

# Interpreting politically-charged numerical information: The influence of numeracy and problem difficulty on response accuracy

S. Glenn Baker, Ph.D.\* Niraj Patel, Ph.D.† Curtis Von Gunten, Ph.D.‡ K. D. Valentine, Ph.D.§  
Laura D. Scherer, Ph.D.¶

## Abstract

People who possess greater mathematical skills (i.e., numeracy) are generally more accurate in interpreting numerical data than less numerate people. However, recent evidence has suggested that more numerate people may use their numerical skills to interpret data only if their initial interpretation conflicts with their worldview. That is, if an initial, intuitive (but incorrect) interpretation of data appears to disconfirm one's beliefs, then numerical skills are used to further process the data and reach the correct interpretation, whereas numerical skills are not used in situations where an initial incorrect interpretation of the data appears to confirm one's beliefs (i.e., motivated numeracy). In the present study, participants were presented with several data problems, some with correct answers confirming their political views and other disconfirming their views. The difficulty of these problems was manipulated to examine how numeracy would influence the rate of correct responses on easier vs. more difficult problems. Results indicated that participants were more likely to answer problems correctly if the correct answer confirmed rather than disconfirmed their political views, and this response pattern did not depend on problem difficulty or numerical skill. Although more numerate participants were more accurate overall, this was true both for problems in which the correct answer confirmed and disconfirmed participants' political views.

Keywords: numeracy, motivated numeracy, partisan politics, judgment and decision-making

## 1 Introduction

In the United States political landscape of 2020, not much is as apparent and consequential as partisan political polarization (Kaltenthaler & Miller, 2012). One longstanding explanation of entrenched political polarization is that people selectively search for, interpret, and integrate information that confirms and bolsters their political worldview and identity while also discounting information that challenges or disconfirms it — a phenomenon often called “motivated reasoning” (Dawson, Gilovich & Regen, 2002; Epley & Gilovich, 2016; Gilovich, 1991; Hastorf & Cantril, 1954; Holyoak & Simon, 1999; Klayman & Ha, 1987; Kunda, 1990; Kunda, 1987; Lord, Ross & Lepper, 1979).<sup>1</sup> Moreover, motivated reasoning has been found to influence a wide range of judgments, including stereotype activation and assessment (Kundra & Sinclair, 1999), science communication (Hart & Nisbett, 2012), critical reasoning skills

(Dawson, et al., 2002), and even time predictions (Buehler, Griffin & MacDonald, 1997). Thus, it is easy to understand why motivated reasoning has been widely implicated as the best way to understand why political partisans are often unable to come to agreement on many (or most) topics (Lebo & Cassino, 2007; Leeper & Slothuus, 2014; Redlawsk, 2002; Slothuus & De Vreese, 2010; Vallone, Ross & Lepper, 1985; Westen et al., 2006).

One important question is the extent to which motivated reasoning is affected by individual differences in cognitive ability, such mathematical ability (i.e., numeracy; Peters 2012; Stanovich & West, 2007; Toplak, West & Stanovich, 2011) or other components of intelligence. If a person possesses the cognitive proficiencies necessary to understand the evidence either in support of or against a given claim, then they might be more likely to consider and incorporate evidence that conflicts with their preexisting beliefs and attitudes. However, those same intelligent and highly numerate individuals might be among those most able to counter argue and dismiss evidence with which they disagree. Therefore, cognitive abilities could reasonably be expected to either protect against or exacerbate motivated reasoning. However, there is also considerable evidence indicating that the tendency to interpret information in favor of preexisting beliefs and attitudes is no more or less likely among individuals with greater cognitive abilities (Stanovich & West, 2000; Stanovich & West, 2007; Stanovich & West, 2008b; Toplak,

Copyright: © 2020. The authors license this article under the terms of the Creative Commons Attribution 3.0 License.

\*Mount Holyoke College, South Hadley, MA. Email: sg-baker@mtholyoke.edu.

†University of Missouri, Columbia, MO.

‡School of Medicine, Duke University, Durham, NC.

§Massachusetts General Hospital, Boston, MA.

¶University of Colorado-Denver, Denver, CO.

<sup>1</sup>Later we discuss why we put “motivated reasoning” and similar terms in quotes.

West & Stanovich, 2011). For example, college entrance exam scores, which are associated with both numeracy and intelligence test scores, have been found to be unassociated with many cognitive biases (Stanovich, West, & Toplak, 2013; West, Meserve & Stanovich, 2012; Stanovich & West, 2008a; 2008b). Thus, even the most intelligent among us may be biased by preexisting beliefs and are capable of making sub-optimal judgments that are not exclusively based upon the available evidence.

Many studies have shown that numeracy (i.e., the ability to process probability and numerical concepts; Peters, Västfjäll, Slovic, Mertz, Mazzocco & Dickert, 2006), unlike more general ability measures, is associated with optimal judgment and decision-making and the avoidance of heuristic errors. For example, research has shown that more numerate individuals perform better on metacognitive tasks, (Cokely, Ghazal & Garcia-Retamaro, 2014), comprehension of medical information (Peters, Hibbard, Slovic & Dieckmann, 2007), and are less susceptible to framing effects (Peters, Västfjäll, Slovic, Mertz, Mazzocco & Dickert, 2006). Numerical ability has also been cited as a key protective factor against committing the conjunction fallacy; more numerate individuals are less likely to display this fallacy (Liberali, Freyna, Furlan, Stein & Pardo, 2012), particularly when they engage in deliberative thinking strategies (Scherer, Yates, Baker & Valentine, 2017). The tendency for more numerate individuals to make better decisions might lead us to expect that numeracy is also associated with a reduced tendency to engage in motivated reasoning when interpreting numerical evidence and data. In particular, when confronted with evidence or data that violates their expectations or worldview, people with greater numerical ability might be expected to interpret evidence or data fairly and accurately, compared with their less numerate peers. Likewise, less numerate individuals might be especially inclined to interpret evidence and data as if it were consistent with their preexisting beliefs, because they lack the skills necessary to correctly interpret the data.

