Behavioural Public Policy (2024), 8: 1, 24–46
 doi:10.1017/bpp.2020.32
First published online 5 August 2020

Motivated numeracy and active reasoning in a Western European sample

PAUL CONNOR
University of California, Berkeley, CA, USA
EMILY SULLIVAN*
Eindhoven University of Technology, Eindhoven, The Netherlands
MARK ALFANO
Macquarie University, Macquarie Park, NSW, Australia
NAVA TINTAREV
Delft University of Technology, Delft, The Netherlands

Abstract: Recent work by Kahan *et al.* (2017) on the psychology of motivated numeracy in the context of intracultural disagreement suggests that people are less likely to employ their capabilities when the evidence runs contrary to their political ideology. This research has so far been carried out primarily in the USA regarding the liberal–conservative divide over gun control regulation. In this paper, we present the results of a modified replication that included an active reasoning intervention with Western European participants regarding both the hierarchy–egalitarianism and individualism–collectivism divides over immigration policy ($n = 746$; considerably less than the preregistration sample size). We reproduce the motivated numeracy effect, though we do not find evidence of increased polarization of high-numeracy participants.

Submitted 9 August 2019; revised 1 April 2020; accepted 16 April 2020

Introduction

People disagree about key societal issues in the face of compelling scientific evidence. Such disagreements have significant societal impacts not only with regard to decision-making (e.g., whether to vaccinate children), but also with regard to political polarization between groups. Why do seemingly intractable disagreements about policy arise? According to the 'identity-protective cognition thesis', the answer is that human reasoning is negatively affected when new information threatens one's social identity. In a previous study with

* Correspondence to: Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands. E-mail: e.e.sullivan@tue.nl

American participants, Kahan *et al.* (2017) found support for this hypothesis. When the topic about which participants were asked to exercise their reasoning skills was unrelated to their political identities (whether a skin cream cured rashes), high-numeracy liberals and conservatives both performed well. However, when the topic was related to their political identities (whether gun control is effective policy), high-numeracy liberals tended to successfully reason about accuracy only when the evidence suggested that gun control is effective, whereas high-numeracy conservatives tended to successfully reason about accuracy only when the evidence suggested that gun control is not effective. It may not be surprising that responses became politically polarized when answering questions about a gun control ban, but what was remarkable in Kahan *et al.* (2017) was that polarization was higher among high-numeracy individuals than among low-numeracy individuals. This suggests that the quantitative reasoning skills of participants with high numeracy skills can become more identity protective, which portends starker disagreement between more numerate partisans than between less numerate partisans.

In this study, we investigated whether a similar result can be found in a Western European sample of participants and for a different controversial topic (migration policies).[1] In addition, we were interested to see whether encouraging active reasoning in one of two ways might mitigate the effect. We thus examine the following two research questions:

RQ1: Do some active reasoning interventions do a better job than others at improving numeric reasoning overall?
RQ2: Can we replicate the polarizing effect of identity-protective cognition on numeracy for a different controversial topic in a different population?

Here is the plan for this paper: in the next section, we contextualize our study in the published literature on motivated numeracy and active reasoning. Then, we explain the methodology used for the current study. Following this, we lay out our results and address RQ1 and RQ2. Finally, we discuss the limitations of the current study and explore opportunities for future work on this important topic.

Related work

In this section, we summarize the extant research in the area of motivated numeracy. We also explain our use of active reasoning inductions and why we believe such inductions may help temper the ill effects of motivated

1 The preregistration for this study is available at https://osf.io/65z4h. We ended up diverging from several details of the preregistration, which we note when relevant below.

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

numeracy. To the best of our knowledge, this is the first study to investigate the effect of active reasoning interventions on motivated numeracy.

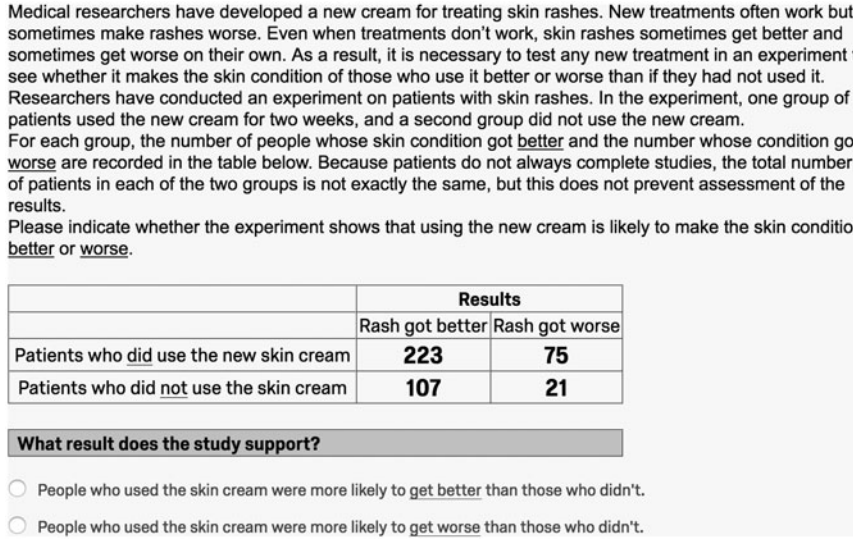
Motivated numeracy

Motivated numeracy is a species within the larger genus of motivated cognition. The overarching category includes processes and dispositions related to seeking out evidence, trusting and distrusting sources of information, interpreting evidence and counterevidence, weighting competing criteria in decision-making, remembering information, noticing inferential connections and so on. Much motivated cognition is normatively unobjectionable, or even desirable. There is nothing wrong with people seeking out information related to topics and issues they care about rather than those they do not. Additionally, if someone lacks epistemic motivation entirely, they are unlikely to engage in inquiry. However, motivated reasoning can turn vicious when it leads people to disregard or misinterpret – for identity-protective reasons – key evidence that they would otherwise be well positioned to process.

Motivated numeracy specifically concerns the way numeracy skills are affected by motivated cognition. Numeracy is a specific measure that encompasses mathematical ability and the disposition to engage in reflective quantitative reasoning (Peters *et al.*, 2006; Liberali *et al.*, 2012). Motivated numeracy crops up in those cases in which people need to exercise their learned capacity to interpret data, tables and figures. In such a context, there is typically a clear right answer dictated by the evidence. This makes the study of motivated numeracy more interpretable than the study of, for instance, risk perception. When social scientists such as Kahan *et al.* (2005) study attitudes towards new technologies like nanoparticles, it is often difficult even for experts to say exactly how the risks and benefits should be weighed against one another. If some people focus more on the risks while others focus more on the benefits, they may come to different conclusions and yet both be reasoning unobjectionably. Indeed, Alfano (2019) argues that the same person may come to opposite evaluations if they approach the evidence first skeptically, then in a trusting mode. When it comes to interpreting a graph or a contingency table, though, there is a definitive correct answer. This means that researchers can use numeracy tasks to examine not just faultless differences in risk aversion, but also outright errors in reasoning, which brings us to Kahan *et al.* (2017).

