

EMPIRICAL ARTICLE

Overcoming ideology-consistent biases: does it help to make things easier?

Philip U. Gustafsson¹, Torun Lindholm¹, Freja Isohanni¹, Ola Svenson^{1,2}, and Sophia Appelbom³

¹Department of Psychology, Stockholm University, Stockholm, Sweden; ²Decision Research at Oregon Research Institute, Eugene, OR, USA and ³Health Informatics Centre, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

Corresponding author: Philip U. Gustafsson; Email: philipgustafssonresearch@gmail.com

Received: 28 June 2024; **Revised:** 12 November 2024; **Accepted:** 14 November 2024

Keywords: analytic thinking; ideological bias; motivated reasoning; numeracy

Abstract

In 2 experiments, we attempted to reduce belief-consistent biases in interpretations of a polarized problem by making information easier to interpret. In the experiments, participants solved numerical problems that were either framed in a politically polarized (the effects of Muslim prayer rooms on support for Islamic extremism) or a neutral setting (the effects of a skin cream on skin rash). In both studies, the problems were presented twice, with the second presentation accompanied with an aid to facilitate problem-solving. In Experiment 1, this aid came in the form of an informative text on how to calculate the numbers to solve the problem. In Experiment 2, the aid provided participants with the first calculus necessary to solve the problem: transforming frequencies to percentages. Overall, results demonstrated belief-consistent responses in the polarized scenario when participants attempted to solve the first problem (higher accuracy when the correct conclusion was in line with participants' ideology). Information on how to calculate the problem (Experiment 1) only slightly reduced the biased responses, whereas the added percentages (Experiment 2) led to a substantial reduction of the bias. Thus, we demonstrate that the facilitation of complex information on a polarized topic reduces biases in favor of rational reasoning.

People sometimes hold beliefs that contradict the best available evidence. While some of these beliefs are trivial, they can also concern serious issues. For example, disbelieving the efficacy of COVID-19 vaccines, or the threat (and cause) of climate change, may not only impact one's own life but also the lives of others. One explanation for the rejection of well-based facts ('knowledge resistance') is *motivated reasoning*. A vast body of research demonstrates that people often selectively seek out, attend to, and accept information that is consistent with their beliefs, and conversely, downplay, distort, and discredit information that contrasts these beliefs (e.g., Baker et al., 2020; Epsley & Gilovich, 2016; Hameleers & van der Meer, 2020; Kunda, 1999; Lord et al., 1979; Taber & Lodge, 2006). However, people do not conclude whatever they want to in a given situation regardless of the actual information available. Indeed, the ability to correctly perceive and interpret the world from observation is what allows us to learn and acquire knowledge, and in the extreme case, to survive. Our motivation to view the world in ways that confirm our beliefs thus competes with a motivation to be accurate. A critical question for anyone attempting to communicate important information is to understand what factors may affect people's tendency to use any of these 2 strategies.

Carefully scrutinizing information can be costly in terms of time and effort, and the importance of understanding a situation correctly can be expected to be balanced against the effort needed to do so (e.g., Kunda, 1990). When it comes to interpretations of complex information, such as estimating distributions from ranges (Dieckmann et al., 2017), or drawing conclusions about the relation between 2 variables from a 2×2 table (Kahan et al., 2017), a strong desire to be accurate should be required to motivate the effort to understand properly. If such motivation is lacking, people may turn to their general understanding of the world, and if the target information concerns a topic they have previous beliefs about, they will use these beliefs as heuristics for understanding. However, if the conclusion from a set of information is more directly evident, requiring little, or no effort to understand, then one should expect people's interpretations to be less influenced by their previous beliefs even when the need to be accurate is low. Hence, a possible strategy to reduce people's tendency for biased understanding of new information when accuracy motivation is low could be to try to reduce the complexity of the target information. In 2 experiments, we evaluate this hypothesis by examining if we can reduce people's biased interpretations of politically polarized numerical data by reducing information complexity. Specifically, we reduce the number of calculations needed to arrive at the correct interpretation for a set of numerical data. If motivated biases can be reduced—or even eliminated—by facilitating interpretations of information, efforts to disambiguate publicly conveyed information could be an important step toward curing knowledge resistance.

1. Biased conclusions from numerical information

In everyday life, drawing conclusions about facts frequently involves assessing numerical data, as evidence often manifests in this format (e.g., the degrees to which climate on earth is expected to rise due to global warming, or the likelihood of being contaminated with COVID-19). Numerical data are sometimes complex (e.g., the basic reproductive number [R0] in a pandemic), requiring not only calculations in several steps but also an understanding of what specific calculations would be needed to understand the data. This complexity opens up for reliance on other available sources, such as one's own beliefs about the dangers of the COVID-19 virus.

People's ability to draw conclusions from complex numerical information based on political affiliation was examined by Kahan et al. (2017). In this study, participants were presented with a numerical problem regarding the relation between 2 variables in the form of a 2×2 contingency table (see versions used in our experiments in Figures 1a–d). The value-neutral form of the problem concerned the effects of a new unknown skin cream and displayed frequencies of people whose skin had improved and frequencies of people whose skin had gotten worse after having used the cream. As a control, the frequencies of people who got better and worse after *not* having used the cream were also displayed. Participants then had to conclude whether the rash in the group using the skin cream improved or worsened skin compared to the group that did not use the cream.

Other participants were presented with a scenario involving an intervention that was *politically polarized*; whether crimes in cities that had banned concealed carry of handguns had increased or decreased, compared to cities without such a ban. The results showed that participants were less accurate in their conclusions in the polarized scenario compared with the neutral scenario. Importantly, however, in the polarized scenario, participants performed better when the correct conclusion was in line with their political beliefs than when it countered these beliefs. To specify, conservatives (generally pro-gun) were more accurate in their conclusions when gun control led to more crime, while liberals (generally anti-gun) showed more accurate conclusions when gun control reduced crime. Follow-up studies have replicated the main effect of lower accuracy in polarized compared to neutral scenarios (Baker et al., 2020; Connor et al., 2024; Lind et al., 2022; Persson et al., 2021; see also Dieckmann et al., 2017; Pennycook & Rand, 2019). However, the increased bias among numerically proficient participants has not been replicated (Baker et al., 2020; Lind et al., 2022; but see Strömbäck et al., 2022). Thus, it appears that numeracy facilitates correct conclusions from numerical problems, but may not affect belief bias.

	Rash got better	Rash got worse
Number of patients who <u>did use</u> the new skin cream	181	61
Number of patients who <u>did not use</u> the new skin cream	87	17

A) Solution: Rash got worse

	Rash got worse	Rash got better
Number of patients who <u>did use</u> the new skin cream	181	61
Number of patients who <u>did not use</u> the new skin cream	87	17

B) Solution: Rash got better

	Support for Islamic extremism has decreased	Support for Islamic extremism has increased
Number of German towns that adopted <u>more generous rules</u> for allowing Muslim prayer rooms	181	61
Number of German towns that <u>did not change the rules</u> for allowing Muslim prayer rooms	87	17

C) Solution: Extremism increased

	Support for Islamic extremism has increased	Support for Islamic extremism has decreased
Number of German towns that adopted <u>more generous rules</u> for allowing Muslim prayer rooms	181	61
Number of German towns that <u>did not change the rules</u> for allowing Muslim prayer rooms	87	17

D) Solution: Extremism decreased

Figure 1a-d. Versions of the numerical problem at ‘Time 1’ in Experiment 1.

Note: These are the problems used in our experiment. In the Kahan et al. (2017) study, the polarized scenario concerned ‘gun control’ rather than ‘Muslim prayer rooms’.

