Motivated Numeracy and Active Reasoning in a Western European Sample

PAUL CONNOR, University of California-Berkeley

EMILY SULLIVAN, MARK ALFANO, and NAVA TINTAREV, Delft University of Technology

Recent work by Kahan et al. [15] on the psychology of motivated numeracy in the context of intra-cultural 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 United States, regarding the liberal-conservative divide over gun control regulation. In this paper, we present the results of a conceptual replication with Western European participants regarding both the hierarchy-egalitarianism and individualism-communalism divides over immigration policy (n=746). We reproduce the motivated numeracy effect, though we do not find evidence of increased polarization of high-numeracy participants.

Additional Key Words and Phrases: motivated numeracy, active reasoning, cultural cognition, conceptual replication

ACM Reference format:

Paul Connor, Emily Sullivan, Mark Alfano, and Nava Tintarev. 2019. Motivated Numeracy and Active Reasoning in a Western European Sample. 1, 1, Article 4 (August 2019), 13 pages.

https://doi.org/10.1145/nnnnnnn.nnnnnnn

Manuscript submitted to ACM

1 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" (ICT), the answer is that human reasoning is negatively affected when new information threatens prior beliefs and values. In a previous study with American participants, Kahan et al. [2017] found support for this hypothesis. People with high numeracy seem to use their reasoning skills selectively: when the topic about which they 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 exercise their capabilities only when the evidence suggested that gun control is effective, whereas high-numeracy conservatives tended to successfully exercise their capabilities 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 individual than among low-numeracy individuals. This suggests that the quantitative reasoning skills of participants with high numeracy skills can become suppressed, 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 Section 2, we contextualize our study in the published literature on motivated numeracy and active reasoning. Then, in Section 3 we explain the methodology used for the current study. In Section 4, we lay out our results and address RQ1 and RQ2. Finally, in Section 5 we discuss limitations of the current study and explore opportunities for future work on this important topic.

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

2.1 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 counter-evidence, weighting competing criteria in decision-making, remembering information, noticing

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

inferential connections, and so on. Much motivated cognition is normatively unobjectionable, 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 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 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 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. The key to interpreting a table like this is to compare not the absolute numbers but the conditional probabilities. For instance, the table pictured here indicates that 223 out of 298 patients who used the cream got better (74.8 percent), whereas 107 out of the 128 patients who did not use the cream got better (83.6 percent). 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.

Kahan et al. [2017] found that higher-numeracy participants were better able to interpret the contingency table than lower-numeracy participants. 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: high-numeracy 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 partisans than between low-numeracy partisans. 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.

To our knowledge, there have been five attempts to reproduce this result – some direct replications, others conceptual replications. First, 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. Second, 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. Third, Washburn and Skitka [2018] (N=1347) replicate and extend the original result by showing that it crops up for both