Recently, Kahan, Peters, Dawson and Slovic (2017a) reported a striking and somewhat counterintuitive example of the relationship between numeracy and motivated reasoning. These researchers found that when evidence presented in the form of numerical data tacitly threatens an individual’s political worldview, more numerate individuals appeared to display *more* bias in their responses than less numerate individuals when answering politically-charged data interpretation problems (Kahan, et al., 2017a). In this study, bias was identified as a differential response rate between participants who answered a data interpretation problem in which the correct answer confirmed their worldview and participants who answered a problem in which the correct answer disconfirmed their worldview. In other words, participants were more likely to answer correctly when their pre-existing beliefs were concordant with the correct re-

sponse than participants whose pre-existing beliefs conflicted with the correct response.

In their study, participants were presented with four numerical values arranged in  $2 \times 2$  table where the key to correctly interpreting the data involved ratio comparisons between rows; thus, requiring the use of the values in all four cells. When participants were presented with problems in an apolitical context (i.e., concerning the effectiveness of skin cream), more numerate participants were more likely to answer the problems correctly than were less numerate participants (Kahan et al., 2017a). However, when the problems presented a politicized topic (e.g., concealed carry handgun laws and their effects on crime), numeracy did not straightforwardly increase correct responding. Instead, responses of more numerate participants were simply more consistent with their prior beliefs.

This phenomenon has become known as “motivated numeracy”. A recent replication with a large sample of online participants reported similar results, further supporting the initial findings (Kahan et al., 2017).<sup>2</sup>

It is important to note that interpreting responses in the research of Kahan et al. (2017a) as a form of motivated reasoning, or more generally a form of *bias*, is debatable and has been challenged. Recent research has argued that incorporating prior beliefs (i.e., Bayesian reasoning) is, in fact, a rational thought process (Baron & Jost, 2019; Bullock, 2009; Druckman & McGrath 2019). Thus, weighing one’s previous beliefs (i.e., “strong gun laws don’t save lives”) into a data interpretation problem regarding the impact of efficacy of gun laws and arriving at a partisan-confirming conclusion does not necessarily constitute a form of bias, because it is not irrational to consider one’s prior beliefs in the face of new evidence (Bullock, 2009).

Furthermore, research has argued that people struggle to make a distinction between the implications of the information they are given, putting aside their prior beliefs, and the conclusion about what is true, all things considered (Kuhn, 1993; Kuhn, Shaw & Felton, 1997; Kuhn, 2001). Therefore, when confronted with information that disconfirms one’s political views, people may fail to interpret it correctly not because they were motivated to discount it, but because they interpret the question they were asked as referring to their current beliefs after seeing the evidence, beliefs that may not have changed much, given that they are consistent with many other prior beliefs. Although results from Kahan et al. (2017a) have been interpreted as a motivated process, the same pattern could be interpreted as a non-motivated process that occurs as a form of intuitive Bayesian updating (Baron & Jost, 2019; Bullock, 2009). Although the debate over what constitutes motivated reasoning and biased re-

<sup>2</sup>Although one recent conference presentation reported being unable to reproduce these findings (Ballarini & Sloman, 2017), Kahan and Peters (2017) argued that it had insufficient statistical power and differed from the original study in other relevant ways.

sponding is worth exploring in more detail it was not within the scope of this project as the main purpose of the present research was to extend the work of Kahan et al. (2017a). Henceforth we use the term “partisan-consistent responses” as opposed to “bias” to describe the difference in accuracy between problems in which the correct answer *agrees* with participants’ partisan political views and problems in which the correct answer *disagrees* with participants’ partisan political views.

## 2 Present Research

The present research further explored the relationship between numeracy and response accuracy in the context of the  $2 \times 2$  data interpretation problems used in prior research (Kahan et al., 2017a). As in the previous literature we examined response accuracy when participants were presented with problems in which the correct answer confirmed their political views (e.g., liberals: banning guns *decreases* crime) and problems in which the correct answer disconfirmed their political views (e.g., liberals: default welfare enrollment *increases* spending). However, we extended the previous research design by providing participants with several items regarding different politically charged topics, allowing this study to test whether the findings of Kahan et al. (2017a) extend to politicized issues besides gun control, and whether a pattern of partisan-consistent responses still emerged when participants answer multiple questions.

We also examined the importance of problem difficulty. Because the data interpretation problems utilized by Kahan et al. (2017a) were very challenging, it was assumed that most participants would rely on a “heuristic” interpretation of the data, without taking all four cells into account, as they would struggle to solve the problems correctly. Kahan et al. (2017a) designed problems such that the heuristic interpretations always produced an incorrect answer, so participants had to have the motivation and numerical skill in order to reason beyond that conclusion and reach the correct answer. Therefore, a natural limitation to that research is that the reported motivated numeracy effect has been established only under circumstances where the numerical data are challenging to interpret, and perhaps even misleading insofar as the heuristically produced answer was always incorrect. Thus,  $2 \times 2$  data interpretation problems similar to those used by Kahan et al. (2017a) were systematically varied in difficulty, in order to examine how the difficulty of data interpretation influences response accuracy, for both partisan-confirming and partisan-disconfirming problems.

