



# A preregistered replication of motivated numeracy

Emil Persson<sup>a</sup>, David Andersson<sup>a</sup>, Lina Koppel<sup>a</sup>, Daniel Västfjäll<sup>b,c</sup>, Gustav Tinghög<sup>a,d,\*</sup>

<sup>a</sup> Department of Management and Engineering, Division of Economics, Linköping University, 581 83 Linköping, Sweden

<sup>b</sup> Department of Behavioural Sciences and Learning, Linköping University, 581 83 Linköping, Sweden

<sup>c</sup> Decision Research, 1201 Oak Street, Suite 200, Eugene, Oregon 97401, USA

<sup>d</sup> The National Center for Priority Setting in Health Care, Department of Health, Medicine and Caring Sciences (HMC), 581 83 Linköping, Sweden

## ARTICLE INFO

### Keywords:

Motivated numeracy  
Motivated reasoning  
Identity-protective cognition  
Replication

## ABSTRACT

*Motivated numeracy* refers to the idea that people with high reasoning capacity will use that capacity selectively to process information in a manner that protects their own valued beliefs. This concept was introduced in a now classic article by Kahan, Peters, Dawson, & Slovic [2017, *Behavioral Public Policy* 1, 54–86], who used numeracy to index reasoning capacity, and demonstrated that the tendency to engage in ideologically congruent interpretation of facts increased substantially with people's numeracy. Despite the importance of this finding, both from a theoretical and practical point of view, there is yet no consensus in the literature about the factual strength of motivated numeracy. We therefore conducted a large-scale replication of Kahan, Peters, Dawson, and Slovic (2017), using a pre-specified analysis plan with strict evaluation criteria. We did not find good evidence for motivated numeracy; there are distinct patterns in our data at odds with the core predictions of the theory, most notably (i) there is ideologically congruent responding that is not moderated by numeracy, and (ii) when there is moderation, ideologically congruent responding occurs only at the highest levels of numeracy. Our findings suggest that the cumulative evidence for motivated numeracy is weaker than previously thought, and that caution is warranted when this feature of human cognition is leveraged to improve science communication on contested topics such as climate change or immigration.

## 1. Introduction

A fundamental premise of modern democratic society is that people are willing and able to accept empirical facts and to interpret new information in an unbiased manner. However, and in contrast to this ideal, a large literature in behavioral and cognitive sciences has established that people often assess information to protect their own valued beliefs. This is commonly referred to as motivated reasoning (Epley & Gilovich, 2016; Kunda, 1990; Lord, Ross, & Lepper, 1979). A surprising finding in the literature is that motivated reasoning, or partisan disagreement more generally, seems to be more pronounced in people with relatively high levels of scientific knowledge, education, or analytic ability (Drummond & Fischhoff, 2017; Kahan et al., 2012; Kuru, Pasek, & Traugott, 2017; Taber & Lodge, 2006; van Boven et al., 2019; van der Linden, Leiserowitz, & Maibach, 2018). This finding has been rationalized as a type of identity-protective cognition, where individuals use their intelligence and reasoning skills selectively when assessing new information, seeking support for their own valued beliefs (Kahan, 2013; Kahan et al., 2012; Kahan et al., 2017; Kahan, Jenkins-Smith, & Braman,

2011; Landrum, Lull, Akin, Hasell, & Jamieson, 2017).

A highly influential source of empirical support for the hypothesis of identity-protective cognition is the paper by Kahan et al. (2017), who link motivated reasoning to numeric ability, thus establishing the concept of *motivated numeracy*. This paper has had a substantial impact on the subsequent literature and it is widely cited, >500 times on Google Scholar. However, the replicability of this now classic pattern has not been thoroughly investigated. To date, a handful of conceptual replications have been conducted, together with one unpublished replication by authors of the original study; and the results so far have been mixed, with evidence both in favor (Guay & Johnston, 2021; Kahan & Peters, 2017; Nurse & Grant, 2020) and against (Baker, Patel, von Gunten, Valentine, & Scherer, 2020; Connor, Sullivan, Alfano, & Tintarev, 2020; Lind, Erlandsson, Västfjäll, & Tinghög, 2018; Strömbäck, Västfjäll, & Tinghög, 2021) the original findings (see also Ballarín & Sloman, 2017). We therefore conducted a large-scale preregistered replication of Kahan et al. (2017) to investigate the replicability of the motivated-numeracy effect.

\* Corresponding author at: Department of Management and Engineering, Division of Economics, Linköping University, 581 83 Linköping, Sweden.  
E-mail address: [gustav.tinghog@liu.se](mailto:gustav.tinghog@liu.se) (G. Tinghög).