Participants in Kahan and colleagues' study were presented with a contingency table like the one pictured in Figure 1. The table represented either the results of a (fictional) pharmaceutical study or the results of a (fictional) policy on gun control. In addition, some participants saw a contingency

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press


Medical researchers have developed a new cream for treating skin rashes. New treatments often work but sometimes make rashes worse. Even when treatments don't work, skin rashes sometimes get better and sometimes get worse on their own. As a result, it is necessary to test any new treatment in an experiment to see whether it makes the skin condition of those who use it better or worse than if they had not used it.
Researchers have conducted an experiment on patients with skin rashes. In the experiment, one group of patients used the new cream for two weeks, and a second group did not use the new cream.
For each group, the number of people whose skin condition got better and the number whose condition got worse are recorded in the table below. Because patients do not always complete studies, the total number of patients in each of the two groups is not exactly the same, but this does not prevent assessment of the results.
Please indicate whether the experiment shows that using the new cream is likely to make the skin condition better or worse.

	Results	
	Rash got better	Rash got worse
Patients who did use the new skin cream	223	75
Patients who did not use the skin cream	107	21

What result does the study support?

- People who used the skin cream were more likely to get better than those who didn't.
- People who used the skin cream were more likely to get worse than those who didn't.

Figure 1. Example stimulus, representing the rash condition.

table that indicated that the skin cream (gun control policy) was effective, while others saw a table that indicated that the cream (policy) was ineffective.

As mentioned by Kahan *et al.* (2017), correctly interpreting the data was expected to be difficult. The key to interpreting a table like this is to compare not the absolute numbers, but the ratios between them. Comparing these ratios is essential to detecting covariance between the treatment and the two outcomes, a necessary element of causal inference that confounds even many intelligent people (Stanovich, 2009; Stanovich & West, 1998). For instance, the table pictured in Figure 1 indicates that 223 out of 298 patients who used the cream got better (74.8%), whereas 107 out of the 128 patients who did not use the cream got better (83.6%). Thus, even though more patients who used the cream got better, the likelihood of getting better given that one used the cream was lower than the likelihood of getting better given that one did not.

Based on previous studies using the design reflected in this experiment, it is known that most people use one of two heuristic alternatives to this approach. The first involves comparing the number of outcomes in the upper left cell to the number of outcomes in the upper right one ('1 vs 2'). The other ('1 vs 3') involves comparing the numbers in the upper left and lower left cells (Wasserman *et al.*, 1990).

Kahan *et al.* (2017) found that higher-numeracy participants – those that scored highly on the numeracy scale – were better able to interpret the

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

contingency table than those with low scores. In the skin cream conditions, participants' political partisanship had no effect on their responses. However, in the gun control conditions, partisan participants tended to answer correctly only when they saw ideologically friendly data: liberal Democrats gave the correct answer primarily when the table suggested that gun control worked, whereas conservative Republicans gave the correct answer primarily when the table suggested that gun control did not work. Moreover, polarization was more evident between high-numeracy participants than between low-numeracy participants for both liberal Democrats and conservative Republicans. Kahan and colleagues explain these results, and in particular the polarization, as stemming from identity-protective cognition. Essentially, the idea is that identity-related commitments (e.g., to minimal regulation of firearms or to strong regulation of firearms) can bump up against the facts, and that when such clashes occur, people tend to hold tight to their commitments and ignore or misinterpret the facts, even if they are ordinarily disposed to do well on quantitative tasks involving reflective thinking.

To our knowledge, there have been four central attempts to reproduce this result – some direct replications, others modified replications.[2] First, Kahan and Peters (2017) report a successful direct replication of the original study with a large ($n = 1596$), demographically diverse sample of participants, though of course replications by different labs are even more persuasive than self-replications. Second, Washburn and Skitka (2018) ($n = 1347$) replicate and extend the original result by showing that it crops up for both conservatives and liberals across a range of controversial issues, including not only gun control, but also health care reform, nuclear power and same-sex marriage. Third, Khanna and Sood (2018) conduct three studies – all using some form of firearms regulation as the controversy – that again replicate the original finding. Finally, Nurse and Grant (2019) conduct a conceptual replication with Australian participants ($n = 504$) using anthropogenic climate change rather than gun control as the controversial topic; this conceptual replication also succeeded in finding the effect of motivated numeracy.

Thus, to date, all but one of the studies of motivated numeracy have involved participants from the USA. Direct replications will presumably continue to employ American participants, since gun control is not nearly as controversial in the vast majority of other countries as it is in the USA. In addition, all four of these replication studies used a unidimensional measure of political ideology,

2 In addition to the four replication studies discussed, Ballarini and Sloman (2017) conducted a small-scale ($n = 55$) replication and extension. Though they did not find evidence of a motivated numeracy effect, the very low statistical power of this study and the fact that almost all participants were politically liberal suggest that it should not be accorded much evidential weight.

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

along the traditional left–right spectrum. While the unidimensional measure is adequate for many purposes, we suspect that it may obscure some interesting differences. For that reason, in the current study, we chose to use Kahan *et al.* (2007)'s two-dimensional measure of ideology. As we explain in more detail below, this scale employs two orthogonal dimensions: hierarchy–egalitarianism (H–E) and individualism–collectivism (I–C). The H–E subscale measures the respondent's attitude towards vertically structured hierarchies, such as are seen in the military, the church and most large corporations. The I–C subscale measures the respondent's attitude towards group solidarity. So, for example, someone who scores high on H–E but low on I–C would be supportive of a society characterized by steep hierarchy and strong communal obligations enforced by governmental regulation, whereas someone who scores high on both subscales would be supportive of a society characterized by steep hierarchy and unregulated communal obligations. Traditional left–right partisan measures tend to conflate these two dimensions.

Active reasoning

Critical thinking – and avoiding the ill effects of motivated reasoning – is a highly valued skill, but a difficult one to teach or nurture. Unfortunately, critical thinking is a skill that is often missing even among people holding a degree in a scientific field of study (Shtulman, 2013). It is difficult to undermine unfounded beliefs by simply pointing out alternative explanations. Indeed, trying to correct such beliefs might even strengthen people's initial beliefs (Nguyen *et al.*, 2007; Lewandowsky *et al.*, 2012). In particular, such backfiring is liable to occur when the argument threatens someone's identity or falls outside the boundaries of what they consider acceptable. One way to address this problem is to present information with sufficient support and guidance. Additionally, it is crucial to support critical thinking early, as it is most likely to exert an influence at the time of message exposure (Lewandowsky *et al.*, 2012).