2. Reducing belief bias by facilitating complex information processing

An important aspect of the problem presented by Kahan et al. (2017) and others is that it is difficult to solve correctly, with participants often performing close to the chance level. Specifically, the correct solution requires several steps: (1) understanding that ratios are needed for the solution and that these are made most accessible through proportions expressed as percentages, (2) correctly calculating the proportion for each cell, that is, dividing the value in a cell with the total amount in that row (not column), and (3) comparing the size of the proportions (column-wise, not row-wise) to find the cell with the largest/smallest proportion. Thus, the calculations involve several steps and a basic understanding of what numbers to use in each step. Given that participants in psychology experiments presumably lack strong incentives to be accurate, it can be assumed that participants often skip these and use their prior beliefs about the target topic to arrive at an answer (Pennycook & Rand, 2019). A question that follows from previous findings, then, is whether biases could be reduced if the interpretation of information is facilitated. Would people then put their previous beliefs aside, and simply let the data lead their conclusions? Some support for this idea comes from a study by Dieckmann et al. (2017). These authors presented participants with forecasts involving uncertainty. For example, the participants were presented with a forecast on how a law allowing citizens to carry concealed guns would decrease the number of sexual assaults, displayed as a range on the expected decrease of assaults from 500 to 9000. Participants then indicated which values within this range they thought were most likely, by choosing between distributions that were either normal, uniform, left/or right-skewed. The correct estimate varied, but in most situations, the distributions were roughly normal or uniform. A person who is unfamiliar with likelihood distributions might either conclude that each value is equally likely or, by motivated reasoning, conclude a skewed distribution that is in line with one’s own opinion (e.g., a higher expected decrease in assaults by someone with pro-gun beliefs and vice versa for someone with less pro-gun attitudes). The results largely supported the latter outcome. Hence, people who were pro-gun were more prone to perceive a normal- or a right-skewed distribution (higher values more likely) as more

likely than a uniform, or left-skewed distribution (lower values more likely). Conversely, participants with less pro-gun attitudes instead typically indicated a left-skewed distribution as more likely. A follow-up study included a graphics condition where a visual aid indicating a normal distribution was added, clarifying that values in the middle of the range were more likely than values at the ends. While belief-consistent biases remained in the condition without clarification, it was substantially reduced in the condition with the clarifying graphic. Thus, even if people's interpretations of data may often be biased by previous beliefs and motivations, they seem more willing to accept the correct interpretation when it is made salient.

Other results suggest that attempts to make information easier to interpret may not help. Baker et al. (2020) tried to reduce belief bias by simplifying numerical problems like those used by Kahan et al. (2017). Using similar contingency tables as Kahan et al. (2017), on the relation between 2 variables on 1 neutral (using skin cream and decrease or increase of rash) and 5 different polarized problems (e.g., guns and criminality, human-caused climate change), they randomized participants to either easy, intermediate, or difficult versions of the 2×2 tables (the difficult versions being on par with the version in Kahan et al., 2017). The difficulty was varied by changing relations between the numbers in the table, where the easy versions presented a clear decrease or increase in the outcome variable (e.g., easy version: A: 20 vs. 10, B: 15 vs. 15; difficult version; A: 25 vs. 7, B: 32 vs. 18). Although participants were better at solving the easy tasks, the belief bias was only slightly attenuated, resulting in preserved belief-consistent responses. However, a possible limitation of this attempt to facilitate interpretations of the problem is that only the difference between numbers in the 2×2 table was manipulated. It is possible that participants presented with the easy problems still had problems with transforming the numbers in the 2×2 table to proportions (see Figures 1a-d). That is, even though the difference between the numbers had been amplified, participants still had to understand that the numbers should be transformed to proportions, select the right digits to transform, and calculate the proportions correctly (see above). Hence, the task was still fairly complex which might have led participants to forgo any analytic processing in favor of a heuristic, belief-aligned response.

To evaluate how a reduction in information complexity affects reasoning, the current study examines 2 alternative ways to facilitate the interpretation of difficult numerical information. In Experiment 1, we provide participants with information on how to do the arithmetic needed to understand the data. Specifically, we show participants how to use the numbers in the 2×2 table to calculate the ratios necessary to arrive at the correct conclusion. In Experiment 2, we facilitate the problem further by adding percentages to the raw frequencies in each cell in the table. In this study then, the first 2 steps to arrive at the correct answer, choosing the right calculation and the right numbers to compare, are already conducted. Both means should facilitate interpretations by reducing the steps needed to solve the problem, and as a consequence reduce the use of heuristics biasing responses.

3. Experiment 1

In the first experiment, we studied the role of participants' ideology in solving a [politically polarized or neutral] numerical problem inspired by Kahan et al. (2017). We used Right-Wing Authoritarianism (RWA) as our measure of political ideology. This scale has been conceptualized as measuring 3 distinct but related ideological attitude constructs; authoritarianism, conservatism, and traditionalism, and is generally agreed to explain individual differences in social, collective, and intergroup behavior (Duckitt et al., 2010, 2018; Zakrisson, 2005). Given the current study's focus on potential collective threats from a religious outgroup, we saw this scale as the best proxy for the relevant ideological attitudes in our study. Furthermore, in a 2-step procedure, we examined the extent to which belief-consistent responses could be reduced by facilitating participants' interpretations of the problem. In addition to examining the accuracy of the solutions, we investigated participants' confidence in their responses (see Supplemental Materials). Participants' numeric ability was used as a control variable.

The setup of the experiment was similar to that of Kahan et al. (2017), with a numerical problem with a politically polarized or a neutral topic presented in a 2×2 table (Figures 1a-d; see detailed description

under *Methods: General procedure*). The neutral scenario was identical to that used by Kahan et al. (2017), and concerned whether using a new skin cream decreased or increased skin rash. The polarized scenario was constructed to be politically relevant to a European sample and concerned the effects of regulations for allowing Muslim prayer rooms on support for Islamic extremism in German towns (see Figures 1c–d). The topic is regularly debated in European countries (e.g., The Times, 2023; The Independent, 2016), and opinions generally divide along the right- and left-wing dimension, with anti-Muslim sentiments being stronger among right- than left-wing supporters.

Participants were first randomly presented with one of the 2×2 tables shown in Figure 1 (including counterbalanced versions). In this first presentation ('Time 1'), the 2×2 table was shown as in previous studies, with only the frequencies in each cell. Participants were then presented with the identical problem again. This second time ('Time 2'), additional information was provided on how to calculate the numbers in the table to arrive at the correct conclusion (see detailed description under *Methods: General procedure*).

We expected a main effect of numeracy (H1), such that people with a high numeric ability would be better overall at solving the problem. Second, we expected several interactions between ideology (i.e., RWA), scenario, outcome, and presentation time on conclusion accuracy (see preregistration). Here, 'outcome' refers to whether a given table showed an increase or a decrease in the target variable (e.g., 'the rash got better/got worse with the new skin cream'), and 'time' refers to participants' performance at the first and at second exposure of the table (i.e., before or after receiving our information how to calculate to arrive at the correct solution). To facilitate reading, we present results for the 2 most important hypotheses (for the remaining tests, see [Supplemental Materials](#)), namely; (H2) a 3-way interaction between RWA, scenario, and outcome on conclusion accuracy, such that participants would be better at solving the polarized problem in the belief-consistent compared to belief-inconsistent conditions, whereas performance in the neutral problem would be unaffected by ideology; and (H3) a 4-way interaction between RWA, scenario, outcome and time on conclusion accuracy, wherein the above-mentioned 3-way interaction was expected at the first presentation of the problem, but that it would be attenuated when participants were informed on how to calculate the problem in the second presentation. That is, at Time 2 (after being provided with calculation help), we did not expect participants to perform better in the belief-consistent condition compared to the belief-inconsistent condition. We expected no corresponding interactions in the neutral-scenario conditions.

3.1. Method

3.1.1. Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All materials and data, including code to analyses, have been made publicly available at the Open Science Framework (osf.io/9532s). The study's design and its analyses were preregistered (osf.io/9532s), but note that we have deviated from some of the hypotheses and analysis choices. This includes a reversed H2, an updated power analysis, and a different dichotomization of RWA.

3.1.2. Participants

We initially recruited 1006 participants from the online crowdsourcing platform Prolific (<https://www.prolific.co/>). Of these, 32 were excluded due to failing the attention check. The final sample consisted of 974 participants ($M_{\text{age}} = 37.52$, $SD = 12.46$, range = 18–76 years), of which 481 identified as female, 486 as male, and 7 as 'other'. All participants were UK residents and non-Muslim. Approximately 13% had a high school education or less, 38% attended vocational school, some university, or A-level education, 34% held a bachelor's degree, and 13% held more advanced degrees. Participation was compensated with £1.50.

The sample size was determined to be in accordance with previous studies using this experimental design (Kahan et al., 2017: $n = 1111$; Lind et al., 2022: $n = 1015$). The main analysis of interest was the difference in performance between high and low-RWA participants in the polarized scenarios with

outcomes that [dis]aligned with the RWA ideology. Previous studies found differences in performance between belief-consistent and belief-inconsistent conditions around 40%, with higher performance for participants in ideology-congruent conditions (e.g., Lind et al., 2022; belief-consistent group: 69.3% accurate, inconsistent group: 50.5% accurate). For our power analysis, we utilized the `power.prop.test` function in the `stats` package in R (R Core Team, 2016). An analysis with $n = 250^1$, group 1 proportion = .70 and group 2 proportion = .50, an alpha level of .05 and 2-sided test suggested a power of 99.61%. Note that the power analysis assumes no exclusions and that the calculated power is for a main effect rather than interactions.