We anticipated observing partisan-consistent responses, at least on difficult problems. That is, when problems were challenging, it was expected that participants would answer more questions correctly when the answer confirmed their pre-existing political beliefs relative to problems in

which the answer disconfirmed their pre-existing political beliefs. Furthermore, it was expected that the motivated numeracy effect would be observed consistent with the results of Kahan et al., (2017a). That is, more numerate participants would be more likely to provide partisan-consistent responses, but less numerate participants would not.

Additionally, if the motivated numeracy effect is contingent upon the data being sufficiently difficult to interpret (as pilot data collected prior to this study indicated<sup>3</sup>), then more numerate participants should respond in a pattern consistent with the results of Kahan et al., (2017a) only on the difficult problems. On easier problems, no such pattern would be observed for more numerate participants as the correct answer would be easily identifiable. However, if more numerate participants are systematically answering partisan-disconfirming problems incorrectly even when problems are easy to interpret, this would suggest that “motivated numeracy” is not about applying cognitive skills per se, but instead involves a more complex mental process that may involve, for example, taking prior beliefs into account, or rejecting the data presented as “fake”.

## 3 Methods

### 3.1 Participants

A total of 1202 participants recruited from Amazon’s Mechanical Turk (MTurk) completed this study in exchange for \$0.75. All demographics, including political affiliation, were recorded at the end of the study to ensure political identities were not primed before participants viewed the data interpretation problems. Participants ranged in age from 18–77 ( $M=35.07$ ). Approximately half (49.2%) of the sample was female, 83.4% reported being White/European American, 9.7% Black/African-American, and 7.5% Asian/Asian-American. A large portion (48.1%) of participants reported having a bachelor’s degree or a higher degree.

About half (51%) of participants indicated they were somewhat liberal, liberal or very liberal; 21% were moderate, and 29% were somewhat conservative, conservative, or very conservative. Political moderates were not excluded from this study because exclusion criteria on MTurk are often difficult to apply, and people on MTurk are attentive to screening and may lie in order to bypass exclusion criteria and participate in studies (Hauser & Schwarz, 2015). By including participants across the political spectrum, we eliminated the pressure to lie regarding their actual political orientation and were able to observe whether political moderates were more (or less) accurate than their partisan

<sup>3</sup>Numbers in the easy and intermediate conditions were pilot tested (both  $Ns \sim 100$ ) before conducting this study in order to verify that the base rate of correct responding was higher for the easy and intermediate versions relative to the difficult version.

	Rash got better	Rash got worse	Total
<b>Patients who did use the new skin cream</b>	20	10	30
<b>Patients who did not use the new skin cream</b>	15	15	30
<b>Total</b>	35	25	

Figure 1a. Skin Cream problem; easy version.

	Rash got better	Rash got worse
<b>Patients who did use the new skin cream</b>	25	7
<b>Patients who did not use the new skin cream</b>	32	18

Figure 1b. Skin Cream problem; difficult version.

peers. Furthermore, Mturk has been found to be a valid recruitment tool for research on political ideology (Clifford, Jewell & Waggoner, 2015).

### 3.2 Design

This study employed a mixed design with between-subjects manipulations of problem difficulty (easy, intermediate, difficult) and of problem sets (set A, set B), as well as one within-subjects factor of problem type (liberal-confirming, conservative confirming, apolitical). Difficulty was varied by manipulating the numbers presented to be easier (or harder) to mentally calculate, as well as the inclusion (or exclusion) of rows and columns displaying total values to make calculation easier still. Figures 1a and 1b display values from the easy and difficult versions of the “skin cream” problem, respectively. As it was necessary to counterbalance the political leaning of each of the six problems, two counterbalanced problem sets were created. In total, participants were presented with two items in which the correct answer confirmed a liberal worldview, two that confirmed a conservative worldview, and two problems which were apolitical. Figure 2 details problem sets A and B.

### 3.3 Procedure and Measures

Participants were instructed that their goal was to correctly interpret numerical data from hypothetical studies. Participants were first randomized to one of three difficulty conditions: easy, intermediate, or difficult. Following this, participants were randomized to one of the two problem sets (Figure 2). Both problem sets began with participants responding to the skin cream problem and the gun control problem from Kahan et al. (2017), presented individually in a randomized order. Next, participants were individually shown the remaining four problems that were developed for this study (regarding Obamacare, default welfare, human caused climate change, and migraine medication side-effects) in an order randomized for each participant.

After responding to all six problems, participants completed the 8-item Rasch Numeracy scale (Weller et al., 2013), and demographic measures including a single-item measure of political orientation (1 = very liberal, 4 = neutral/centrist, 7 = very conservative). A two-dimensional cultural worldview scale was also measured (Kahan et al., 2012; Kahan et al., 2017) but is not reported here for the sake of brevity and because analyses with the two-dimensional scale showed similar findings as the political orientation measure (See <https://osf.io/f2jqx/>).

The main dependent measure was correct vs. incorrect responding on each of the six problems. As previously described (and shown in Figures 1a and 1b), all problems presented information in a 2 × 2 contingency table. To arrive at the correct answer participants were required to examine the column and row information and compare ratios, and therefore, use of all the information provided in the table to arrive at the correct conclusion (although the numbers were simplified for the easy problems).

Specifically, participants were asked to select which of the two conclusions provided for each problem was supported by the data presented. Importantly, for overtly political questions, one conclusion supported a conservative-worldview and the other supported a liberal-worldview; whether the conservative or liberal conclusion was supported for a given problem was dependent on the set of problems (Figure 2) to which participants were randomly assigned. Additionally, difficult problems were designed to be on par with the difficulty level of the original problems of Kahan and colleagues (2017a) and did not include column or row totals. Both easy and intermediate versions of the problems included column and row totals and these problems presented numbers that were easier to mentally calculate and compare between columns and rows (e.g., “100”, “50” as opposed to “116”, “92”).