Extant research documents encouraging evidence for various active reasoning approaches that support critical thinking. In the classroom, an effective method to foster active reasoning has been to ask students to themselves generate counterarguments for unfounded beliefs (Miller & Wozniak, 2001). Teaching such active reasoning skills and pointing out the flawed argumentation techniques used by providers of misinformation has also been shown to be effective at reducing belief in false information (Cook *et al.*, 2017). The results suggested a slight increase in item acceptance. Other work introduced a lightweight but effective protocol for supporting debate in a classroom activity with university students. The findings suggest that this intervention led to a

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

statistically significant belief change, and that this change was in the direction of the position best supported by scientific evidence. However, the intervention combined several aspects (including exposure to a lecture on critical thinking and seeing the arguments of peers), which does not allow us to draw conclusions about the effects of individual aspects (Holzer *et al.*, 2018).

Furthermore, some authors argue that online debate could reduce beliefs in pseudoscientific claims (Holzer *et al.*, 2015; Tsai *et al.*, 2015), possibly leveraging the fact that arguments from peers can be more persuasive than those coming from more authoritative figures (Garrett, 2011). In this vein, *rbutr* is a software solution that scaffolds peer debates on controversial information right where it appears.[3] It does so by allowing users to post and rate rebuttals for webpages through a browser plugin. In this way, any webpage can become a live debate platform. This is in line with a view that there should be a World Wide Argument Web, connecting arguments with each other online (for a review, see Schneider *et al.*, 2013).

In light of this previous work, we posit that a procedure that encourages active reasoning could decrease the extent to which identity-protective cognition manifests. To clarify this issue, we designed a replication study measuring identity-protective cognition with two active reasoning manipulations (one with online argumentation and the other using online search).

Experiment

This experiment is a modified replication of the study by Kahan *et al.* (2017) that includes an active reasoning intervention. While this study was preregistered on OSF,[4] two difficulties resulted in deviations from the preregistration. First, because of the funds available, we were not able to collect the full sample indicated in the preregistration ($n = 1600$). Second, there were technical issues with the *rbutr* platform, as we outline in the discussion. We are unable to assess how many participants experienced technical issues.

Participants

Participants were recruited on the Prolific platform, with a filter for participants registered as British or Dutch to ensure a European sample with high English comprehension. In total, 746 participants completed the study (61%

3 http://rbutr.com
4 https://osf.io/59uv7/?view_only=a4d7c4bc42a8475f9c40a0d24cf6631

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

female).[5] The majority (68%) were British, and a small minority (2%) were Dutch, though 28% did not specify a nationality. The mean age was 34.75 years (SD = 11.61). The majority of participants had either completed a college (30%) or a bachelor's degree (39%), but there were participants at an elementary school level (1%), high school (15%), master's (12%) and PhD/JD/MD (2%).

Stimulus

As in the original study, the stimulus consisted of four versions of a problem involving the interpretation of data and causal inference. Those results were reported in a 2×2 contingency table, the columns of which specified the number of cases that reflected positive and negative results, respectively, and the rows of which reflected the experimental treatment (see Figure 1). These were on two different topics: *Medicine* and *Policy*, and both used the same numbers as the original study.

Medicine

For the skin rash treatment topic, there were two versions of the experiment. These two versions differed only in terms of which result they supported. This meant that labels at the tops of the columns ('Rash got better' versus 'Rash got worse') in the table were reversed. The contingency table below the labels describes a number of patients suffering from skin rashes, where some have received treatment and others have not. The table indicates how many patients got better, and the participant is asked to indicate either that 'the people who used the skin cream were more likely to *get better* than those who didn't' or that 'the people who used the skin cream were more likely to *get worse* than those who didn't'. These stimuli are identical to those used in the original Kahan *et al.* (2017) study.

Policy

Two conditions of the experiment involved a new immigration policy. The contingency table describes the effectiveness of a strict new immigration policy; in one condition, the stricter policy is effective, and in the other it is not. The table indicates the number of people whose level of radicalization decreased and the

5 This was fewer than the target of 1600 participants in our preregistration. Unfortunately, we ran out of money to pay participants and so were not able to collect the full sample.

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

number of people whose level of radicalization increased. The wording was kept as comparable as possible to that of the original Kahan *et al.* study:

> Terrorism researchers have developed a new policy for identifying radicalization in recent immigrants. New policies often work but sometimes lead to additional radicalization. Even when policies don't work, radicalization sometimes decreases and sometimes increases randomly. As a result, it is necessary to test any new policy in an experiment to see whether it leads to more or less radicalization. Researchers have conducted an experiment on recent immigrants at risk of radicalization. In the experiment, one group of border security officers applied a stricter entrance policy and a second group did not apply the stricter entrance policy. For each group, the number of people whose level of radicalization decreased and the number whose level of radicalization increased are recorded in the table below. Because security officers do not always complete studies, the total number of participants in each of the two groups is not exactly the same, but this does not prevent assessment of the results. *Please indicate whether the experiment shows that using the strict new policy is likely to make radicalization decrease or increase.*

Procedure

In a between-subjects design, participants were assigned to one out of eight conditions ($2 \times 2 \times 2$ design):

- Result polarity (2): intervention caused improvement, intervention caused decline
- Topic (2): medical treatment, immigration policy
- Active reasoning (2): browser search, *rbutr*

Participants first supplied basic demographic information. Then they were asked to spend some time on actively and critically researching their topic (medical treatment or immigration policy).

Depending on the condition, participants were either asked to use the *rbutr* website or to use their preferred method for finding information online. The *rbutr* system is a website and plugin where users supply links to articles that 'rebut' or argue against the points made in other articles.

The instructions given for *rbutr* were:

> We are testing a new online tool called '*rbutr*' that aims to help people see both sides of an argument or debate. Users of *rbutr* supply links to articles about various topics, such as health, politics and religion. Users can also supply links to additional articles that 'rebut' or argue against the points made in the original articles. We would like to ask you to use *rbutr* to

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

> investigate the quality of medical research. Do modern medical treatments work? How effective are they? What strengths or flaws do they have?[6]
>
> To answer these questions, please follow this link or copy and paste the following url into your browser: http://rbutr.com/rbutr/WebsiteServlet?requestType=browse&tagId=25. You'll find links to original articles listed beneath the text "All rbutls tagged as health." Links to rebuttals appear to the right of the original articles. Please spend approximately 10 minutes using the *rbutr* tool to learn about the quality of medical research.