The study was conducted in full in accordance with the ethical principles outlined on <https://www.codex.vr.se/>, and with the 1964 Helsinki declaration and its later amendments.

3.1.3. General procedure

Participants completed a survey, created in Qualtrics (<https://qualtrics.com>). In the survey, participants first answered demographic information and responded to a short version of the RWA scale (Zakrisson, 2005). They also answered questions about political orientation (left/right) and political partisanship (Labour/Conservative). Next, participants were randomly assigned to 1 of 4 versions of the numerical problem ('Time 1'; see Figures 1a–d), detailed below. Participants' numeric ability was then assessed, and they were randomly presented again with the numerical problem ('Time 2')—this time with instruction of the steps needed to calculate the correct answer:

Below, you find a description of how to combine the numbers in order to draw a correct conclusion about the relation:

The total number of towns that has adopted more generous regulations for allowing Muslim prayer rooms are $181 + 61 = 242$. Support for extremism has increased in 181 of the 242 towns, $181/242 = 75\%$, while the support has decreased in 61 of the 242 towns, $61/242 = 25\%$.

The total number of towns with unchanged rules for allowing Muslim prayer rooms are $87 + 17 = 104$. Support for extremism has increased in 87 of the 104 towns, $87/104 = 84\%$, while support has decreased in 17 of the 104 towns, $17/14 = 16\%$.

Both after the first and the second presentation of the problem, participants stated their conclusion and also provided confidence in these conclusions. Finally, they answered an open-ended question asking them to justify their decision and were thanked for participating.

3.1.4. Measurements

3.1.4.1. Ideology

RWA was measured using a RWA short-scale (Zakrisson, 2005), containing 15 items rated on a scale from 1 (*very negative*) to 7 (*very positive*), such as 'God's laws about abortion, pornography and marriage must be strictly followed before it is too late, violations must be punished' and 'Facts show that we have to be harder against crime and sexual immorality, in order to uphold law and order' ($M = 3.33$, $SD = 0.92$, $n_{>1SD} = 296$, $n_{<1SD} = 350$). We then dichotomized this scale by selecting people 1 SD above or below the mean ($n_{included} = 323$). Additionally, participants rated the extent to which they (dis)agreed with the statement 'Islam is generally a threat to the British way of life' (1—Strongly disagree to 7—Strongly agree).

3.1.4.2. Numeracy

Numeric ability and cognitive reflection abilities were assessed with 3 items from the Berlin Numeracy Test (Schwartz et al., 1997, further developed by Cokely et al., 2012) and 3 items from the Cognitive Reflection Test (Frederick, 2005). In order to create a 'numeracy' index, we added the scores from the 2 tests. Thus, higher values indicated higher numeric/cognitive reflection ability (Lind et al., 2022).

¹ $n = 250$ for scenario: polarized, and outcome: increase[decrease], given total sample size $n = 1000$.

3.1.4.3. Numerical problem

The central part of this study was the numerical problem used to assess people's ability to draw conclusions from data. The numerical problem was adapted from Kahan et al. (2017) and involved results from a [bogus] study with data presented as frequencies in a 2×2 contingency table (see Figure 1). There were 4 versions of this table; a scenario that was either *politically polarized* in nature (the effect of regulations for Muslim prayer rooms on support for Islamic extremism), or *politically neutral* (the effect of a new skin cream to treat skin rashes; henceforth referred to as '*neutral*'). The outcome of the problem (i.e., correct conclusion) was either an *increase* (increased support for Islamic extremism/the rash got worse), or a *decrease* (decreased support for Islamic extremism/the rash got better; see Figure 1). The numerical problem was prefaced with a text that described the target scenario. In the polarized scenario, the text read that scientists had examined if more generous rules for allowing Muslim prayer rooms in German towns led to an increase or decrease in support for Islamic extremism. In the neutral scenario, the text read that scientists had examined if the use of a new skin cream led to increased or decreased amount of skin rashes. Participants' task was to conclude, based on the data presented in the table, whether the study showed an *increase* or *decrease* in the outcome variable (i.e., rash/extremism). Responses to these numerical problems ('conclusion accuracy') were classified as correct or incorrect.

The scenarios were presented twice. In the first presentation (Time 1), only the frequencies in each cell of the table were presented, as in previous studies using this paradigm (e.g., Baker et al., 2020, Kahan et al., 2017; Lind et al., 2022; Persson et al., 2021). The second time participants were shown the problem (Time 2), the tables in all scenarios included a text that explained how to calculate the numbers to arrive at the correct conclusion (see *General procedure*). The text also showed the percentages resulting from these calculations, thus providing participants with the correct percentages for the 4 cells.

3.2. Results

As all participants responded to the numerical problem twice, we analyzed data using logistic mixed-effect multilevel modeling (Wright & London, 2009), with responses nested within participants (i.e., as random effect), and all predictors entered as fixed effects. Our procedure for the analyses followed Mansour et al. (2017), wherein predictors are entered in a stepwise fashion. This was done with Rstudio (RStudio Team, 2020) in R (R Core Team, 2016), using the *lme4* (Bates et al., 2009) and *lmerTest* (Kuznetsova et al., 2017) packages. Demographic variables (age, gender, education) were controlled for and did not affect any of the results. We therefore present results without these variables (the interested reader can access these analyses in the code document referenced under 'Transparency and Openness').

3.2.1. Conclusion accuracy

We had 3 main hypotheses: (H1) Participants high in numeric ability would outperform participants low in numeracy on the numerical problem; (H2) there would be a 3-way interaction between RWA, scenario, and outcome, such that participants would be better at solving the polarized scenario in the belief-consistent condition (e.g., high RWA, polarized scenario, 'increased extremism' outcome), compared to a belief-inconsistent condition (e.g., high RWA, polarized scenario, 'decreased extremism' outcome), with no such difference in the neutral scenarios; and (H3) a 4-way interaction between RWA, scenario, outcome and time, where we expected the 3-way interaction at Time 1 (H2), but that this effect would be eliminated at Time 2, when participants were informed about how to calculate the correct response. Given the hypothesis that biases would decrease in the polarized problem, we further expected that performance would improve overall after presentation of the calculation. To test this, we compared models containing the predictors and interactions. We entered predictors in a stepwise fashion, starting with Numeracy, followed by main effects of RWA, Scenario and Outcome, with Conclusion accuracy as dependent variable. Next, we examined the 3-way interaction (RWA, Scenario, Outcome), then included Time as predictor, and finally added the 4-way interaction between RWA, Scenario, Outcome, and Time.

Results showed that the model containing numeracy (Model 2) was superior to a baseline, intercept-only model (Model 1, see Table 1), $\chi^2(1) = 143.36$, $p < .001$, $w_{aic} > .99$. In line with H1, participants high in numeracy (i.e., $>1 SD$) performed significantly better (78.6% accurate) than participants low in numeracy (i.e., $<1 SD$; 43.6% accurate, $p < .001$). Next, we added main effects of RWA, Scenario, and Outcome to the model (Model 3, see Table 1), which outperformed the previous model, $\chi^2(3) = 34.27$, $p < .001$, $w_{aic} > .99$. Numeracy, Scenario, and Outcome were significant, unique predictors in the model (see Table 1); participants performed better in the neutral (63.0% accurate) compared to polarized scenario (53.7% accurate, $p < .001$), and in conditions with the ‘decrease’ (61.6% accurate), compared to the ‘increase’ outcome (55.1% accurate, $p = .004$). We then added the 3-way interaction (Model 4, see Table 1) which was better than Model 3, $\chi^2(1) = 7.21$, $p = .007$, $w_{aic} = .93$. Numeracy, Scenario, Outcome, as well as RWA were all significant main effects (low RWA [i.e., $<1 SD$] = 68.3% accurate; High RWA [i.e., $>1 SD$] = 48.0% accurate, $p < .001$). In line with predictions, the 3-way interaction between RWA, Scenario, and Outcome also reached significance ($p = .008$). Breaking down the 3-way interaction, we found a significant 2-way interaction between RWA and Outcome in the *polarized* scenario ($p = .025$). More specifically—in line with H2—participants low in RWA performed significantly better in the condition where extremism *decreased* in towns where regulations for Muslim prayer rooms were more generous (72.7% accurate) compared to *increased* (56.8%, $p = .040$). The opposite was found for participants high in RWA, who performed better in the *increase* outcome (54.9% accurate) compared to the *decrease* outcome (38.9% accurate, $p = .069$), but this was not statistically significant. We also found an unpredicted interaction between RWA and Outcome in the *neutral* scenario ($p < .001$), such that the high-RWA group performed better in the condition where the skin cream led to a decrease in rashes (67.9% accurate) compared to an increase (36.0% accurate, $p < .001$). There was no effect for the low-RWA group (decrease = 67.0% accurate, increase = 78.4% accurate, $p = .139$).