It is important to note that Kahan and colleagues (2017a) designed each problem such that participants could use two different heuristic problem solving strategies to answer the

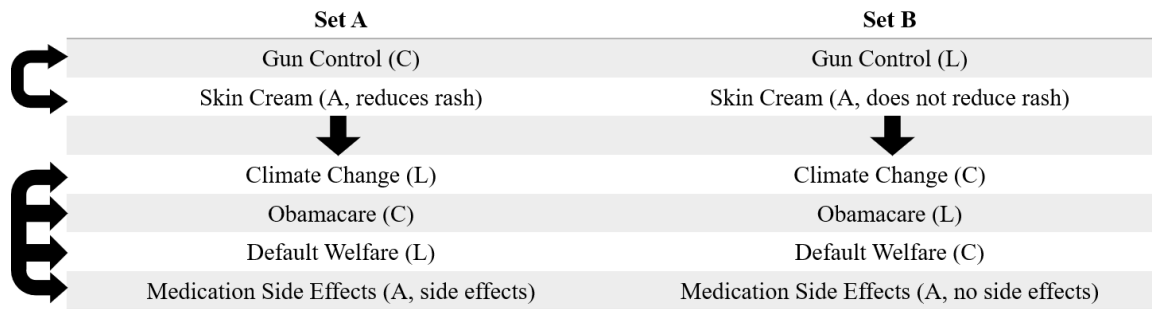


FIGURE 2: Set A and Set B full list. First, as indicated by the arrows, Gun Control and Skin Cream items were presented in a randomized order. Then, the remaining four items were presented, also in a randomized order. (L) = Liberal-confirming. (C) = Conservative-confirming. (A) = Apolitical.

problems, but these strategies always led to incorrect answers. These strategies, the “1 vs. 2” strategy (in which participants compare only the numbers adjacent to each other in one row) and the “1 vs. 3” strategy (in which participants compare only the numbers adjacent to each other in one column) do not use all the values present (Kahan et al., 2017a). This is important, as it indicates that one factor that may contribute to partisan-consistent responses is the fact that heuristic strategies may inhibit information processing. In the current study, use of the “1 vs. 2” heuristic strategy always led to the correct answer regardless of difficulty level. The “1 vs. 3” heuristic strategy led to the correct answer on all easy problems, except for the Obamacare problem, in which the strategy was inconclusive because the values compared were equivalent. However, for the intermediate problems, use of the “1 vs. 3” strategy led to three correct and three incorrect responses, and for the difficult problems this strategy led to four incorrect and two correct responses.

The inconsistency between the previous research and the present study in regards to the use of the heuristic strategies leading to incorrect responses was an oversight in developing materials that resulted in easier ratio calculations. The main implication of these differences is that reliance on heuristic strategies in the present study would, in most cases, lead participants to a correct response. Thus, incorrect responses observed in the present study may not necessarily reflect reliance on one of these heuristic strategies, especially for problems in the easy condition. For more information, see the supplemental materials (<https://osf.io/f2jqx/>).

## 4 Results

Because each participant responded to multiple problems within the dataset, a multilevel model (MLM) was used to control for correlated error (Gelman, 2006). Bates et al. (2018)’s “lme4” package in R was used for analyses. A single maximum likelihood logistic multilevel regression

model was used to test hypotheses and examine interactions between problem-type (apolitical, liberal-confirming, conservative-confirming), problem difficulty (easy, intermediate, difficult), political orientation, and numeracy. Participants were included as a random intercept factor. Political orientation and numeracy were grand mean centered before analyses to adjust for multicollinearity among interaction terms. Problem-type was dummy-coded such that the conservative-confirming problems served as the base comparison (i.e., conservative-confirming vs. liberal-confirming; conservative-confirming vs. apolitical). Difficulty levels were coded and included in the model (easy = -1, intermediate = 0, difficult = 1). These four key variables were used to predict responses on the data interpretation problems (incorrect = 0, correct = 1), with participant serving as the random variable. This model tested all main effects, 2-way, 3-way, and 4-way interactions.

If the motivated numeracy effect as reported in Kahan et al. (2017a) is contingent upon the difficulty of the problems as well as participants’ numerical ability, then 4-way interactions between all relevant variables should be observed. If numeracy (but not problem difficulty) is responsible for the production of partisan-consistent responses, then 3-way interactions between numeracy, problem-type, and political orientation should be observed. If problem difficulty (but not numeracy) influences partisan-consistent responding, then 3-way interactions between numeracy, problem-type, and political orientation should be observed. Theoretically irrelevant components from this MLM are not discussed below, though full details about this model can be found in the supplement (<https://osf.io/f2jqx/>).

First, a main effect<sup>4</sup> of difficulty was observed, ( $\gamma = -0.47$ ,  $SE = 0.07$ ,  $p < .001$ ) offering support for successful manipulation of difficulty. A main effect of numeracy was also observed, indicating that more numerate participants made fewer errors relative to less numerate participants ( $\gamma = 0.18$ ,  $SE = 0.03$ ,  $p = 0.001$ ), overall. Furthermore, there

<sup>4</sup> $\gamma$  symbols represent a slope value for all reported results within the MLM.