The active control was described in the following way:

> We are testing the idea that searching for information online helps people see both sides of an argument or debate. Searching can be done using a search engine such as Google or Yahoo, browsing trusted websites or following forum discussions.
>
> We would like to ask you to search online to investigate the quality of medical research. Do modern medical treatments work? How effective are they? What strengths or flaws do they have?[7]
>
> To answer these questions, please use your preferred method for getting information online. Please spend approximately 10 minutes searching, reading or watching videos to learn about the quality of medical research.

Both active reasoning interventions were accompanied by a 10-minute timer that prevented participants from moving to the next stage before they had done some research.

Next, participants completed a questionnaire about their political affiliation and a questionnaire assessing their numeracy skills. The experiment was concluded with a free-text comment box for remaining questions or comments from participants.

Political orientation

The Kahan *et al.* study that we are replicating used self-reports on the continuum between conservative Republican and liberal Democrat. To broaden the study to European political views, we used a questionnaire containing two validated scales to measure political affiliation (Kahan, 2012). In this questionnaire, participants indicate the level of their disagreement or agreement

6 For the policy domain, these questions are rephrased: "We would like to ask you to use *rbutr* to investigate the effects of immigration policy. Do immigrants benefit or harm the communities that they join? Do they pose safety or security risks? What immigration policies should countries adopt?"

7 These were also rephrased for the policy domain: "We would like to ask you to search online to investigate the effects of immigration policy. Do immigrants benefit or harm the communities that they join? Do they pose safety or security risks? What immigration policies should countries adopt?"

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

with each item on a Likert response measure. Responses are then aggregated (with appropriate reverse coding of the 'E' and 'C' items) to form continuous H–E (13 items) and I–C (17 items) worldview scores. Here is an example item from the I–C scale associated with high individualism: "People who are successful in business have a right to enjoy their wealth as they see fit." And here is an example item from the H–E scale associated with high hierarchy: "It seems like the criminals and welfare cheats get all the breaks, while the average citizen picks up the tab." A full list of items can be found in Kahan *et al.* (2007). This scale avoids measuring political ideology along the traditional left–right spectrum, which may obscure interesting differences between worldviews. It is also more generally applicable across cultures. For example, this scale was used successfully in both UK (Marris *et al.*, 1998) and Dutch populations (Steg and Sievers, 2000; Poortinga *et al.*, 2002), which were the target populations in our study.

Numeracy

To assess numeracy competence, participants completed the questions in a numeracy questionnaire. In the original study by Kahan *et al.* (2017), the questionnaire by Weller *et al.* (2013) was used. We also updated the two critical reflection test (CRT) questions to questions from CRT-2 (see Toplak *et al.*, 2014; Thomson & Oppenheimer, 2016) to decrease the risk of previous exposure to the questions on crowdsourcing platforms.

Questions range in difficulty to make it possible to distinguish between participants with various levels of numeracy. Following Kahan *et al.* (2017), we define high-numeracy individuals as those whose numeracy score is above the mean and low-numeracy individuals as those who scored below the mean. The mean was 5.37 correct answers. The used questionnaire can be found in Appendix 1.

Results

All analyses were conducted in R (R Core Team, 2018). Following Kahan and colleagues, primary analyses used multiple imputation to handle missingness (the maximum amount of missingness for any variable used was seven missing responses for two items within the I–C scale, less than 1% missing). Multiple imputation was performed using the *mice* R package (van Buuren & Groothuis-Oudshoorn, 2011).[8]

8 All code is available at the OSF website associated with this project: https://osf.io/59uv7/?view_only=a4d7c4bc42a8475f9c40a0d24cf66313

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

Preliminary analysis

We first investigated whether numeracy skills were different based on mean splits of political scores. Our measure of numeracy consisted of seven questions, on which participants were scored as answering correctly or incorrectly. We summed correct responses to calculate numeracy scores (M = 5.37, SD = 1.56) out of a maximum of 7. Correlations between study variables are presented in Table 1.

Welch two-sample *t*-tests indicated that numeracy scores differed between high and low scorers on H–E ($p < 0.001$, Cohen's $d = 0.28$) and high and low scorers on I–C ($p < 0.001$, Cohen's $d = 0.27$). In each case, more liberal participants (who scored below the mean on the political scales) scored higher on numeracy.

RQ1: Do some active reasoning interventions do a better job than others at improving numeric reasoning overall?

Overall, participants selected the correct interpretation of the data table only 43% of the time, which was significantly lower than chance ($t(742.9) = -3.94$, $p < 0.001$, 95% confidence interval (CI) = 0.39–0.46). However, this depended on respondents' numeracy. Participants below the mean on numeracy were correct 40% of the time (95% CI = 0.35–0.47), while participants who scored above the mean on numeracy were correct 46% of the time (95% CI = 0.41–0.51). This is similar to the result in Kahan *et al.* (2017), who found 41% correct interpretation.

To test whether the active reasoning manipulation affected the accuracy of responses, we fit a logistic regression predicting correct responses (1 = correct, 0 = incorrect) from a dummy indicating condition (1 = active reasoning manipulation, 0 = control). The active reasoning condition had no significant effect on response accuracy ($b = 0.09$, SE = 0.15, $t(741.9) = 0.62$, $p = 0.53$, $r^2_{pseudo} = 0.0003$[9] fitted to the non-missing data only (we could not find a package or information on computing pseudo-r^2 for logistic regressions fit to imputed data)). Moreover, there were also no significant two-way interaction effects between active reasoning and topic or result polarity, as well as no three-way active reasoning by topic by polarity interaction effects (all $p > 0.16$).

These results suggest that there were no significant differences between the two active reasoning interventions. However, there were some issues with the used platform (*rbutr*), which are addressed in the 'Discussion' section.

9 In our manuscript, r^2_{pseudo} refers to McFadden's pseudo *r*- for logistic regression models (McFadden, 1979).

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

Table 1. Correlations between study variables. For both Ideology measures (hierarchy–egalitarianism and individualism–collectivism), higher values indicate more conservative responses. 'Correct' is a dummy indicating correct response to the numerical reasoning problem. 'Policy' is a dummy indicating the topic of the problem (0 = medicine, 1 = policy). 'Increase' is a dummy indicating result polarity (0 = intervention decreases outcome, 1 = intervention increases outcome). '*rbutr*' is a dummy indicating active reasoning condition (1 = active reasoning, 0 = control).