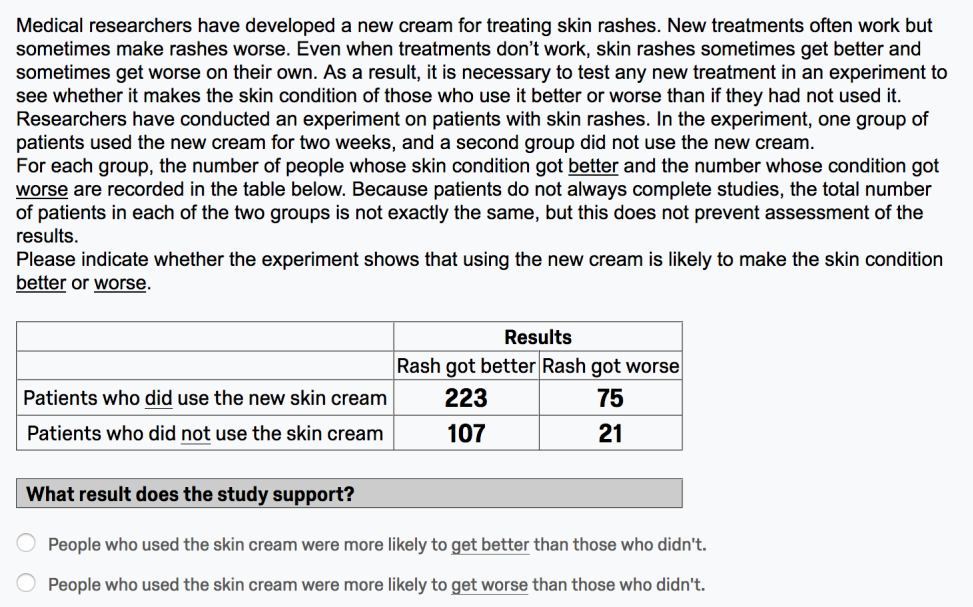

Fig. 1. Example stimulus . This represents the rash condition

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. Fourth, Khanna and Sood [2018] conducted three studies – all using some form of firearms regulation as the controversy – that again replicated the original finding. Finally, Nurse and Grant [2019] conducted 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 United States. 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 States. In addition, all five of these replication studies used a unidimensional measure of political ideology, 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-communitarianism (**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.

2.2 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 [Lewandowsky et al. 2012; Nguyen et al. 2007]. 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 counter-arguments for unfounded beliefs [Miller and Wozniak 2001]. Teaching such active reasoning skills and pointing out flawed argumentation techniques used by providers of misinformation has also been shown to be effective to reduce belief in false information [Cook et al. 2017]. The results suggested a slight increase in item acceptance. Other work introduced a light-weight but effective protocol for supporting debate in a classroom activity with university students. The findings suggest that this intervention led to a 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].

Further, 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 vien, *rbutr* is a software solution that scaffolds peer debates on controversial information right where it appears.[2] It does so by allowing users to post and rate rebuttals for web pages through a browser plugin. In this way, any web page 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 (see Schneider et al. [2013] for a review).

In light of this previous work, we posit that a procedure that encourages active reasoning could decrease the extent that 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, the other using online search).

3 EXPERIMENT

This experiment is a conceptual replication of the study by Kahan et al. [2017].

3.1 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 two-by-two contingency table, the columns of which specified

[2] http://rbutr.com, retrieved August 2019

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*.

Medicine. For the skin rash treatment topic, there were two of the 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" vs. "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 use 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, in the other not. The table indicates the number of people whose level of radicalization decreased and the number whose level of radicalization increased. The wording was kept as comparable as possible to 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.*

3.2 Procedure

In a between-subjects design participants were assigned to one out of 8 conditions (2 by 2 by 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), e.g.,*"Do modern medical treatments work? How effective are they? What strengths or flaws do they have?"*

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 active control is described in the following way: "*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 (see Section 3.2.1) and a questionnaire assessing their Numeracy skills (see Section 3.2.2). The experiment was concluded with a free text comment box for remaining questions or comments from participants.

3.2.1 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 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 "Hierarchy - egalitarianism" (**H-E**, 13 items) and "Individualism - communitarianism" (**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].

3.2.2 Numeracy. To assess Numeracy competence, participants completed the questions in a validated numeracy questionnaire [Weller et al. 2013]. Questions range in difficulty to make it possible to distinguish between participants with various levels of numeracy. A relatively easy question is, "In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?" A relatively difficult question is, "Which of the following numbers represents the biggest risk of getting a disease? (1 in 12 or 1 in 37)."

4 RESULTS

All analyses were conducted in 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 7 missing responses for two items within the Individualism-Collectivism scale, less than 1 percent missing). Multiple imputation was performed using the 'mice' R package [van Buuren and Groothuis-Oudshoorn 2011], and type-II Analyses of Variance (ANOVAs) were performed using the 'miceadds' R package [Robitzsch et al. 2018].[3]

4.1 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 percent female).[4] The majority (68 percent) were British, and a small minority (2 percent) were Dutch, though 28 percent did not specify a nationality. The mean age was 34.75 (StD = 11.61). The majority of participants had either completed a College (227) or a Bachelors degree (294), but there were participants at Elementary school level (7), High school (111), Masters (90), and PhD/JD/MD (17). The mean numeracy score was 5.37 (StD=1.56) out of a maximum of 7.