Moving to the next step in the model comparison, we added Time as predictor (Model 5, see Table 1) to test H3. Results showed improved performance with the new model, $\chi^2(1) = 34.64$, $p < .001$, $w_{aic} > .99$. All main effects (and the 3-way interaction) were statistically significant, including Time; participants performed better at Time 2 when they had been shown how to calculate the numbers to get the correct answer (63.3% accurate), compared to Time 1 where this information was absent (53.4% accurate, $p < .001$). Finally, we added the 4-way interaction (RWA, Scenario, Outcome, Time) to the model (Model 6, see Table 1). This model did not increase explained variance compared to the previous model, $\chi^2(1) = 0.17$, $p = .684$, $w_{aic} = .29$. Indeed, examining the predictors in the model revealed that the 4-way interaction was not statistically significant ($p = .525$). As we had hypothesized a 4-way interaction (H3), we nonetheless decided to break down analyses separately for Time 1 and Time 2 (see Figure 2). First, we examined the results at Time 1, that is, the results for the first numerical problem participants solved. In line with the results of the overall analysis, there was a 3-way interaction at Time 1 between RWA, scenario, and outcome ($p = .005$). In the polarized scenario, there was a significant 2-way interaction between RWA and Outcome ($p = .027$); low-RWA participants performed better in the condition where extremism *decreased* in towns where regulations for Muslim prayer rooms were more generous (70.5% accurate) compared to the condition where extremism *increased* (52.3%, $p = .040$). High-RWA participants tended to perform better when extremism *increased* in towns where regulations for Muslim prayer rooms were more generous (53.7% accurate) rather than *decreased* (30.6% accurate, $p = .069$). As in the overall analysis, at Time 1, there was an unpredicted interaction between RWA and Outcome ($p = .003$) in the neutral scenario. Here, contrasts revealed that high-RWA participants performed better in the condition where skin rashes increased (60.7% accurate) compared to the condition where they decreased (32.6% accurate, $p = .036$), whereas there was no significant difference for the low-RWA participants (increase = 81.1%, decrease = 64.0% accurate, $p = .133$). Next, we examined the results at Time 2, where the numerical problem was presented again, this time with n how to calculate the problem and the outcome of the calculation. At Time 2 there was no 3-way interaction between RWA, Scenario, and Outcome ($p = .241$; see Figure 2). Hence, there was no interaction between RWA and Outcome in the polarized scenario ($p = .221$). Instead and unpredicted,

Table 1. Parameter estimates (and standard error) for predictors in models of conclusion accuracy in Experiment 1.

Predictor	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Fixed effects						
Intercept	0.57 (0.09)***	0.59 (0.09)***	1.24 (0.16)***	1.23 (0.16)***	0.21 (0.24)	0.33 (0.24)
Numeric ability		0.55 (0.06)***	0.54 (0.06)***	0.54 (0.06)***	0.57 (0.06)***	0.57 (0.06)***
Scenario			-0.73 (0.17)***	-0.73 (0.17)***	-0.78 (0.18)**	-0.81 (0.18)***
Outcome			-0.54 (0.17)**	-0.53 (0.17)**	-0.56 (0.18)**	-0.59 (0.18)***
RWA			-0.17 (0.10)	-0.32 (0.11)**	-0.34 (0.12)**	-0.32 (0.12)**
Scenario*Outcome*RWA				0.55 (0.21)**	0.59 (0.22)**	0.80 (0.45)
Time					0.73 (0.13)***	0.68 (0.13)***
Scenario*Outcome*RWA*Time						-0.17 (0.26)

Note: For exact p-values, see [Supplementary Table S1](#). n = 323, *p < .05, **p < .01, ***p < .001

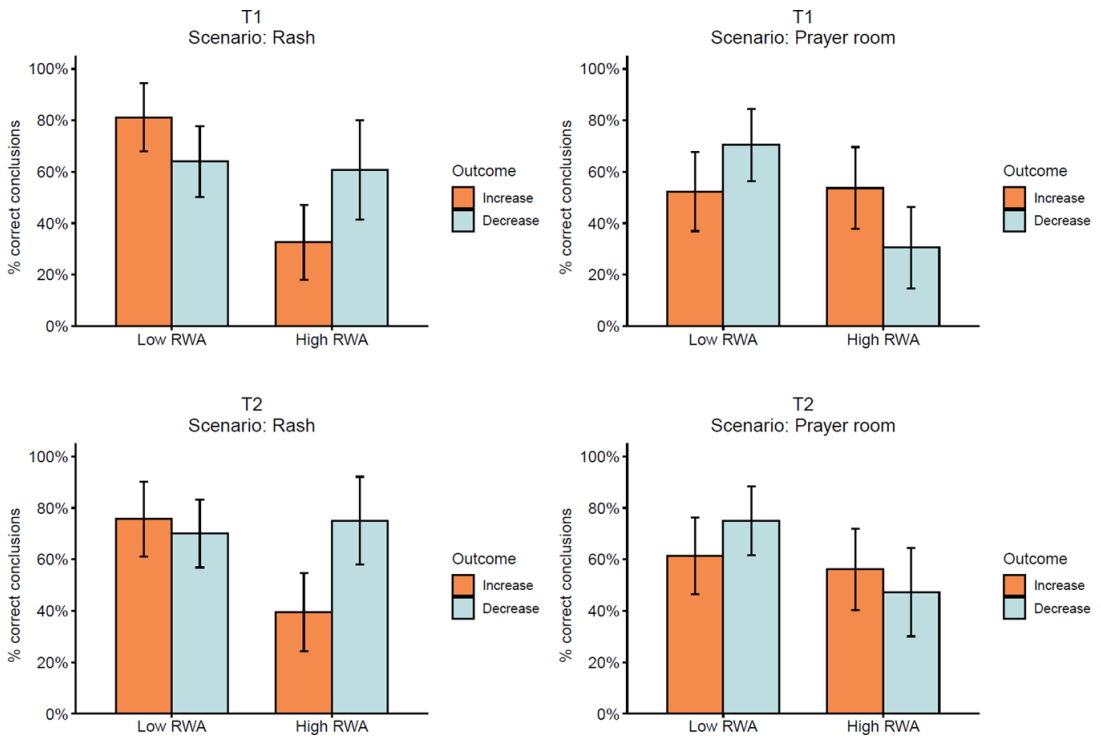


Figure 2. Percentage of accurate conclusions in Experiment 1 across conditions. Leftmost column displays results for the neutral scenario (effect of skin cream on skin rash), rightmost column shows results for the polarizing scenario (effect of prayer room on support for extremism). Top row displays results when the participants were presented with the problem for the first time ('T1'), bottom row displays the second time the problem was presented, now containing the calculations needed to reach the correct conclusion. Legend ('Outcome') displays conditions in which the correct conclusion was an increase (e.g., increased support for extremism) or decrease, respectively. High/low RWA indicates participants ± 1 SD above mean ($n_{rash} = 158$; $n_{prayer\ room} = 165$).

the RWA by Outcome interaction was significant in the neutral scenario ($p = .005$), in which high-RWA participants performed better when skin rashes decreased (75.0% accurate) compared increased (39.5% accurate, $p = .007$). There was no significant difference for the low-RWA participants (increase = 75.7% accurate, decrease = 70.0% accurate, $p = .732$).