		1	2	3	4	5	6
1	Numeracy						
2	Hierarchy–egalitarianism	–0.14					
3	Individualism–collectivism	–0.12	0.64				
4	Increase	–0.06	–0.03	–0.05			
5	Policy	0.02	0.05	0.04	0.004		
6	*rbutr*	–0.004	0.03	–0.004	0.02	–0.05	
7	Correct	0.09	–0.05	0.01	–0.01	–0.05	0.02

Given the similar performance across the active reasoning interventions, we collapsed across these two conditions in further analyses. We also compared whether the topic manipulation (medicine and policy) affected the accuracy of responses. The average number of correct responses was lower for the policy topic (40%) compared to the medicine topic (45%), but this difference was not statistically significant ($b = –0.21$, SE = 0.15, $t(741.9) = –1.41$, p = 0.16, $r^2_{pseudo} = 0.002$).

RQ2: Can we replicate the polarizing effect of identity-protective cognition on numeracy for a different controversial topic in a different population?

Based on the findings of Kahan *et al.* (2017), we hypothesized that individuals' political orientations would interact with topic (medicine versus policy) and result polarity (intervention leads to increase versus decrease in rashes/radicalization) in determining the probability of correct responses among individuals higher in numerical reasoning ability. This hypothesis entails a four-way interaction between political ideology, topic, polarity and respondent numeracy. Specifically, we hypothesized that liberal-leaning respondents high in numerical reasoning would be more likely to respond correctly in the policy condition when the data supported a more liberal policy stance (i.e., when the stricter entrance policy increased radicalization), while more conservative-leaning

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

respondents high in numerical reasoning would be more likely to respond correctly in the policy condition when the data supported a more conservative political stance (i.e., when the stricter entrance policy decreased radicalization). By contrast, we expected that we would not see a similar effect in the medicine condition, with the ideology of high-numeracy respondents showing little relationship with correct interpretations of the data.

To test this hypothesis, we fit two separate logistic regression models predicting correct responses from a dummy indicating the topic (0 = medicine, 1 = policy), a dummy indicating response polarity (0 = intervention decreases outcome, 1 = intervention increases outcome), respondents' numeracy scores and respondents' political ideology (one model used H–E scores for political ideology, the other model used I–C scores), as well as all interactions between these predictors. Following Kahan and colleagues, we modeled both numeracy and political ideology as *z*-scored continuous variables, and also modeled nonlinear effects of numeracy by including the squared term of numeracy in each model, as well as its interactions with the experimental conditions.

Full model results are presented in Table 2. The four-way interaction term between topic, polarity, political ideology and numeracy did not reach significance in either the full H–E model ($b = -0.07$, SE = 0.21, $t(722.2) = -0.36$, p = 0.72, $\Delta r^2_{pseudo} = 0.00003$) or the full I–C model ($b = -0.04$, SE = 0.22, $t(720.1) = -0.17$, p = 0.87, $\Delta r^2_{pseudo} = 0.00004$).[10] This was also the case even in simplified models removing numeracy squared and its higher-order interactions. Therefore, our findings are slightly different from those of Kahan and colleagues.

However, our results did provide some support for a weaker version of the identity-protective cognition hypothesis. Specifically, reduced models removing numeracy and its squared term indicated significant three-way interaction terms between topic, polarity and political ideology in both the H–E model ($b = -0.67$, SE = 0.30, $t(734.1) = -2.20$, p = 0.03, $\Delta r^2_{pseudo} = 0.005$) and the I–C model ($b = -0.66$, SE = 0.31, $t(733.1) = -2.11$, p = 0.04, $\Delta r^2_{pseudo} = 0.004$). Figure 2 displays the predicted probabilities of answering correctly for each topic and polarity type.[11]

As is shown in Figure 2, more egalitarian and communitarian respondents were generally more likely to select the correct answer, but consistent

10 Δr^2_{pseudo} indicates the change in r^2_{pseudo} between models with and without the effect in question.

11 In our preregistration, we indicated that we would produce additional visualizations based on the ones in Kahan and colleagues' paper. However, we found these overly complex and difficult for readers to interpret, so we have left them out of this paper.

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

Table 2. Multivariate regression analysis ($n = 746$). Outcome variable is 'correct', a binary variable coded 1 for correctly interpreting the data and 0 for incorrectly interpreting them. Predictor estimates are logit coefficients with *t*-statistics indicated parenthetically. 'Policy' is a dummy indicating the topic (0 = medicine, 1 = policy). 'Increase' is a dummy indicating result polarity (0 = intervention decreases outcome, 1 = intervention increases outcome). Both Ideology (hierarchy–egalitarianism (H–E) or individualism–collectivism (I–C), with higher values indicating more conservative responses) and Numeracy are *z*-scored for ease of interpretation. Bold typeface indicates a coefficient is significant at $p < 0.05$.

	Hierarchy–egalitarianism			Individualism–collectivism		
Intercept	–0.149 (–0.997)	–0.156 (–1.015)	–0.279 (–1.38)	–0.139 (–0.922)	–0.149 (–0.971)	–0.283 (–1.392)
Policy	–0.318 (–1.443)	–0.287 (–1.248)	–0.296 (–0.994)	–0.308 (–1.401)	–0.363 (–1.558)	–0.283 (–0.939)
Increase	–0.105 (–0.510)	–0.088 (–0.413)	0.095 (0.332)	–0.122 (–0.587)	–0.106 (–0.497)	0.111 (0.385)
Ideology	–0.266 (–1.903)	–0.229 (–1.594)	–0.235 (–1.629)	**–0.400 (–2.585)**	**–0.358 (–2.256)**	**–0.377 (–2.351)**
Increase × Policy	0.179 (0.594)	0.152 (0.489)	–0.035 (–0.084)	0.162 (0.536)	0.223 (0.708)	–0.054 (–0.128)
Policy × Ideology	**0.559 (2.653)**	**0.503 (2.296)**	**0.518 (2.352)**	**0.579 (2.467)**	**0.584 (2.338)**	**0.602 (2.408)**
Increase × Ideology	0.121 (0.578)	0.092 (0.428)	0.098 (0.456)	0.187 (0.900)	0.104 (0.481)	0.118 (0.540)
Increase × Policy × Ideology	**–0.672 (–2.204)**	**–0.642 (–2.030)**	**–0.669 (–2.105)**	**–0.658 (–2.109)**	–0.602 (–1.830)	–0.625 (–1.892)
Numeracy		0.108 (1.132)	0.155 (1.442)		0.113 (1.168)	0.161 (1.500)
Policy × Numeracy		–0.141 (–0.966)	–0.152 (–0.965)		–0.110 (–0.736)	–0.147 (–0.925)
Increase × Numeracy		0.131 (0.999)	0.066 (0.454)		–0.110 (–0.736)	0.063 (0.432)
Ideology × Numeracy		–0.03 (–0.333)	0.004 (0.037)		–0.054 (–0.497)	–0.028 (–0.252)
Increase × Policy × Numeracy		0.054 (0.283)	0.136 (0.638)		0.010 (0.051)	0.117 (0.545)
Policy × Ideology × Numeracy		0.067 (0.476)	0.058 (0.398)		–0.138 (–0.820)	–0.152 (–0.883)
Increase × Ideology × Numeracy		0.141 (1.020)	0.105 (0.731)		0.241 (1.626)	0.219 (1.453)
Numeracy^2			0.051 (0.946)			0.053 (1.015)
Policy × Numeracy^2			0.003 (0.036)			–0.030 (–0.384)
Increase × Numeracy^2			–0.074 (–0.973)			–0.084 (–1.126)
Increase × Policy × Numeracy^2			0.064 (0.606)			0.097 (0.930)
Increase × Policy × Ideology × Numeracy		–0.101 (–0.503)	–0.075 (–0.363)		–0.056 (–0.260)	–0.037 (–0.167)