[3] All code is available at the OSF website associated with this project: https://osf.io/59uv7/?view_only=a4d7c4bc42a8475f9c40a0d24cf66313.
[4] This was fewer than the target of 1600 participants in our pre-registration. Unfortunately, we ran out of money to pay participants and so were not able to collect the full sample.

4.2 Preliminary analysis

We first investigated whether numeracy skills were different based on mean-splits of political scores. 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.

4.3 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 percent of the time, which was significantly lower than chance, .95 CI = [.39, .46]. This is similar to the result in Kahan et al. [2017], who found 41 percent 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). Active reasoning condition had no significant effect on response accuracy, $b = 0.09$, $SE = 0.15$, $t(741.9) = 0.62$, $p = 0.53$. Moreover, there were also no significant two-way interaction effects between active reasoning and topic or result polarity, and no three-way active reasoning by topic by polarity interaction (all $p > 0.16$).

These results suggest that there is no significant difference in the two active reasoning interventions, however there were some issues with the used platform (Rbutr) which are addressed in the discussion. Given the similar performance across the active reasoning interventions, we also collapsed across these two conditions in further analyses. We also compared whether the topic manipulation (medicine and policy) affected accuracy of responses. The average number of correct responses was lower for the policy topic (40 percent) compared to the medicine topic (45 percent), but this difference was not statistically significant ($p = 0.16$).

4.4 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 vs. policy) and result polarity (intervention leads to increase vs. decrease in rashes/radicalization) in determining the probability of correct responses among individuals higher in numerical reasoning ability. Specifically, we hypothesized that liberal-leaning respondents high in numerical reasoning ability would be more likely to respond correctly when the data supported a more liberal policy stance (i.e., when the stricter entrance policy increased radicalization), while more conservative-leaning 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 in the medicine condition, result polarity would have no effect on response accuracy, regardless of respondents' ideology or numeracy. This hypothesis entails a predicted four-way interaction: topic by polarity by respondent numeracy by respondent political ideology.

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). Following Kahan and colleagues, we modelled both numeracy and political ideology as continuous variables, and also modelled non-linear effects of numeracy by