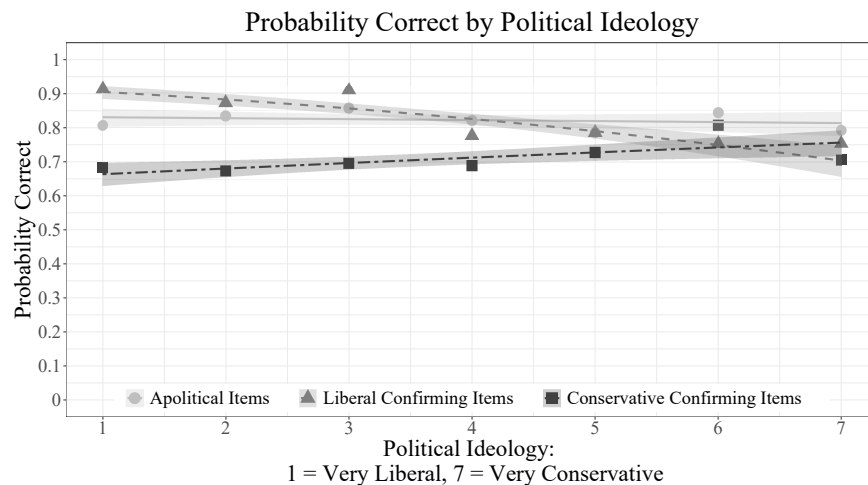


FIGURE 3: Lines indicate the probability of correct responding ( $y$ -axis) between problem-type, across political orientation ( $x$ -axis). Points indicate average probability correct for each level of political orientation from 1 to 7.

was a significant 2-way interaction between political orientation and conservative-confirming vs. liberal-confirming problems ( $\gamma=0.13$ ,  $SE=0.04$ ,  $p<.001$ ). To interpret this interaction, simplified models looking at each problem type individually were conducted. When examining apolitical problems, response accuracy was not predicted by political orientation ( $\gamma=0.13$ ,  $SE=0.04$ ,  $p<.001$ ), as anticipated. Political orientation was significant, however, when examining liberal-confirming ( $\gamma=-0.28$ ,  $SE=0.04$ ,  $p<.001$ ) items, such that liberal-leaning participants were more accurate than conservative-leaning participants. Conversely, liberal-leaning participants were less accurate relative to conservative-leaning participants on conservative-confirming ( $\gamma=-0.08$ ,  $SE=0.03$ ,  $p=.001$ ) items. Thus, the pattern in Figure 3 indicated political partisans were more likely to answer questions correctly when the correct answer agreed with their worldview than when it disagreed with their worldview.

If this pattern of partisan errors is dependent on both one’s numerical ability and problem difficulty, 4-way interactions should have been observed. However, these interactions were non-significant ( $ps > 0.31$ ). For a figure displaying the relationship between all relevant variables, see supplemental materials (<https://osf.io/f2jqx/>).

All 3-way interactions between numeracy, political orientation, and problem type were non-significant,  $\gamma<-0.01$ ,  $SEs = 0.03$ ,  $ps>0.20$ . This indicates that the pattern of partisan-consistent responses — making more errors on problems that conflict with one’s worldview than on problems that are consistent with it — did not depend on numerical ability (Figure 4). Instead, more numerate participants across the political spectrum outperformed their peers on all problem types, without a change in differential accuracy between partisan-confirming and partisan-disconfirming problems.

Finally, all 3-way interactions between difficulty, problem-type, and political orientation were not significant,  $\gamma s < -0.01$ ,  $SEs = 0.05$ ,  $ps > 0.79$  (Figure 5). This indicated that problem difficulty did not attenuate the observed pattern of partisan-consistent responses; participants were more likely to answer partisan-confirming items correctly relative to partisan-disconfirming items regardless of how difficult the problems were to solve.

## 5 Discussion

The present study extended previous research by Kahan and colleagues (2017a) in two main ways. First, participants were asked to respond to multiple data interpretation problems: the correct answer to some of these problems confirmed, disconfirmed, or unrelated to their political worldviews. Second, the difficulty level of these data interpretation problems were directly manipulated to examine if problem difficulty operates as a boundary condition for the expression of the motivated numeracy effect. The inclusion of multiple items allowed for within-subject comparisons and aggregate measures for problem-type (liberal-confirming, conservative-confirming, apolitical) that reduced the chance that outcomes could be the result of a unique scenario (e.g., banning concealed carry guns).

Results indicated that partisan-consistent responses were observed even when the problems were easy to solve. In addition, regardless of numerical skill, participants were more likely to produce the incorrect answer when the correct answer disconfirmed their worldview than when the correct answer supported their worldview. Furthermore, this differential response pattern was found at all levels of problem difficulty (i.e., how difficult the problems were to solve did not attenuate the observed pattern of partisan-consistent re-

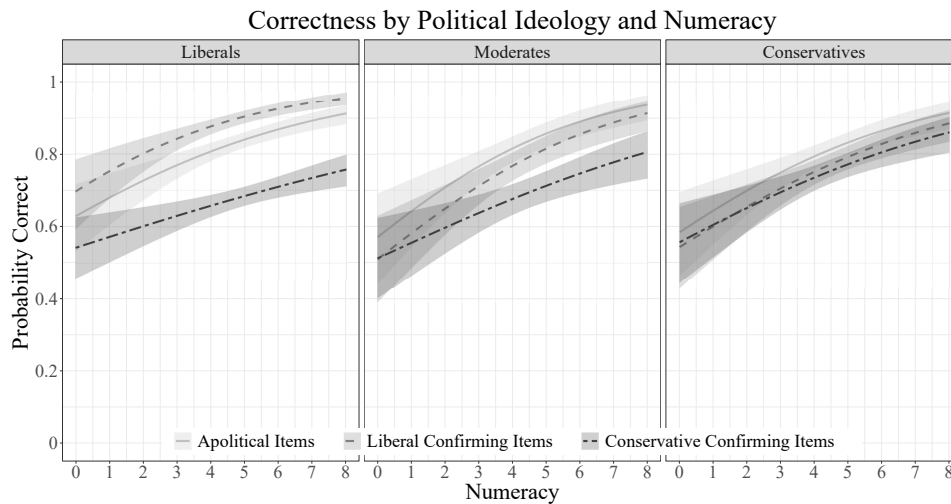


FIGURE 4: Lines indicate the probability of correct responding (y=axis) between condition and problem-type, across numeracy (x-axis).