3.2.2. Ancillary analyses

Given the varying findings in previous studies regarding the effects of numeracy in biased interpretations of numeric information (Baker et al., 2020; Kahan et al., 2017; Lind et al., 2022), we finally explored if participants' numeric ability interacted with our main variables. Thus, we checked for possible interactions with numeracy, compared to just examining its main effect in the analyses above. We created a model containing conclusion accuracy as dependent variable, and numeracy, RWA, scenario, and outcome as predictors, including all 2-way, one 3-way, and one 4-way interaction (see Table 2). Results showed that numeracy significantly interacted with scenario ($p < .001$), such that high-numeracy participants (i.e., >1 SD) performed better in the neutral (88.5% accurate) compared to the polarized scenario (68.3% accurate, $p < .001$), while low-numeracy participants (i.e., <1 SD) did not significantly differ in performance between scenarios (neutral = 40.8% accurate, polarized = 47.0% accurate, $p = .222$; see Figure 3). There were no other significant interactions involving numeracy.

Table 2. Parameter estimates (and standard error) for predictors in models of conclusion accuracy in answers in Experiment 1.

Predictor	Model 1
Fixed effects	
Intercept	1.24 (0.17)***
Numeric ability	0.62 (0.10)***
RWA	-0.17 (0.10)
Scenario	-0.80 (0.17)***
Outcome	0.52 (0.17)**
Numeric ability*RWA	-0.05 (0.06)
Numeric ability*Scenario	-0.32 (0.13)*
Numeric ability*Outcome	0.23 (0.13)
Numeric ability*Scenario*Outcome	-0.16 (0.19)
Numeric ability*RWA*Scenario*Outcome	-0.06 (0.13)

Note: For exact p-values, see [Supplementary Table S3](#). n = 323, *p < .05, **p < .01, ***p < .001

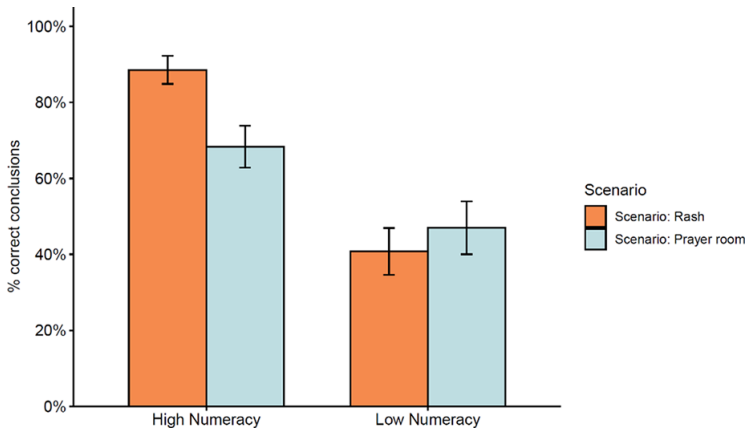


Figure 3. Conclusion accuracy in solving the numerical problem among participants with high and low numeric ability (+/-1 SD above mean, n = 508), in the polarizing and neutral scenario in Experiment 1.

Hence, the main effect of the scenario that we found in the first overall set of analysis seems to have been due to high-numeracy participants drastically dropping in performance when the scenario was polarized.

3.3. Discussion

Overall, the results supported our predictions. We replicated previous findings, showing that participants are better at solving the polarized problem when the correct outcome aligns rather than not aligns with their political ideology (see top right [Figure 2](#); Baker et al., 2020; Connor et al., 2024; Kahan et al., 2017; Lind et al., 2022; Persson et al., 2021). Moreover, also in line with predictions and previous findings, high-numeracy participants performed better overall than low-numeracy participants (see [Figure 3](#)).

Contrary to our prediction, we were unable to significantly change participants biased responses in the polarized scenario by informing them how to calculate the numbers in the table to arrive at

the correct conclusion. However, the bias did decrease slightly (see Figure 2, bottom right). We had expected that this instruction would guide participants in their interpretation of the table, making the correct answer more salient without having to expend much cognitive effort. Indeed, as noted by other scholars (e.g., Pennycook & Rand, 2019), even when people have the tools to understand complex information, they sometimes simply choose to not engage in analytic reasoning due to lack of interest or motivation (i.e., low accuracy goals). The clarification we added at Time 2—showing how to calculate the problem—required participants not only to read the entire added text but also to choose the correct numbers and the right calculation to compare them. In hindsight then, given that an erroneous answer in a research study probably has no particular detrimental consequences for a participant, it may not be surprising if many participants did not ponder too long over the instructions, but simply bypassed it by continuing to rely on their beliefs. Thus, we believe that participants would have been able to arrive at the correct solution to the numerical problem in the second presentation (Time 2) if they had thoroughly read the added text on how to use the numbers, but they apparently did not. Hence, given participants' (lack of) motivation, the added text was presumably too cumbersome to be taken into account.

Finally, exploratory analyses showed an interaction between numeracy and scenario, such that the performance of high-numeracy participants decreased significantly in the polarized compared to the neutral problem scenario (Figure 3). One possible explanation is that participants' motivation to engage in analytical processing for some reason decreased when presented with a polarized problem. That is, if facing a polarized problem activates preconceived ideas, and these ideas in turn allow for responses in line with preconceptions, then also participants with high numerical skills appear to use these as heuristics. However, we did not find any interaction between numeracy, RWA, scenario, and outcome, meaning that we could not replicate the motivated numeracy effect as found by Kahan et al. (2017).

4. Experiment 2

In Experiment 1, we hypothesized that participants' biases in the numerical problem would vanish after they had been shown how to calculate the numbers in the table to arrive at the correct answer. While performances improved, the tendency to interpret information in a belief-consistent way still largely remained (Figure 2). A possible explanation is that many participants ignored, or did not fully process the additional information, perhaps due the text being too long and complex (e.g., Pennebaker & Rand, 2018). Therefore, in Experiment 2, we designed the information so that it would be even more easily comprehended. Specifically, we put the *percentages* from the calculations, along with the frequencies, in the contingency tables at Time 2 (see Figure 4). This reduces the number of steps, hence the effort needed to draw the correct conclusions from the numbers. This should increase accuracy in responses, even when the answer conflicts with a participant's ideological belief about the topic.

	Support for Islamic extremism has <u>increased</u>	Support for Islamic extremism has <u>decreased</u>
Number of German towns that adopted <u>more generous rules</u> for allowing Muslim prayer rooms	181 (75%)	61 (25%)
Number of German towns that did <u>not change the rules</u> for allowing Muslim prayer rooms	87 (84%)	17 (16%)

Figure 4. The numerical problem presented at Time 2 in Experiment 2, with percentages.

Thus, our hypotheses were largely identical to those in Experiment 1: We expected a higher-conclusion accuracy for high-numeracy participants (H1); a 3-way interaction between RWA, scenario, and outcome (H2); a 4-way interaction between RWA, scenario, outcome, and time (H3). Furthermore, given our exploratory finding in Experiment 1, we hypothesized that numeracy and scenario would interact, such that high-numeracy participants would perform better in the neutral compared to the polarized scenario (H5), while there would be no such difference for participants low in numeric ability.

4.1. Method

4.1.1. Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All materials and data, including code to analyses, have been made publicly available at the Open Science Framework (osf.io/5q9xw). The study's design and its analyses were preregistered (osf.io/5q9xw), but note that we deviated from this initial preregistration by using an updated power analysis, and a different dichotomization of RWA.

4.1.2. Participants

One-thousand and sixty-eight participants took part in this experiment, again using Prolific (<https://www.prolific.com/>) as a recruitment platform. Of these, 64 participants were excluded due to failed attention checks and/or incomplete responses. The final sample consisted of 1004 participants ($M_{age} = 36.26$, $SD = 12.77$, range = 18–82 years), of which 504 identified as female, 495 as male, and 5 as 'other'. All participants were UK residents and non-Muslim. The distribution of education was similar to Experiment 1: 16% High school education or less/35% Vocational school, some university, or A-level education/35% Bachelor's degree/12% more advanced degrees. Participation was compensated with £1.50.

As in Experiment 1, the sample size was determined to be in accordance with previous studies using this experimental design (Kahan et al., 2017; Lind et al., 2022). Moreover, given that the results in Experiment 1 largely mirrored the effects in Kahan et al. (2017) and Lind et al. (2022), we utilized the same values to calculate power for this experiment. Thus, the `power.prop.test` function in the *stats* package in R (R Core Team, 2016) displayed 99.61% power, given $n = 250$, group 1 proportion = .70 and group 2 proportion = .50, and an alpha level of .05. Note that the power analysis assumes no exclusions and that the calculated power is for a main effect rather than interactions.

The study was conducted in full in accordance with the ethical principles outlined on <https://www.codex.vr.se/>, and with the 1964 Helsinki Declaration and its later amendments.