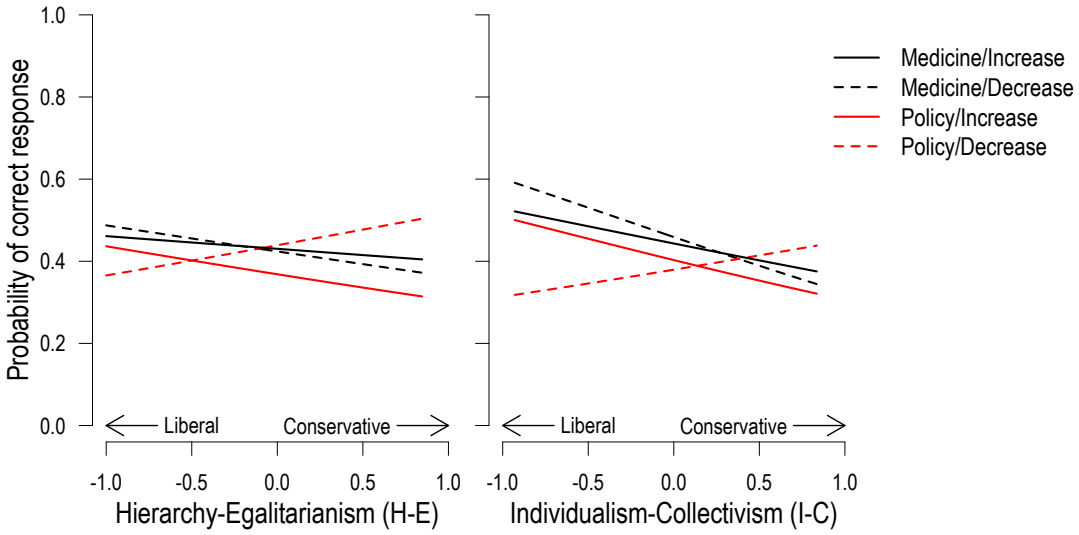

Fig. 2. Predicted probabilities of answering correctly for each topic and polarity type by H-E scores (left panel) and by I-C scores (right panel)

including the squared term of numeracy in each model, as well as its interactions with the experimental conditions. Each model was therefore identical to Kahan and colleagues' full model, with **H-E** and **I-C** scores, respectively, standing in as measures of respondents' political ideology.

For each model, we performed type-II ANOVAs to test each main and interaction effect. Results suggested that the four-way interaction was not significant in either the **H-E** model, $F(1,721.7) = 0.24$, $p = 0.62$, partial eta-squared = .0003, or the **I-C** model, $F(1,723.8) = 0.04$, $p = 0.84$, partial eta-squared = .00007. We therefore did not fully conceptually replicate Kahan and colleagues' results.

However, our results did provide some support for a weaker version of the identity-protective cognition hypothesis. Specifically, ANOVA results indicated significant three-way interactions between topic, polarity, and political ideology in both the **H-E**, $F(1,721.7) = 4.66$, $p = 0.03$, partial eta-squared = .006, and **I-C** models $F(1,723.8) = 4.26$, $p = 0.04$, partial eta-squared = .006. Figure 2 displays the predicted probabilities of answering correctly for each topic and polarity type.[5]

Consistent with a motivated numeracy account, more egalitarian and collectivist respondents were generally more likely to select the correct answer, but when results ran counter to an egalitarian world-view – 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.

No other interaction effects were significant in either model, but we did observe a main effect of numeracy in both the **H-E** model, $F(1, 721.7) = 7.44$, $p = 0.006$, partial eta-squared = .01, and the **I-C** model, $F(1,723.8) = 6.77$, $p = 0.009$, partial eta-squared = .01, with more numerate respondents more likely to give the correct answer. There was also a significant main effect of political ideology in the **I-C** model, $F(1,723.8) = 7.81$, $p = .005$, partial eta-squared = .01, with more liberal/collectivistic respondents more likely overall to select the correct response. Political ideology was not a

[5]In our pre-registration, 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 – and the authors! – to interpret, so we have left them out of this paper.

significant predictor in the **H-E** model, F(1, 721.7) = 2.06, p = 0.15, partial eta-squared = .002. Predicted probabilities of correct response by numeracy scores and political ideology in the **H-E** and **I-C** models are depicted in Figure 3.