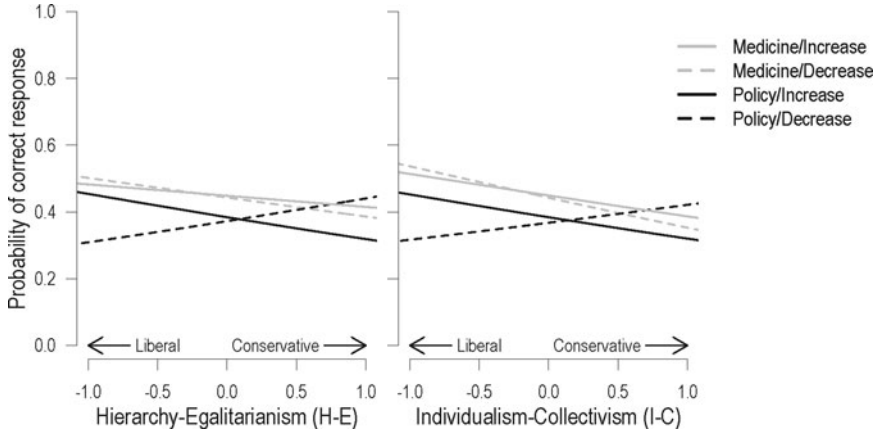


Figure 2. Predicted probabilities of answering correctly for each topic and polarity type by *z*-scored hierarchy–egalitarianism (H–E) scores (left panel) and by *z*-scored individualism–collectivism (I–C) scores (right panel).

with a motivated numeracy account, when results ran counter to an egalitarian worldview – the policy/decrease condition, in which stricter border policies led to reduced radicalization – more egalitarian and collectivist respondents became less likely to select the correct answer and more hierarchical and individualistic respondents became more likely to select the correct answer.

Power

Given our failure to fully replicate the results of Kahan *et al.* (2017) and the fact that we did not achieve our desired sample size, we assessed *post hoc* the statistical power of our study to detect the four-way interaction between ideology, polarity, topic and numeracy. No study to date has reported a standardized effect size for this four-way interaction, so we performed a power sensitivity analysis to estimate our power to detect effects of varying size. To achieve this, we simulated populations of $n = 1{,}000{,}000$ based on the distributions described in Kahan and colleagues' study, and we systematically varied the effect size of the four-way interaction within these populations. We then took 10,000 unique samples of $n = 756$ from each simulated population, refitted our full model using each sample and recorded the proportion of models returning significant four-way interaction coefficients at each effect size. The results suggested that, at our sample size, we were adequately powered to detect an effect of $\Delta r^2_{pseudo} = 0.0083$ (see Figure 3).

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

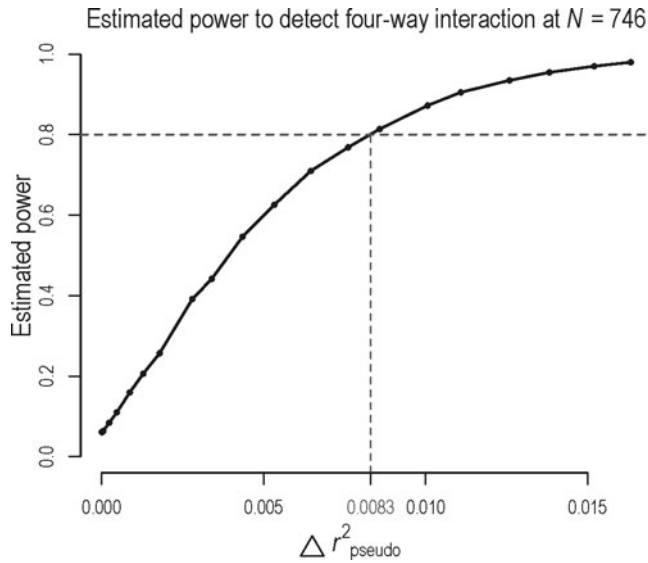


Figure 3. Estimated power to detect four-way interactions of varying effect sizes within full models based on simulated data and our achieved sample size of 746.

Discussion

The main finding of this study is that a motivated numeracy effect can be reproduced in a Western European sample using immigration policy rather than gun control as the controversial topic. In addition, we find that both the H–E and the I–C dimensions of political orientation are associated with this motivated numeracy effect. However, we were not able to reproduce the four-way interaction (involving greater polarization among high-numeracy than low-numeracy participants) indicative of *increased* polarization among high-numeracy partisans. This may be due to differences between the American participants in the original study and our European participants, to the difference between the gun control controversy and the immigration controversy or to some other (set of) factor(s). We also note that there is evidence that high-numeracy partisans tend to place different evaluative emphasis on the same conditional probabilities (Van Boven *et al.*, 2019), which might partially explain our results. That said, we also found no evidence of *convergence* among high-numeracy participants with opposing ideologies – that is to say, we found no evidence that being high in numeracy led to *reduced* polarization, which is what one might naively hope for.

Another possibility that could explain the lack of greater polarization among high-numeracy individuals is our introduction of active reasoning tasks for all

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

participants. Each participant was asked either to do research on *rbutr* or their preferred search engine. Scherer *et al.* (2017) found that high-numeracy individuals performed better on the well-known conjunction fallacy when asked to provide reasons in favor of each answer (correct and incorrect). However, the active reasoning in our study was importantly different. Our active reasoning task was open ended and given before participants saw the test question. In the Scherer *et al.* (2017) study, active reasoning was given alongside the test question and participants were explicitly instructed to give reasons in favor of the answers they chose.