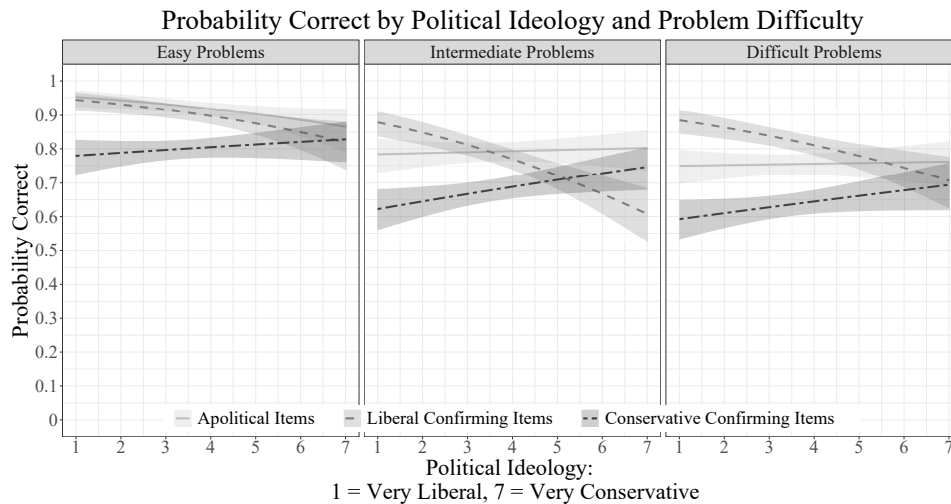


FIGURE 5: Lines indicate the probability of correct responding (y=axis) between condition and problem-type, across political orientation (x-axis).

sponses). Although participants had the opportunity to improve and/or ignore the overt political context of the problems given the within-subjects design, supplemental analyses indicated that participants became less accurate on later problems; this probable fatigue effect occurred regardless of numerical ability (see supplemental materials <https://osf.io/f2jqx/>).

Although numerical skill was associated with increased accuracy, numeracy did not moderate the expression of partisan-consistent responses. This result is consistent with previous research demonstrating the lack of an association between cognitive biases and cognitive ability (Stanovich & West, 2008a; West, Meserve & Stanovich, 2012). In this study, the correct answers were often able to be derived through heuristic responses strategies (especially for the

easiest problems). Therefore, making partisan-consistent responses does not appear to be the result of limitations in information processing. It is possible that these results are somewhat inconsistent with the findings of Kahan and colleagues (2017a) because their study deliberately intended for the problem to be highly challenging, and the heuristic responses strategies always led to incorrect responses. The original problems from Kahan et al. (2017a) may, in fact, have been so difficult that less numerate participants resorted to guessing, whereas in the present study less numerate participants may not have done so.

Although these data cannot definitively speak to the mechanism(s) responsible for partisan-consistent responses, they show that even people possessing considerable numerical ability use their prior beliefs to inform their responses

when interpreting politically-charged information. People may preferentially weight partisan-consistent information as more pertinent than relevant, but threatening, information from consideration. Alternatively, participants may reject the information at hand because it doesn't align with what they already believe to be true. It is also possible that participants who possessed the ability to identify the correct answer nonetheless deliberately provided incorrect responses in order to arrive at a preferred, partisan-consistent response. If this is the case then such responding could reflect something much more intentional than mere selective skill application suggested by previous findings especially for more numerate participants (Kahan et al., 2017a).

Furthermore, partisan-consistent responses could be the result of different processes for less numerate participants and more numerate participants. For example, less numerate participants may use partisan-consistent information when responding due to simply not understanding how to answer the question correctly. If so, then less numerate individuals may make biased judgment due to a lack of skill, whereas more numerate individuals may make errors due to actively counter arguing and/or rejecting threatening information to avoid arriving at undesired conclusions, or not wanting to give the appearance of being on the wrong side.

Notably, these findings are distinct from classical work on motivated reasoning (Dawson & Gilovich & Regen, 2002; Eply & Gilovich, 2016; Kunda, 1990), which suggests that problems that are more difficult for people to “reason their way out of” should also result in fewer partisan-consistent responses. Indeed, one could imagine that, at least for the easy problems in the present research, more numerate individuals must have found it challenging to *not* notice the objectively correct answer. Thus, these results may instead support arguments made by some researchers (Baron & Jost, 2019; Bullock, 2009) that political partisans engage in a form of Bayesian belief updating, in which they may reject information presented that disagrees with their prior beliefs. However, if such a Bayesian process (i.e., responding based on what they already think is true, regardless of the correct answer) is what participants chose to do in the present research, it would be difficult to conclude that such a process is *rational*, assuming that participants tried to follow our instructions, as participants were instructed to interpret the data at hand, not determine how true/untrue they believe the conclusion to be in the real world (Baron & Jost, 2019).