4.1.3. Materials and procedure

The procedure was largely identical to Experiment 1; participants filled in demographic information, the RWA scale ($M = 3.41$, $SD = 0.84$, $n_{>1SD} = 294$, $n_{<1SD} = 340$), tried to solve one of the 2 numerical problems, and filled out the numeric ability test. As in Experiment 1, we dichotomized RWA by selecting people 1 SD above or below the mean ($n_{included} = 317$). The new manipulation in Experiment 2 displayed *percentages* in addition to the original frequencies in the second presentation of the numerical problem. Figure 4 shows an example of the text for the condition where the scenario that was *polarized* (i.e., prayer room), and with *decrease* as the correct interpretation of the outcome (i.e., decreased support for Islamic extremism).

4.2. Results

As in Experiment 1, data were analyzed using logistic mixed-effect multilevel modeling, nesting responses within participants (i.e., as random effect), and all predictors entered as fixed effects. All results and statistics are presented in Table 3. Demographic variables (age, gender, education) were

Table 3. Parameter estimates (and standard error) for predictors in models of correct answers in Experiment 2.

Predictor	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Fixed effects							
Intercept	0.67 (0.08)***	0.69 (0.08)***	0.76 (0.12)***	0.79 (0.12)***	0.79 (0.12)***	-1.29 (0.23)***	-1.31 (0.24)***
Numeric ability		0.48 (0.04)***	0.47 (0.05)***	0.59 (0.06)***	0.57 (0.06)***	0.72 (0.08)***	0.72 (0.08)***
Scenario			-0.41 (0.14)**	-0.46 (0.14)***	-0.44 (0.14)**	-0.81 (0.18)***	-0.54 (0.17)**
Outcome			0.26 (0.14)	0.26 (0.14)	0.26 (0.14)	0.34 (0.17)*	0.34 (0.17)*
RWA			-0.04 (0.09)	-0.03 (0.09)	-0.19 (0.10)	-0.22 (0.13)	-0.24 (0.13)
Numeric ability*Scenario				-0.23 (0.08)**	-0.19 (0.08)*	-0.27 (0.10)***	-0.25 (0.10)**
Scenario*Outcome*RWA					0.49 (0.18)**	0.51 (0.22)*	1.62 (0.47)***
Time						1.52 (0.15)*	1.52 (0.15)***
Scenario*Outcome*RWA*Time							-0.69 (0.28)*

Note: For exact p-values, see [Supplementary Table S4](#). n = 317, *p < .05, **p < .01, ***p < .001

controlled for and did not affect any of the results. We therefore present results without these variables (the interested reader can access these analyses in the code document referenced under ‘Transparency and Openness’).

4.2.1. Conclusion accuracy

Model 2, with numeracy as predictor outperformed Model 1 containing only the intercept, $\chi^2(1) = 166.70, p < .001, w_{aic} > .99$ (high numeracy = 80.2% accurate, low numeracy = 42.8% accurate, $p < .001$; see Table 3). Similarly, Model 3 including RWA, scenario, and outcome outperformed Model 2, $\chi^2(3) = 13.19, p = .004, w_{aic} = .97$. In addition to numeracy, scenario was also significant in the model (see Table 3; neutral = 64.8% accurate, polarized = 58.2% accurate, $p = .003$). Model 4, including the interaction between numeracy and scenario, proved better fit than Model 3, $\chi^2(1) = 9.32, p = .002, w_{aic} = .97$. In line with expectations (H5), high-numeracy participants performed better in the neutral scenario (85.4% accurate) compared to the polarized scenario (75.4%, $p = .005$). Also as expected, there was no difference between scenarios for low-numeracy participants (neutral = 39.9% accurate, polarized = 45.5% accurate $p = .232$). Next, we added the 3-way interaction between RWA, scenario, and outcome (Model 5), which outperformed Model 4, $\chi^2(1) = 7.56, p = .006, w_{aic} = .94$ (Table 3). As in Experiment 1, the 3-way interaction significantly predicted conclusion accuracy. Follow-up analyses showed an interaction between RWA and Outcome in the *polarized* scenario ($p = .001$). More specifically, in line with H2, participants low in RWA performed better in the condition where extremism decreased in towns where regulations for Muslim prayer rooms were more generous (76.3% accurate) compared to the outcome where generous rules increased extremism (60.2% accurate, $p = .037$). Conversely, participants high in RWA performed better in the condition where extremism increased with more generous rules (64.1% accurate) compared to the condition in which it decreased (46.7% accurate, $p = .035$).

Also as expected, we found no interaction between RWA and Outcome in the *neutral* scenario ($p = .450$), with small differences between the groups (high RWA_{outcome: increase} = 50.0% accurate; high RWA_{outcome: decrease} = 54.2% accurate; low RWA_{outcome: increase} = 77.9%; low RWA_{outcome: decrease} = 70.00%). The next model (Model 6) including Time was superior to Model 5, $\chi^2(1) = 7.56, p < .001, w_{aic} > .99$ (Time 1 = 51.1% accurate, Time 2 = 71.9% accurate). Finally, in Model 7, we added the 4-way interaction between RWA, scenario, outcome, and time, which provided a better fit than the previous one, $\chi^2(1) = 6.23, p = .013, w_{aic} = .89$. An examination showed that the 4-way interaction was a significant predictor in the model (see Table 3). Clarifying this interaction, we analyzed contrasts by first separating results from Time 1 and Time 2. At Time 1, there was a significant interaction between RWA and Outcome in the polarized scenario ($p < .001$). In line with predictions (H3), contrasts showed a better performance for *high-RWA* participants when extremism increased rather than decreased (increase = 64.1% accurate, decrease = 31.1% accurate, $p = .005$). *Low-RWA* participants showed results in the opposite direction (decrease = 65.8% accurate, increase = 46.9% accurate, $p = .124$, see Figure 5), albeit not statistically significant. In the neutral scenario, there was no interaction between RWA and Outcome ($p = .814$; high RWA_{outcome: increase} = 37.0% accurate; high RWA_{outcome: decrease} = 41.7% accurate; low RWA_{outcome: increase} = 65.1%; low RWA_{outcome: decrease} = 57.5%). Next, we examined the results at Time 2, where participants were shown the percentages in each cell of the 2×2 table, in addition to raw frequencies (see Figure 4). Here we expected a reduced difference in conclusion accuracy between belief-consistent and belief-inconsistent conditions. Indeed, at the second problem presentation (Time 2), results showed no interaction between RWA and Outcome in the polarized scenario ($p = .215$), with small differences between the groups (high RWA_{outcome: increase} = 64.1% accurate; high RWA_{outcome: decrease} = 62.2% accurate; low RWA_{outcome: increase} = 73.5%; low RWA_{outcome: decrease} = 86.8%; see Figure 5). Similarly, there was no significant interaction between RWA and Outcome in the neutral scenario ($p = .431$), again with small differences between the groups (high RWA_{outcome: increase} = 63.0% accurate; high RWA_{outcome: decrease} = 66.7% accurate; low RWA_{outcome: increase} = 90.7%; low RWA_{outcome: decrease} = 82.5%; see Figure 5).