As it stands, we are unable to assess whether active reasoning inductions mitigated the motivated numeracy effect in our study. We note that, compared to Kahan and colleagues' study, the conditions could have increased correct responses and thereby diminished the greater political divisiveness among the highly numerate; however, the similarity in the percentage of correct responses (40%) to that of Kahan *et al.* (2017) (41%) suggests that these did not have a strong effect.

In the replicated paper, Kahan and colleagues pit the 'science comprehension thesis' against the 'identity-protective cognition thesis'. Strictly speaking, these are not inconsistent. Problems in public discourse and deliberation could be due to multiple causes, including both poor overall science comprehension and identity-protective cognition on the part of those who would otherwise be well positioned to understand and interpret scientific evidence. Our results suggest that both may be in play. The participants who were low in numeracy would have done better to flip a coin than to trust their own reasoning. The participants higher in numeracy did slightly better than chance, but showed signs of identity-protective cognition and resulting polarization. Together, these results suggest that both improving education and dampening the effects of identity-protective cognition are worth pursuing.

We conclude by discussing the prospects of active reasoning inductions, several limitations of the current study and directions for future research.

Active reasoning

Motivated numeracy about politically contentious issues presents a serious challenge to democratic deliberation and decision-making. In this study, we compared two active reasoning inductions to see whether either was more successful than the other at mitigating the motivated numeracy effect: inviting participants to use their own preferred method of information-seeking about the topic versus using the *rbutr* interface. The results were inconclusive. We found no evidence that either approach is more effective than the other.

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

In both conditions, participants displayed the motivated numeracy effect at a similar level as in the original study. This suggests that the active reasoning conditions introduced unlikely improved numeric reasoning, although a more complex interaction may have occurred. This could be due to any number of reasons. For instance, several participants in the *rbutr* condition reported that the interface was hard to use or broke down. Unfortunately, we are unable to assess how many participants experienced these technical issues. We hold out hope that a different active reasoning induction may help mitigate the motivated reasoning effect. Incorporating a non-active reasoning condition in addition to active reasoning conditions in future work could be helpful in seeing more subtle differences that active reasoning may have on numeracy. However, we should hope that an active reasoning condition would increase the percentage of correct responses over the 41% established in Kahan *et al.* (2017), which did not occur in this study.

Limitations

Our study has several limitations. First, as mentioned above, numeracy and political orientation were confounded for both ideology subscales. Participants with egalitarian (collectivist) politics tended to score higher on the numeracy scale than those with hierarchical (individualist) politics. A follow-up study using stratified sampling would address this limitation. Second, we deviated from our preregistered data collection plan. In the preregistration, we aimed to collect data from 1600 participants. In the end, we could only afford to collect data from 746 participants. This is still a sizable dataset, but with a larger sample we may have been able to detect a potential four-way interaction as in Kahan *et al.*'s original study – though it is worth pointing out that the four-way interaction was nowhere near the threshold for statistical significance in our data.

A further limitation is that our study did not include a condition without active reasoning against which to compare the active reasoning tasks. In future work, it would be worthwhile to include this control. Moreover, it could be beneficial not to control for the amount of time taken in the active reasoning task. It is possible that higher-numeracy participants are more motivated to actively research compared to lower-numeracy participants. Thus, allowing participants to spend more or less time on this task might provide informative information about the relationship between active reasoning effects on motivated numeracy.[12]

12 We would like to thank an anonymous referee for this point.

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

Future directions

We close on a pessimistic and skeptical note about the prospects of dampening identity-protective cognition. In their original paper, Kahan and colleagues suggest that this is possible, and they point to a book review by Kahan *et al.* (2006) of Sunstein (2005) as providing a method for overcoming identity-protectiveness. However, that method turns out to be self-affirmation exercises, which were first developed in the context of responding to stereotype threat (Cohen *et al.*, 2000). Alas, the literature on stereotype threat seems not to be replicating well (Flore & Wicherts, 2014; Paulette *et al.*, 2019), which indicates that self-affirmation is a solution in search of a problem. Of course, this does not mean that self-affirmation cannot be the solution to a different problem. Does self-affirmation dampen identity-protective cognition? Further research is needed to shed light on this question.

We are more enticed, though, by the prospect of using identity itself to dampen identity-protective cognition. Paradoxical as this might sound, it seems quite promising. The way this would work is by cultivating identities that incorporate epistemic aims (e.g., accuracy, reliability, reasonableness). Someone who embodies such an identity would presumably find it threatening to employ lazy heuristics (Van Bavel & Pereira, 2018). More research is needed on this proposal.

Acknowledgments

We are grateful to Paul Slovic, Elliot Berkmann and Jay Van Bavel for comments on a draft of this paper.

Financial support

This study was funded by the Delft Design for Values Institute and the Australian Research Council (grant DP190101507).

References

Alfano, M. (2019), 'Nietzsche's affective perspectivism as a philosophical methodology', in *Nietzsche's Metaphilosophy*, Cambridge University Press.

Ballarini, C. and S. Sloman (2017), Reasons and the "motivated numeracy effect". *Conference paper draft available at* https://pdfs.semanticscholar.org/fa45/7ac478b3f77069cddb75a3017e666495f749.pdf.

Cohen, G., J. Aronson and C. Steele (2000), 'When beliefs yield to evidence: Reducing biased evaluation by affirmation of the self', *Personality and Social Psychology Bulletin*, **26**(9): 1151–64.

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

Cook, J., S. Lewandowsky and U. K. Ecker (2017), 'Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence', *PLoS ONE*, **12**(5): e0175799.

Flore, P. and J. Wicherts (2014), 'Does stereotype threat influence performance of girls in stereotyped domains? a meta-analysis', *Journal of School Psychology*, **53**(1): 25–44.

Garrett, R. K. (2011), 'Troubling consequences of online political rumoring', *Human Communication Research*, **37**(2): 255–274.

Holzer, A., S. Govaerts, S. Bendahan and D. Gillet (2015), 'Towards mobile blended interaction fostering critical thinking', in *MobileHCI'15*, 735–742. ACM.

Holzer, A., N. Tintarev, S. Bendahan, B. Kocher, S. Greenup and D. Gillet (2018), 'Digitally scaffolding debate in the classroom', in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*,ACM, LBW054.

Kahan, D. and E. Peters (2017), 'Rumors of the 'nonreplication' of the 'motivated numeracy effect' are greatly exaggerated', *SSRN electronic journal*, Yale Law & Economics Research (584).