<https://doi.org/10.1016/j.cognition.2021.104768>

Received 25 November 2020; Received in revised form 5 May 2021; Accepted 6 May 2021

Available online 26 May 2021

0010-0277/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

### 1.1. Summary of method and main findings in the original study

In Kahan et al. (2017), subjects were asked to interpret data generated by a fictional experiment. There were four scenarios (containing one fictional experiment each) and subjects were randomly assigned to one of them. Two of the scenarios concerned the effects of a new skin cream developed for treating skin rashes, and the data that was presented indicated either that the skin rash got better (rash decreased) or that it got worse (rash increased) when the new cream was used. The other two scenarios concerned the effects of a gun ban on violent crime, which, unlike the effects of a skin cream, is an issue where there is a substantial political disagreement between liberals and conservatives in the US. In these two scenarios, the data indicated that crime either increased or decreased following a ban on carrying concealed handguns in public. The main finding in Kahan et al. was that subjects were better at interpreting data when the correct interpretation was congruent with their political ideology, and, in particular, that this effect was more pronounced among subjects that had a high numeric ability. Thus, conservatives were better than liberals at interpreting the data in the crime-increase scenario, which is congruent with a conservative view on the probable effects of a gun ban (easier for criminals when law-abiding citizens cannot defend themselves), and the difference between conservatives and liberals was greatest among subjects highest in numeracy. The opposite pattern was found in the crime-decrease scenario, which is congruent with liberal ideology, with liberals performing better than conservatives, and the difference between them was again greater among subjects high in numeracy.

### 1.2. Overview of replication

We conducted a replication of Kahan et al. (2017) by following their original protocol as closely as possible. The replication protocol was vetted (and approved) by authors of the original study (Kahan & Peters) who kindly responded to any query about the original material and also helped us develop three new items for the numeracy scale. All deviations from the original protocol are described in detail in the preregistration. These deviations concerned minor details, possibly with the exception of our decision to run the study on MTurk instead of YouGov, which we knew would result in a slightly different type of sample. For the record, this was also pointed out by Kahan & Peters during the vetting process.

## 2. Method

The study was preregistered at <https://osf.io/pzwta/>. We deviated slightly from the preregistration on a few occasions, which is clearly noted in the text and also listed in Supplementary material Table S1. Data and analysis codes and a transcript of the survey instrument can be found on the project's OSF repository (see link above).

### 2.1. Participants

A total of 3154 MTurk participants completed the study (mean age 40 years, 45% female). We used the  $n \times 2.5$  rule for determining sample size (Simonsohn, 2015) and therefore aimed at  $n = 3200$  responses, but we stopped slightly below our target because the survey was rolled out in batches over a few days and we started to get duplicate responses (some people retake the survey; see Supplementary material Table S1 for details). Subjects were informed upon recruitment that only US citizens were eligible for participation and that we would use an IP-filter to screen out non-US residents during the HIT. Compared to the original study, participants in our sample were younger on average and more educated, and a larger proportion identified as liberals (Table 1). Participants in the replication also scored higher on the nine-item numeracy scale, where the mean number of correct responses was 5.4 (SD = 2.4, range 0–9), compared to 3.7 (SD = 2.1, range 0–9) in the original study. Fig. 1 shows the distribution of numeracy scores in the replication. We

**Table 1**

Sample characteristics for replication and original study.

	Replication	Kahan et al. (2017)
Sample size	3154	1111
Geographic origin	US	US
Medium for data collection	Online, MTurk	Online, YouGov
Year	2020	2013
Age in years	M = 40, SD = 13, range 18–89	M = 48
Sex	45% Female, 55% Male	52% Female, 48% Male
Race	76% White, 9% Black, 4% Hispanic, 7% Asian	73% White, 11% African-American
Education, highest level completed	Med = 4-year college	M = some college
Annual family income, nominal	Med = \$50,000 to \$59,000	M = \$40,000 to \$49,000
Attention check	97% passed, 3% failed	N/A
Numeracy (# correct ans.)	M = 5.4, SD = 2.4	M = 3.7, SD = 2.1
Political party self-identification	46% Democrat 20% Independent 34% Republican	36% Democrat 30% Independent 28% Republican
Political ideology	46% Liberal 18% Moderate 36% Conservative	26% Liberal 27% Moderate 38% Conservative

Notes: Thirty-five respondents (1% of sample size) preferred not to state their income in the replication. Political party self-identification was measured with a seven-point Likert scale in both the original study and in the replication. Political ideology was measured with a seven-point Likert scale in the replication and a five-point Likert scale in the original study. M = mean, Med = median, SD = standard deviation.

can see that the numeracy scores tended to be higher for liberal/Democrat participants than conservative/Republican participants (Fig. 1; mean (SE) difference = 0.87 (0.08) score points). In contrast, in the original study, conservatives/Republicans scored better than liberals/Democrats did, with a mean difference of 0.3 score points between the two groups.