Whether or not participants are able to follow or understand such an instruction, however, has been questioned by researchers as well (Kuhn, 1997) suggesting at the very least that the role of motivation in the commission of partisan-consistent responses is unclear. One possible way future research could address whether or not partisan-consistent responses are a consequence of motivated information processing would be to ask whether people (across the spectrum of numeracy) can be trained to a) unambiguously understand

the instructions presented to them (i.e., “analyze the data in front of you and consider no other evidence”) and b) display increased accuracy over time. Based on research regarding the mental models people construct to answer syllogistic reasoning problems (Klauer, Musch & Naumer, 2000), which present challenging problems to solve that include absurd conclusions (i.e., “whales can walk”), it is possible that the partisan-consistent response pattern observed in the present research were due to participants being unable to decouple their prior beliefs of the world and the evidence presented to them. If so, then partisan-consistent responding in the present study may have more to do with learning a new skill — cognitive decoupling — and less to do with motivated reasoning. If, however, partisan-consistent responses persist even when participants have been properly instructed and trained, then they may instead be motivated: participants may selectively seek out and apply information that confirms their worldview while ignoring or discounting information that threatens it.

The first limitation of the present research concerns the distribution of political views within the sample. As is typical for data collected from MTurk, a disproportionate number of participants were liberal and liberal-leaning relative to conservative and conservative-leaning. Hence, interpretation of partisan-consistent responding across the ideological spectrum should be made with caution. Although a balanced sample is superior and should be pursued in future research, the present sample size provides a large enough number of conservative participants for an initial study. In addition, numeracy scores for participants in the present study were higher on average than those reported by Kahan and colleagues (2017a). This difference in numerical scores may reflect that many Mturk participants have relatively high levels of education, and thus high numerical ability as a population, relative to the population of subjects recruited by Kahan and colleagues (2017a).

Second, although the difficulty level of the “difficult” problems were designed to be on par with the difficulty of problems from Kahan et al. (2017a), a majority of participants (more numerate participants in particular) found our difficult problems easier than those used by Kahan et al. (2017a). This is likely related to the aforementioned fact that numerical ability was overall higher in the present study, leading to higher overall accuracy, though it also indicates that our difficult problems were not as difficult as intended. Thus, results from Kahan et al. (2017a) may occur only at even *higher* levels of difficulty, with the present study indicating a different (and somewhat surprising) pattern of results emerging when problems were easier to solve. Thus, future research utilizing such problems should closely examine the difficulty levels of different types of problems.

A third limitation with the present study involves the degree to which the data interpretation problems nudged participants into selecting a particular response by way of



heuristic strategies. Specifically, unlike the problems constructed by Kahan and colleagues (2017a) in which heuristic strategies always led to incorrect responses, in the present study the “1 vs. 2” heuristic approach always leads to the correct answer, whereas a “1 vs. 3” heuristic approach led to a mix of incorrect and correct response on intermediate and difficult versions of the problems. If it is the case that partisan-consistent responses depends on these heuristic strategies leading to an incorrect response, then it stands to reason that the deck was inadvertently stacked *against* observing a pattern of partisan-consistent responding, especially in the condition with the easiest problems. However, partisan-consistent errors were still observed in all conditions, regardless of how numerate participants were. This suggests one of three possibilities: 1) participants put very little effort into these problems resulting in their guesses being influenced by their prior beliefs, 2) participants were intentionally answering incorrectly when the correct answer disconfirmed their prior beliefs, or 3) participants do not rely on these heuristic strategies as much as suggested by prior research.

These data do not fully support the first possibility. Although analyses showed a small positive association between problem response time and problem accuracy ( $r = .06$ ,  $p < .001$ ,  $n = 7160$ ), on average participants spent one minute per problem, suggesting that most participants were not immediately guessing upon presentation of the problem (see supplemental materials <https://osf.io/f2jqx/>). Unfortunately, these data are not sufficiently rich to distinguish between the second and third possibilities, and future research should address whether partisan-consistent responding is a result of intentionally providing the incorrect answer, reliance on heuristic problem-solving strategies, or both. Further, to definitively rule out the possibility that the present results were due to the construction of the problems with respect to heuristic responses strategies, future research should ensure that all heuristic strategies that participants may employ consistently lead to the same type of conclusion, whether it be correct or incorrect.

Finally, comparisons between liberal, moderate, and conservative participants in the present study demonstrate that the pattern of partisan-consistent responses appears greater among liberal participants relative to conservative participants. One possibility is that, despite best efforts while constructing study materials, liberal participants simply found the partisan-disconfirming items more threatening than conservative participants. An alternative is that, in our sample, liberals were more strongly liberal than the conservatives were conservative. Future research should take care to ensure that politically-charged topics are equivalently threatening (and/or supportive) to participants across the political spectrum.

## 6 Conclusion

The present research found that when the correct answer to an overtly political data interpretation problem disconfirms one’s political beliefs, the probability of answering correctly is likely to decrease. This was true for all types of participants, regardless of their numerical ability or how challenging it was to interpret the data presented to them. Although being more numerate does result in an overall increase in accuracy, more numerate individuals displayed the same degree of response errors as their less numerate peers. Whether people engaged in motivated reasoning or found it challenging to decouple their prior beliefs about the world from the data they were analyzing is unclear. However, these results suggest that numerical ability may not be the most relevant factor to consider when examining when and to what extent peoples’ judgments and decisions will be influenced by partisan political views.