Kahan, D., D. Braman, P. Slovic, J. Gastil and G. Cohen (2005), 'Cultural cognition of the risks and benefits of nanotechnology', *Nature Nanotechnology*, **4**:87–90.

Kahan, D., D. Braman, P. Slovic and C. Mertz (2007), 'Culture and identity-protective cognition: Explaining the white-male effect in risk perception', *Journal of Empirical Legal Studies*, **4**(3): 465–505.

Kahan, D., P. Slovic, D. Braman and J. Gastil (2006), 'Fear of democracy: A cultural evaluation of sunstein on risk', *Harvard Law Review*, **119**(4): 1071–1109.

Kahan, D. M. (2012), 'Cultural cognition as a conception of the cultural theory of risk', in *Handbook of risk theory*, 725–759. Springer.

Kahan, D. M., E. Peters, E. C. Dawson and P. Slovic (2017), 'Motivated numeracy and enlightened self-government', *Behavioural Public Policy*, **1**(1): 54–86.

Khanna, K. and G. Sood (2018), 'Motivated responding in studies of factual learning', *Political Behavior*, **40**(1): 79–101.

Lewandowsky, S., U. K. Ecker, C. M. Seifert, N. Schwarz and J. Cook (2012), 'Misinformation and its correction continued influence and successful debiasing', *Psychological Science in the Public Interest*, **13**(3): 106–131.

Liberali, J. M., V. F. Reyna, S. Furlan, L. M. Stein and S. T. Pardo (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.

Marris, C., I. H. Langford and T. O'Riordan (1998), 'A quantitative test of the cultural theory of risk perceptions: Comparison with the psychometric paradigm', *Risk analysis*, **18**(5): 635–647.

McFadden, D. (1979), 'Quantitative methods for analysing travel behavior of individuals: Some recent developments', in D. A. H. P. R. Stopher (eds), *Behavioural travel modelling*, London, England: Croom Helm, 279–318.

Miller, R. L. and W. Wozniak (2001), 'Counter-attitudinal advocacy: Effort vs. self-generation of arguments', *Current Research in Social Psychology*, **6**(4): 46–57.

Nguyen, H., J. Masthoff and P. Edwards (2007), 'Modelling a receiver's position to persuasive arguments', *Persuasive Technology*, pages 271–282.

Nurse, M. and W. Grant (2019), 'Numeracy in perceptions of climate change risk', *Environmental Communications*.

Paulette, F., J. Mulder and J. Wicherts (2019), 'The influence of gender stereotype threat on mathematics test scores of dutch high school students: A registered report', *Comprehensive Results in Social Psychology*, pp. 1–35.

Peters, E., D. Västfjäll, P. Slovic, C. Mertz, K. Mazzocco and S. Dickert (2006), 'Numeracy and decision making', *Psychological science*, **17**(5): 407–413.

Poortinga, W., L. Steg and C. Vlek (2002), 'Environmental risk concern and preferences for energy-saving measures', *Environment and behavior*, **34**(4): 455–478.
R Core Team (2018), 'A language and environment for statistical computing', *R Foundation for Statistical Computing, Vienna, Austria*.
Scherer, L. D., J. F. Yates, S. G. Baker and K. D. Valentine (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.
Schneider, J., T. Groza and A. Passant (2013), 'A review of argumentation for the social semantic web', *Semantic Web*, **4**(2): 159–218.
Shtulman, A. (2013), 'Epistemic similarities between students' scientific and supernatural beliefs', *Journal of Educational Psychology*, **105**(1): 199.
Stanovich, K. E. (2009), *What intelligence tests miss: The psychology of rational thought*, Yale University Press.
Stanovich, K. E. and R. F. West (1998), 'Who uses base rates andp(d/âĹijh)? an analysis of individual differences', *Memory & Cognition*, **26**(1): 161–179.
Steg, L. and I. Sievers (2000), 'Cultural theory and individual perceptions of environmental risks', *Environment and behavior*, **32**(2): 250–269.
Sunstein, C. (2005), *Laws of Fear: Beyond the Precautionary Principle*, Cambridge University Press.
Thomson, K. S. and D. M. Oppenheimer (2016), 'Investigating an alternate form of the cognitive reflection test', *Judgment and Decision making*, **11**(1): 99.
Toplak, M. E., R. F. West and K. E. Stanovich (2014), 'Assessing miserly information processing: An expansion of the cognitive reflection test', *Thinking & Reasoning*, **20**(2): 147–168.
Tsai, C.-Y., C.-N. Lin, W.-L. Shih and P.-L. Wu (2015), 'The effect of online argumentation upon students' pseudoscientific beliefs', *Computers & Education*, **80**:187–197.
Van Bavel, J. and A. Pereira (2018), 'The partisan brain: An identity-based model of political belief', *Trends in Cognitive Science*, **22**:213–24.
Van Boven, L., J. Ramos, R. Montal-Rosenberg, T. Kogut, D. Sherman and P. Slovic (2019), 'It depends: Partisan evaluation of conditional probability importance', *Cognition*, **188**:51–63.
van Buuren, S. and K. Groothuis-Oudshoorn (2011), 'mice: Multivariate imputation by chained equations in r', *Journal of Statistical Software*, **45**(3): 1–67.
Washburn, A. and L. Skitka (2018), 'Science denial across the political divide: Liberals and conservatives are similarly motivated to deny attitude-inconsistent science', *Social Psychological and Personality Science*, **9**(8): 972–80.
Wasserman, E. A., W. Dorner and S. Kao (1990), 'Contributions of specific cell information to judgments of interevent contingency', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**(3): 509.
Weller, J. A., N. F. Dieckmann, M. Tusler, C. Mertz, W. J. Burns and E. Peters (2013), 'Development and testing of an abbreviated numeracy scale: A rasch analysis approach', *Journal of Behavioral Decision Making*, **26**(2): 198–212.

Appendix 1. Numeracy questionnaire.

Please answer these seven questions to the best of your ability.

Imagine that we roll a fair, six-sided die 1000 times. Out of 1000 rolls, how many times do you think the die would come up as an even number? Answer: ____

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press

In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1000 people each buy a single ticket from BIG BUCKS? Answer: ____ people.

In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percentage of tickets of ACME PUBLISHING SWEEPSTAKES win a car? Answer: ____

If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000? Answer: ____ people

If the chance of getting a disease is 20 out of 100, this would be the same as having a ____% chance of getting the disease.

If you're running a race and you pass the person in second place, what place are you in?

A farmer had 15 sheep and all but 8 died. How many are left? Answer: ____ sheep.

https://doi.org/10.1017/bpp.2020.32 Published online by Cambridge University Press