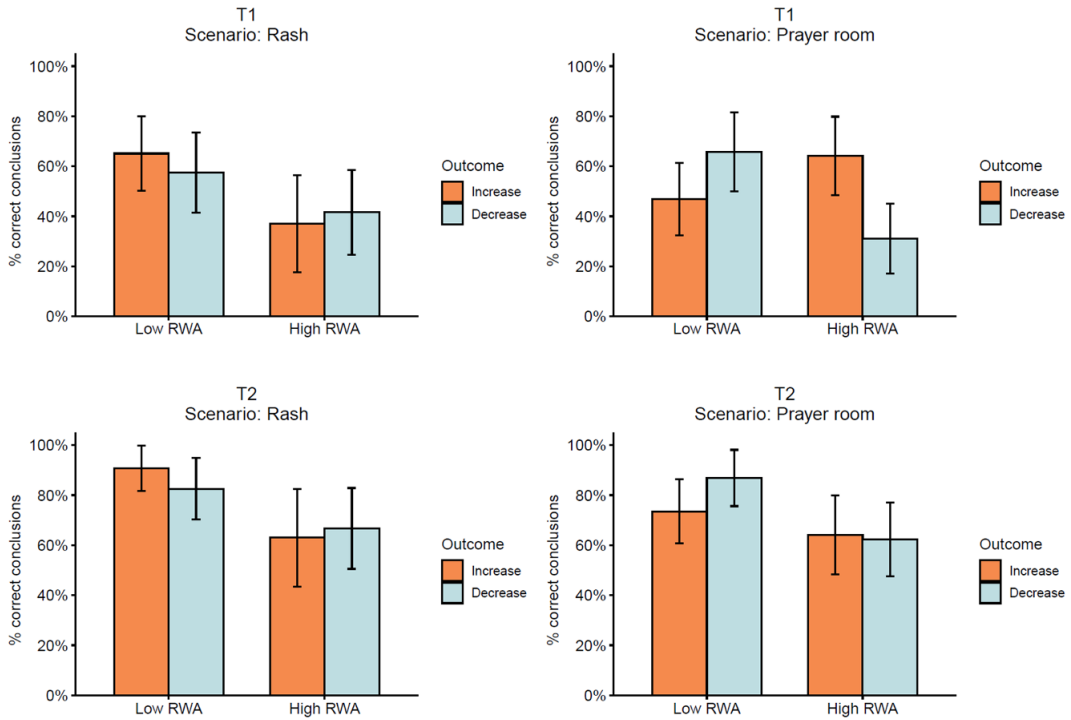


Figure 5. Percentage of accurate conclusions in Experiment 2 in the different conditions. Leftmost column displays results for the neutral scenario (effect of skin cream on skin rash), rightmost column shows results for the polarizing scenario (effect of prayer room on support for extremism). Top row displays results when the participants were presented with the problem for the first time ('T1'), bottom row displays the second time the problem was presented, now containing both frequencies and percentages. Legend ('Outcome') displays conditions in which the correct conclusion was an increase (e.g., increased support for extremism) or decrease, respectively. High/low RWA indicates participants ± 1 SD above mean ($n_{rash} = 146$; $n_{prayer\ room} = 171$).

4.3. Discussion

In this experiment, we first replicated previous studies showing improved performances for participant in belief-consistent scenario outcomes (e.g., high-RWA participant judging the Muslim prayer-room scenario with 'increased support for Islamic extremism' as outcome) compared to belief-inconsistent scenario outcomes (Baker et al., 2020; Connor et al., 2024; Kahan et al., 2017; Lind et al., 2022; Persson et al., 2021), when participants judged the problem the first time. Similar to Experiment 1, we obtained a fairly large difference in performance between consistent and inconsistent conditions (see top right cell in Figure 5). More importantly, we managed to reduce participants' belief-consistent biases when participants judged the problem the second time, where differences between aligned and unaligned conditions ranged between: 3–18% (see Figure 5, bottom right). This was done by reducing the number of steps required to reach the correct solution, that is, by showing percentages in all cells in the numerical problems in addition to the raw frequencies. Baker et al. (2020) previously attempted to reduce belief-consistent biases in a similar numerical problem by enhancing the difference between the frequencies in the cells. Although this tended to lead to more correct answers, belief-consistent biases largely remained. In our experiment, we substantially reduced biases for high-RWA participants and also reduced them for low-RWA participants (see right column in Figure 5). This supports the notion of an accuracy-effort trade-off for understanding motivated reasoning, wherein attempts to reduce the

cognitive effort needed for people to interpret a complex problem leads to more accurate and less biased responses (see e.g., Kunda, 1990). Put differently, our results suggest that people may draw their desired conclusions to the extent that evidence does not blatantly counter them, but that this biased reasoning can be abandoned when evidence for the correct conclusion is sufficiently clear.

Moreover, Experiment 2 replicated the exploratory finding in Experiment 1 regarding the interaction between numeracy and scenario on conclusion accuracy. Thus, whereas conclusion accuracy in the low-numeracy group was similar in both scenarios in Experiment 2, performance among participants with high numeric skills dropped drastically from the neutral to the polarized scenario.

5. General discussion

In 2 experiments, the current study set out to examine whether belief-consistent biases in interpretations of complex and politically polarized numerical data could be reduced if information was made easier to interpret. In Experiment 1, we examined if instructions on how to solve a difficult numerical problem increased people's ability to make accurate conclusions. To some extent it did, but participants still broadly responded in a belief-consistent manner, with better performance in conditions when the correct outcome of the problem was aligned with (as opposed to contrasting with) their political ideology (see Figure 2). In Experiment 2, we opted for another way to reduce biases, showing participants the percentages along with the frequencies in each cell of the numerical problem. With this strategy, participants' biases were substantially reduced (see Figure 5). We now discuss these findings in relation to prior studies, including limitations and potential implications.

5.1. Problem difficulty and motivated reasoning

Research shows that people tend to interpret information in ways consistent with their previous beliefs, even when information is neutral or even contrasts these beliefs (e.g., Hameleers & van der Meer, 2020; Kunda, 1999; Lord et al., 1979; Taber & Lodge, 2006). In particular, when information is not readily comprehensible, people seem inclined to rely on previous beliefs rather than spending effort in trying to decipher its meaning. However, people are also motivated to perceive the world correctly to acquire knowledge that helps them navigate in life. Hence, our tendency to interpret information in belief-consistent directions is conditional on our motivation to be accurate, but importantly, also on the extent to which evidence is ambiguous enough to give room for different interpretations. This pattern demonstrates that human thinking is flexible and that we can choose between multiple cognitive strategies based on our motivations and goals. Sometimes we prioritize speed and ease in our thinking, as illustrated by the cognitive miser metaphor described by Fiske and Taylor (1991). At other times we use thoughtful and sophisticated analyses of information. Hence, people are motivated tacticians who shift their thinking strategies to suit their goals (Crisp & Turner, 2014).

In the current studies then, we assumed that difficulty in comprehending a problem would result in motivated interpretations, but also that means to facilitate comprehension of such stimuli would decrease biased thinking. Indeed, problems that are difficult to interpret have been shown to lead to reliance on ideological beliefs (e.g., Persson et al., 2021). However, research also indicates that motivated reasoning can decrease if people's interpretations of the information are facilitated. For example, Dieckmann et al. (2017) demonstrated that a simple clarification of uncertainty range estimates, showing people a graphic with a normal distribution included, and explaining the likelihood for different outcomes, reduced participants' belief-based biases in estimates of effects regarding a politically polarized decision. We contribute to the field and expand these findings by demonstrating 2 additional ways to present information in an easily comprehensible way, which can reduce biased interpretation of polarized information. Adding to the robustness of these findings, the overall pattern largely replicated across 3 additional measures of ideology/beliefs in addition to RWA, namely participants' view on the compatibility between Islam and the British way of life (compatible/not

compatible), Ideological orientation (left/right), and Party identification (Labour/Conservative), see Figures S1, S14–17. Indeed, we believe our findings match many real-world situations, wherein odds are high that people do not ponder much upon information they encounter, but rather use their previous beliefs as heuristics to guide understanding. As shown by Baker et al. (2020), merely using easier numbers in the 2×2 table for the numerical problem did not inhibit motivated reasoning. Similarly, our instruction to the participants on how to calculate the problem in order to arrive at the correct solution did only slightly reduced bias (Experiment 1; see Figure 2). In line with Pennebaker and Rand (2018), we thus conclude that people, not the least participants in experimental studies, are often motivated to take the easiest way out when it comes to understanding information. Crucially, however, it should also be acknowledged that even in situations where incentives to thoroughly scrutinize information is low, efforts to facilitate interpretation of polarized information can prevent motivated reasoning, which is what we largely demonstrated in Experiment 2. Moreover, in line with recent studies (Baker et al., 2020; Lind et al., 2022, but see Strömbäck et al., 2022), we failed to replicate the motivated numeracy effect, wherein numerically skilled participants display a *greater* bias toward the ideology-consistent response compared to less numerically skilled participants.

5.2. *Motivated reasoning or Bayesian priors?*

As pointed out by Baker et al. (2020), the design of the numerical-problem task does not allow us to deduce whether the biased responses in the polarizing scenarios are mainly driven by a desire to confirm participants' beliefs (i.e., motivated reasoning), or due to priors pointing in a specific direction (i.e., Bayesian, rational reasoning). This study was not set out to explore this topic, hence our results cannot answer this question. However, even if participants were drawing conclusions based on (or influenced by) priors, those priors should have been formed in large part due to motivated reasoning (e.g., choosing to expose oneself to a particular news source). Thus, the biased responses will at the very least be indirectly driven by motivated reasoning, if not directly. Moreover, regarding the rational or irrational nature of the biased responses, it is important to highlight that the task is a deductive-reasoning problem; participants are asked to describe what the numbers in the table show, not describe their beliefs about the state of the world. Thus, the biased responses decisively show irrational behavior, regardless of origin. Nonetheless, a task for future research is to disentangle these potential causes of the biased responses, and further, examine if methods used to extinguish these biases differ in effectiveness depending on bias origin.