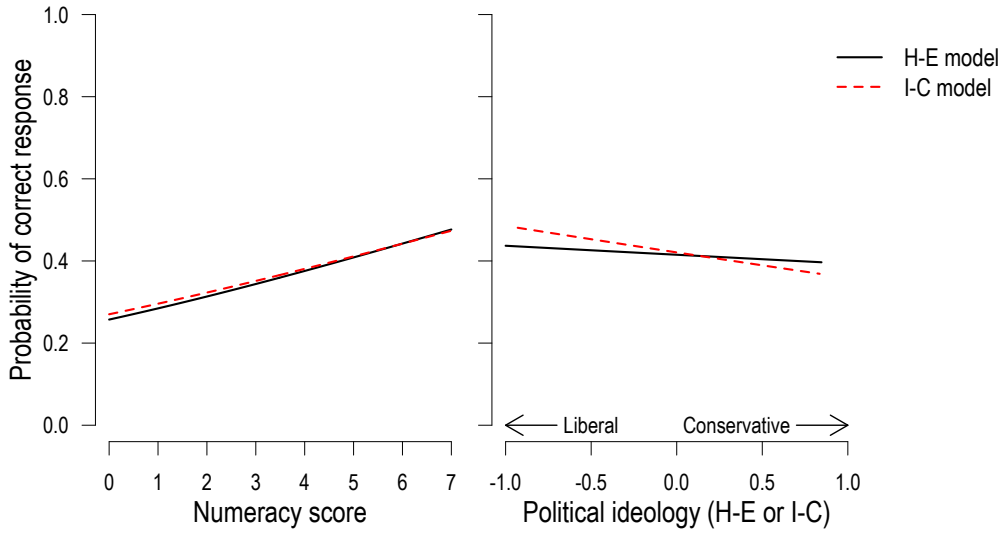

Fig. 3. Predicted probabilities of correct answer by numeracy (with ideology set to its mean) and political ideology (with numeracy set to its mean) from models using **H-E** (black lines) and **I-C** (red dashed lines) as the measure of political ideology.

5 DISCUSSION

The main finding of this study is that a motivated numeracy effect can be conceptually 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. In addition, we found no evidence that different active reading inductions mitigated the motivated numeracy effect differently.

In the replicated paper, Kahan and colleagues pit the "science comprehension thesis" (SCT) against the "identity-protective cognition thesis" (ICT). 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.

5.1 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.

In both conditions participants displayed the motivated numeracy effect, at a similar level as the original study. This suggests that the failure to replicate the 4-way interaction of polarity, topic, numeracy, and ideology is not easily explain by the presence of (some form of) active reasoning for all conditions. That is, active reasoning did not improve numeric reasoning directly, although a more complex interaction may have occurred.

This could be due to any number of causes. For instance, several participants in the rbutr condition reported that the interface was hard to use or broke down. We hold out hope that a different active reasoning induction may help to mitigate the motivated reasoning effect.

5.2 Limitations

Our study has several limitations. First, as mentioned above, numeracy and political orientation were confounded for both ideology subscales. Participants with egalitarian (communitarian) 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 pre-registered data collection plan. In the pre-registration, 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.

5.3 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 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 to not to be replicating well [Flore and 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 and Pereira 2018]. More research is needed on this proposal.

ACKNOWLEDGEMENTS

This study is funded by Delft Design for Values Institute and the Australian Research Council (grant DP190101507). We are also grateful to REDACTED for comments on a draft of this paper.

REFERENCES

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

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

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

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

Core Team, R. (2018). A langugae and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*.

Flore, P. and Wicherts, J. (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., Govaerts, S., Bendahan, S., and Gillet, D. (2015). Towards mobile blended interaction fostering critical thinking. In *MobileHCI'15*, pages 735–742. ACM.

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

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

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

Kahan, D., Slovic, P., Braman, D., and Gastil, J. (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*, pages 725–759. Springer.

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

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

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

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

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

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

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

Robitzsch, A., Grund, S., and Henke, T. (2018). miceadds: Some additional multiple imputation functions, especially for mice. *R package version 2.15-22*.

Schneider, J., Groza, T., and Passant, A. (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.

Sunstein, C. (2005). *Laws of Fear: Beyond the Precautionary Principle*. Cambridge University Press.

Tsai, C.-Y., Lin, C.-N., Shih, W.-L., and Wu, P.-L. (2015). The effect of online argumentation upon students' pseudoscientific beliefs. *Computers & Education*, 80:187–197.

Van Bavel, J. and Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in Cognitive Science*, 22:213–24.

Van Boven, L., Ramos, J., Montal-Rosenberg, R., Kogut, T., Sherman, D., and Slovic, P. (2019). It depends: Partisan evaluation of conditional probability importance. *Cognition*, 188:51–63.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67.

Washburn, A. and Skitka, L. (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.

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C., Burns, W. J., and Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A rasch analysis approach. *Journal of Behavioral Decision Making*, 26(2):198–212.