## References

- Andris, C., Lee, D., Hamilton, M. J., Martino, M., Gunning, C. E., & Selden, J. A. (2015). The rise of partisanship and super-cooperators in the US House of Representatives. *PLOS One*, *10*(4).
- Ballarini, C., & Sloman, S. A. (2017). Reasons and the “motivated numeracy effect.”. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, 1580- 1585.
- 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.
- Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.
- Buehler, R., Griffin, D., & MacDonald, H. (1997). The role of motivated reasoning in optimistic time predictions. *Personality and Social Psychology Bulletin*, *23*(3), 238–247.
- Bullock, J. G. (2009). Partisan bias and the Bayesian ideal in the study of public opinion. *The Journal of Politics*, *71*(3), 1109-1124.
- Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.
- Ghazal, S., Cokely, E. T., & Garcia-Retamero, R. (2014). Predicting biases in very highly educated samples: Numeracy and metacognition. *Judgment and Decision Making*. *9*(1), 15–34.
- Clifford, S., Jewell, R.M., & Waggoner, P.D. (2015). Are samples drawn from Mechanical Turk valid for research on political ideology? *Research and Politics*, *2*(4).
- Cokely, E. T., Ghazal, S., & Garcia-Retamero, R. (2014). Measuring numeracy. *Numerical Reasoning in Judgments and Decision Making about Health*, 11–38.

- Dawson, E., Gilovich, T., & Regan, D. T. (2002). Motivated reasoning and performance on the was on selection task. *Personality and Social Psychology Bulletin*, 28(10), 1379–1387.
- Druckman, J. N., & McGrath, M. C. (2019). The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change*, 9(2), 111–119.
- Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic Perspectives*, 30(3), 133–40.
- Gelman, A. (2006). Multilevel (hierarchical) modeling: what it can and cannot do. *Technometrics*, 48(3), 432–435.
- Hauser, D. J., & Schwarz, N. (2015). It’s a trap! Instructional manipulation checks prompt systematic thinking on “tricky” tasks. *Sage Open*, 5(2).
- Hart, P. S., & Nisbett, E. C. (2012). Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Communication Research*, 39(6), 701–723.
- Evans, J. S. B., & Frankish, K. E. (2009). *In two minds: Dual processes and beyond*. Oxford University Press.
- Gilovich, T. (1991). How we know what isn’t so: The fallibility of human reason in everyday life. New York: Free Press.
- Hastorf, A. H., & Cantril, H. (1954). They saw a game; a case study. *The Journal of Abnormal and Social Psychology*, 49(1), 129.
- Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General*, 128(1), 3.
- 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.
- Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, 1(1), 54–86.
- Kahan, D. M., & Peters, E. (2017). Rumors of the “non-replication” of the “motivated numeracy effect” are greatly exaggerated. Working Paper No. 324. Cultural Cognition Project, Yale University.
- Kalienthaler, K., & Miller, W. J. (2012). The polarized American: Views on humanity and the sources of hyper-partisanship. *American Behavioral Scientist*, 56(12), 1718–1734.
- Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94(2), 211.
- Klauer, K.C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, 107(4), 852.
- Kuhn, D. (1993). Connecting scientific and informal reasoning. *Merrill-Palmer Quarterly*, 74–103.
- Kuhn, D. (2001). How do people know? *Psychological Science*, 12(1), 1–8.
- Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and instruction*, 15(3), 287–315.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, 108(3), 480.
- Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology*, 53(4), 636.
- Kunda, Z., & Sinclair, L. (1999). Motivated reasoning with stereotypes: Activation, application, and inhibition. *Psychological Inquiry*, 10(1), 12–22.
- Lebo, M. J., & Cassino, D. (2007). The aggregated consequences of motivated reasoning and the dynamics of partisan presidential approval. *Political Psychology*, 28(6), 719–746.
- Lecheler, S., de Vreese, C., & Slothuus, R. (2009). Issue importance as a moderator of framing effects. *Communication Research*, 36(3), 400–425.
- Leeper, T. J., & Slothuus, R. (2014). Political parties, motivated reasoning, and public opinion formation. *Political Psychology*, 35(S1), 129–156.
- Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making*, 25(4), 361–381.
- Redlawsk, D. P. (2002). Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making. *Journal of Politics*, 64(4), 1021–1044.
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098.
- Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision-making. *Psychological Science*, 17(5), 407–413.
- Scherer, L. D., Yates, J. F., Baker, S. G., & Valentine, K. D. (2017). The influence of effortful thought and cognitive proficiencies on the conjunction fallacy: Implications for dual-process theories of reasoning and judgment. *Personality and Social Psychology Bulletin*, 43(6), 874–887.
- Slothuus, R., & De Vreese, C. H. (2010). Political parties, motivated reasoning, and issue framing effects. *The Journal of Politics*, 72(3), 630–645.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23(5), 645–665.
- Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reason-*

- ing, 13(3), 225-247.
- Stanovich, K. E., & West, R. F. (2008a). On the relative independence of thinking biases and cognitive ability. *Journal of personality and social psychology*, 94(4), 672.
- Stanovich, K. E., & West, R. F. (2008b). On the failure of cognitive ability to predict myside and one-sided thinking biases. *Thinking & Reasoning*, 14(2), 129–167.
- Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science*, 22(4), 259–264.
- 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.
- Vallone, R. P., Ross, L., & Lepper, M. R. (1985). The hostile media phenomenon: biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of Personality and Social Psychology*, 49(3), 577.
- West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of Personality and Social Psychology*, 103(3), 506.
- Westen, D., Blagov, P. S., Harenski, K., Kilts, C., & Hamann, S. (2006). Neural bases of motivated reasoning: An fMRI study of emotional constraints on partisan political judgment in the 2004 US presidential election. *Journal of Cognitive Neuroscience*, 18(11), 1947-1958.