A strength of this study is that we replicated the bias-consistent responses in the polarizing scenario, using frequency numbers different from those in Kahan et al. (2017; sourced from Lind et al., 2022), thus extending the generalizability of this numerical problem. Moreover, we validate the use of a new polarized scenario for the European population, namely the effect of generous rules for Muslim prayer rooms on support for Islamic extremism (see Figure 1).

5.3. *Limitations (constraints on generality)*

One limitation of this research is that we do not know exactly how participants arrived at their answers to the numeric problem. It is conceivable though, that as long as solving the problem required more than one comparison between 2 numbers—which was the case in all problem presentations except the final one in Experiment 2—participants facing the Muslim prayer room study used their ideological preconceptions to inform them of what the study results showed. Another potential limitation is the classification of the 'skin cream' scenario as neutral. This is how the scenario was originally framed by Kahan et al. (2017), that is, presented as a neutral counterpart to the politically polarizing scenario (gun control in their study, Muslim prayer rooms in our study). Later studies have continued to use the skin-cream scenario as control to a polarizing scenario (Baker et al., 2020; Connor et al., 2024; Lind et al., 2022; Persson et al., 2021). However, in Experiment 1 we found that high-RWA participants were better

at solving this problem when the correct answer is that the skin cream leads to reduced, as opposed to increased skin rashes. A possible explanation to this result is that it may seem more likely that a cream to treat skin rashes actually succeeds in reducing rashes—or at least, that it appears very unlikely that the cream would *increase* rashes. This was not replicated in Experiment 2, but similar findings have occurred in other studies (e.g., Lind et al., 2022; Persson et al., 2021). Thus, future studies could create a new setting for the numerical problem in the neutral condition, where both outcomes are roughly equally likely, to avoid response bias contaminating the results.

Another limitation was that our experimental design lacked a ‘true’ control group wherein participants did not receive any intervention between the first and second experiment. Future studies should try to remedy this, in order to isolate the intervention effect (i.e., calculation instruction) from potential training effects. Furthermore, it would be informative to also test a group that *first* receives the calculation instruction prior to attempting to answer the numerical problem, in order to rule out effects whereby participants choose to keep the same answer over time because they feel a need to be consistent.

5.4. Implications

We believe that our findings could be practically useful, not the least in the realm of scientific communication of information that is politically polarized. To increase the effectiveness of disseminating such information to the public, communicators need to ensure that it is presented in an easily comprehensible manner, which seems to reduce the room for biased interpretations even among individuals with strong countering beliefs. While we acknowledge that information clarification may not overrule biases in every situation—such as when the source of information is immediately rejected as unreliable—we contend that these results constitute an important addition to the limited sets of strategies aimed at combating misinformation (for an overview of strategies see Lewandowsky et al., 2012).

6. Conclusion

When people are confronted with information that is complex and difficult to process, they tend to rely on their previous beliefs. We show one way to make numerical information easier to interpret, which subsequently reduces biases in favor of rational reasoning.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/jdm.2024.44>.

Data availability statement. All materials and data, including code to analyses, have been made publicly available at the Open Science Framework. Experiment 1: osf.io/9532s, Experiment 2: osf.io/5q9xw.

Competing interests. All authors declare that they have no conflicts of interest.

References

- Baker, S. G., Patel, N., Von Gunten, C., Valentine, K. D., & Scherer, L. D. (2020). Interpreting politically-charged numerical information: The influence of numeracy and problem difficulty on response accuracy. *Judgment and Decision making*, 15, 203–213. <https://doi.org/10.1017/S193029750000735X>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2009). Fitting linear mixed effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bizumic, B., & Duckitt, J. (2018). Investigating right wing authoritarianism with a very short authoritarianism scale. *Journal of Social and Political Psychology*, 6, 129–150. <https://doi.org/10.5964/jpspp.v6i1.835>
- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making*, 7, 25–47. <https://doi.org/10.1017/S1930297500001819>
- Connor, P., Sullivan, E., Alfano, M., & Tintarev, N. (2024). Motivated numeracy and active reasoning in a Western European sample. *Behavioural Public Policy*, 8, 24–46. <https://doi.org/10.1017/bpp.2020.32>
- Crisp, R. J., & Turner, R. N. (2014). *Essential social psychology* (3rd ed.). New York: SAGE Publications.

- Dieckmann, N. F., Gregory, R., Peters, E., & Hartman, R. (2017). Seeing what you want to see: How imprecise uncertainty ranges enhance motivated reasoning. *Risk Analysis*, 37, 471–486. <https://doi.org/10.1111/risa.12639>
- Duckitt, J., Bizumic, B., Krauss, S., & Heled, E. (2010). A tripartite approach to right-wing authoritarianism: The authoritarianism-conservatism-traditionalism model. *Political Psychology*, 31, 685–715. <https://doi.org/10.1111/j.1467-9221.2010.00781.x>
- Eplsey, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic perspectives*, 30, 133–140. <https://doi.org/10.1257/jep.30.3.133>
- Fiske, S. T. & Taylor, S. E. (1991). *Social cognition* (2nd ed.). McGraw-Hill.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19, 25–42. <https://doi.org/10.1257/089533005775196732>
- Hameleers, M., & Van der Meer, T. G. (2020). Misinformation and polarization in a high-choice media environment: How effective are political fact-checkers?. *Communication Research*, 47, 227–250. <https://doi.org/10.1177/0093650218818196>
- Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, 1, 54–86. <https://doi.org/10.1017/bpp.2016.2>
- Kunda, Z. (1990). *The case for motivated reasoning*. *Psychological bulletin*, 108(3), 480–498.
- Kunda, Z. (1999). *Social cognition: Making sense of people*. MIT Press.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B.. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82, 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13, 106–131. <https://doi.org/10.1177/1529100612451018>
- Lind, T., Erlandsson, A., Västfjäll, D., & Tinghög, G. (2022). Motivated reasoning when assessing the effects of refugee intake. *Behavioural Public Policy*, 6, 213–236. <https://doi.org/10.1017/bpp.2018.41>
- 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, 2098. <https://doi.org/10.1037/0022-3514.37.11.2098>
- Mansour, J. K., Beaudry, J. L., & Lindsay, R. C. L. (2017). Are multiple-trial experiments appropriate for eyewitness identification studies? Accuracy, choosing, and confidence across trials. *Behavior Research Methods*, 49, 2235–2254.
- 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>
- Persson, E., Andersson, D., Koppel, L., Västfjäll, D., & Tinghög, G. (2021). A preregistered replication of motivated numeracy. *Cognition*, 214, 104768. <https://doi.org/10.1016/j.cognition.2021.104768>
- R Core Team. (2016). *R: A language and environment for statistical computing, ver. 4.0.5*. [computer program]. R Foundation for Statistical Computing. <http://www.R-project.org/>.
- RStudio Team. (2020). *Rstudio: Integrated development environment for R*. RStudio, PBC. <http://www.rstudio.com>
- Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, 127, 966–972. <https://doi.org/10.7326/0003-4819-127-11-199712010-00003>
- Strömbäck, J., Wikforss, Å., Glüer, K., Lindholm, T., & Oscarsson, H. (2022). *Knowledge resistance in high-choice information environments* (p. 328). Taylor & Francis. <https://doi.org/10.4324/9781003111474>
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50, 755–769. <https://doi.org/10.1111/j.1540-5907.2006.00214.x>
- The Independent (2016). <https://www.independent.com.mt/articles/2016-01-25/local-interviews/Regulated-prayer-rooms-and-preachers-will-eliminate-possibility-of-extremism-6736152246>
- The Times (2023). <https://www.thetimes.co.uk/article/irelands-muslim-prayer-houses-are-prey-for-radicals-659gc2f2f>
- Wright, D. B., and London, K. (2009). *Modern regression techniques using R: A practical guide for students and researchers*. Sage Publication Ltd. <https://doi.org/10.4135/9780857024497>
- Zakrisson, I. (2005). Construction of a short version of the Right-Wing Authoritarianism (RWA) scale. *Personality and Individual Differences*, 39, 863–872. <https://doi.org/10.1016/j.paid.2005.02.026>