705. https://doi.org/10.1177/2053168016676705

Kahan, D. M., Landrum, A., Carpenter, K., Helft, L., & Jamieson, K. H. (2017). Science Curiosity and Political Information Processing. *Political Psychology*, *38*(S1), 179–199. https://doi.org/10.1111/pops.12396

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, *1*(1), 54–86. https://doi.org/10.1017/bpp.2016.2

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, *2*(10), 732–735. https://doi.org/10.1038/nclimate1547

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, *107*(4), 852–884. https://doi.org/10.1037/0033-295X.107.4.852

Koehler, J. J. (1993). The Influence of Prior Beliefs on Scientific Judgments of Evidence Quality. *Organizational Behavior and Human Decision Processes*, *56*(1), 28–55. https://doi.org/10.1006/obhd.1993.1044

Kunda, Z. (1987). Motivated Inference: Self-Serving Generation and Evaluation of Causal Theories. *Journal of Personality and Social Psychology*, *53*(4), 636–647. https://doi.org/10.1037%2F0022-3514.53.4.636

Kuru, O., Pasek, J., & Traugott, M. W. (2017). Motivated Reasoning in the Perceived Credibility of Public Opinion Polls. *Public Opinion Quarterly*, *81*(2), 422–446. https://doi.org/10.1093/poq/nfx018

LeBoeuf, R. A., & Shafir, E. (2003). Deep thoughts and shallow frames: On the susceptibility to framing effects. *Journal of Behavioral Decision Making*, *16*(2), 77–92. https://doi.org/10.1002/bdm.433

Leeper, T. J., Arnold, J., & Arel-Bundock, V. (2018). *Marginal Effects for Model Objects (margins)*. https://cran.r-project.org/web/packages/margins/margins.pdf

Lind, T., Erlandsson, A., Västfjäll, D., & Tinghög, G. (2018). Motivated reasoning when assessing the effects of refugee intake. *Behavioural Public Policy*, 1–24. https://doi.org/10.1017/bpp.2018.41

Malka, A., Krosnick, J. A., & Langer, G. (2009). The Association of Knowledge with Concern About Global Warming: Trusted Information Sources Shape Public Thinking. *Risk Analysis*, *29*(5), 633–647. https://doi.org/10.1111/j.1539-6924.2009.01220.x

Markovits, H., & Nantel, G. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory & Cognition*, *17*(1), 11–17. https://doi.org/10.3758/BF03199552

McCright, A. M., & Dunlap, R. E. (2011). The Politicization of Climate Change and Polarization in the American Public's Views of Global Warming, 2001–2010. *The Sociological Quarterly*, *52*(2), 155–194. https://doi.org/10.1111/j.1533-8525.2011.01198.x

Mercier, H. (2012). The social functions of explicit coherence evaluation. *Mind & Society*, *11*(1), 81–92. https://doi.org/10.1007/s11299-011-0095-4

Mercier, H. (2017). How Gullible are We? A Review of the Evidence from Psychology and Social Science. *Review of General Psychology*, *21*(2), 103–122. https://doi.org/10.1037/gpr0000111

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The Generalizability of Survey Experiments*. *Journal of Experimental Political Science*, *2*(2), 109–138. https://doi.org/10.1017/XPS.2015.19

Nurse, M. S., & Grant, W. J. (2019). I'll See It When I Believe It: Motivated Numeracy in Perceptions of Climate Change Risk. *Environmental Communication*, 1–18. https://doi.org/10.1080/17524032.2019.1618364

Nyhan, B., & Reifler, J. (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2016). Is the cognitive reflection test a measure of both reflection and intuition? *Behavior Research Methods*, *48*(1), 341–348. https://doi.org/10.3758/s13428-015-0576-1

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). Everyday Consequences of Analytic Thinking. *Current Directions in Psychological Science*, *24*(6), 425–432. https://doi.org/10.1177/0963721415604610

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

Peterson, C. R., & Miller, A. J. (1965). Sensitivity of subjective probability revision. *Journal of Experimental Psychology*, *70*(1), 117–121. https://doi.org/10.1037/h0022023

Peterson, C. R., Schneider, R. J., & Miller, A. J. (1965). Sample size and the revision of subjective probabilities. *Journal of Experimental Psychology*, *69*(5), 522–527. https://doi.org/10.1037/h0021720

Pew Research Center. (2019, January 24). *Public's 2019 Priorities: Economy, Health Care, Education and Security All Near Top of List | Pew Research Center*. https://www.people-press.org/2019/01/24/publics-2019-priorities-economy-health-care-education-and-security-all-near-top-of-list/

Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, *72*(3), 346–354. https://doi.org/10.1037/h0023653

Shah, P., Harris, A. J. L., Bird, G., Catmur, C., & Hahn, U. (2016). A pessimistic view of optimistic belief updating. *Cognitive Psychology*, *90*, 71–127. https://doi.org/10.1016/j.cogpsych.2016.05.004

Shenhav, A., Rand, D. G., & Greene, J. D. (2012). Divine intuition: Cognitive style influences belief in God. *Journal of Experimental Psychology: General*, *141*(3), 423–428. https://doi.org/10.1037/a0025391

Stagnaro, M. N., Pennycook, G., & Rand, D. G. (2018). Performance on the Cognitive Reflection Test is stable across time. *Judgment and Decision Making*, *13*(3), 260–267.

Sumner, C., Scofield, J. E., Buchanan, E. M., Evans, M.-R., & Shearing, M. (2018). *The Role of Personality, Authoritarianism and Cognition in the United Kingdom's 2016 Referendum on European Union Membership*. https://doi.org/10.31219/osf.io/n5r67

Sunstein, C. R., Bobadilla-Suarez, S., Lazzaro, S. C., & Sharot, T. (2016). How People Update Beliefs about Climate Change: Good News and Bad News. *Cornell Law Review*, *6*, 1431–1444.

Taber, C. S., Cann, D., & Kucsova, S. (2009). The Motivated Processing of Political Arguments. *Political Behavior*, *31*(2), 137–155. https://doi.org/10.1007/s11109-008-9075-8

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

Tappin, B. M., & Gadsby, S. (2019). Biased belief in the Bayesian brain: A deeper look at the evidence. *Consciousness and Cognition*, *68*, 107–114. https://doi.org/10.1016/j.concog.2019.01.006

Tappin, B. M., Pennycook, G., & Rand, D. (2018). Rethinking the link between cognitive sophistication and politically motivated reasoning. *PsyArXiv*, 1–54. https://doi.org/10.31234/osf.io/yuzfj

Tappin, B. M., Pennycook, G., & Rand, D. G. (2020). Thinking clearly about causal inferences of politically motivated reasoning: Why paradigmatic study designs often undermine causal inference. *Current Opinion in Behavioral Sciences*, *34*, 81–87. https://doi.org/10.1016/j.cobeha.2020.01.003

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99–113.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275. https://doi.org/10.3758/s13421-011-0104-1

van der Linden, S., Leiserowitz, A., & Maibach, E. (2018). Scientific agreement can neutralize politicization of facts. *Nature Human Behaviour*, *2*(1), 2. https://doi.org/10.1038/s41562-017-0259-2

Williams, D. (2018). *Hierarchical Bayesian models of delusion*. https://doi.org/10.17863/CAM.23224

Yu, R., & Chen, L. (2015). The need to control for regression to the mean in social psychology studies. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01574