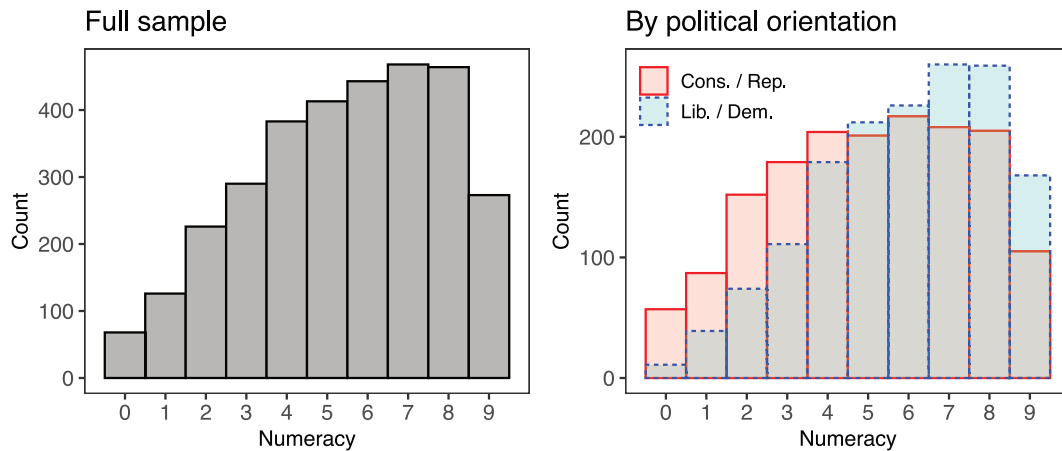
### 2.2. Materials and procedure

Participants were recruited on MTurk for a \$1.25 reward to complete a short survey about social and political issues. The median response time was six and a half minutes and thus the realized median wage for participants was approximately \$11.50 per hour. In the survey, participants read one scenario and answered a question about the fictitious data presented, then followed nine numeracy questions and two questions about political orientation (political ideology and party self-identification, respectively). Finally they answered a few background questions, completed an easy attention check, and received a short debriefing. All participants gave informed consent prior to participation.

Participants were randomly assigned to see one of the four different scenarios: crime increase, crime decrease, rash got better, and rash got worse. The scenarios were the same as in the original study; the two rash scenarios concerned the effects of a new skin cream developed for treating skin rashes and the two crime scenarios concerned the effects of a gun ban on violent crime. The text for the **crime-increase** scenario is reproduced below:

*A city government is trying to decide whether to pass a law banning private citizens from carrying concealed handguns in public. Government officials are unsure whether the law will be more likely to decrease crime by reducing the number of people carrying weapons or increase crime by making it harder for law-abiding citizens to defend themselves from violent criminals.*

*Researchers completed a study of two groups of cities to answer that question. The study involved comparing changes in annual crime rates for one group of cities that had banned concealed handguns with changes in*



**Fig. 1.** Distribution of numeracy scores in the full sample (left panel,  $n = 3154$ ) and by political orientation (right panel). Political orientation is a composite measure of two underlying variables measuring ideology and party self-identification. Cons./Rep. = score above zero on the composite indicator for political orientation ( $n = 1615$ ); Lib./Dem. = score below zero on the composite indicator for political orientation ( $n = 1539$ ).

annual crime rates for a second group of cities that had not banned concealed handguns.

In each group, the number of cities in which the crime rate decreased and the number of cities in which the crime rate increased are recorded in the table below. The exact number of cities in each group is not the same, but this does not prevent assessment of the results.

Please indicate whether the study shows that banning private citizens from carrying concealed handguns in public is likely to decrease crime or to increase crime.

	Result	
	Decrease in crime	Increase in crime
Cities that <u>did</u> ban carrying concealed handguns in public	223	75
Cities that <u>did not</u> ban carrying concealed handguns in public	107	21

*What result does the study support?*

- Cities that enacted a ban on carrying concealed handguns were more likely to have a decrease in crime than cities without bans.
- Cities that enacted a ban on carrying concealed handguns were more likely to have an increase in crime than cities without bans.

The **crime-decrease** scenario was identical to crime increase except the column headings in the results table were reversed, such that the left column was labeled “Increase in crime” and the right column was labeled “Decrease in crime.” The remaining two scenarios, about the skin cream, were constructed using the exact same logic, with the same text and data for both **rash got worse** and **rash got better** but the column headings in the results table were reversed (see section B.1 in the preregistration for more details and a transcript of the scenario text for the rash scenarios). Our main dependent variable (*correct interpretation*) was an indicator variable for correctly interpreting the data presented in the scenario the participant was assigned to.

Numeracy was assessed with nine questions, where the first six were conventional word problems measuring mathematical ability. For example, one of the questions read: “Imagine that we roll a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up as an even number?” The remaining three numeracy questions were taken from the Cognitive Reflection Test (CRT) (Fredrick, 2005), where there is an intuitive but wrong answer that springs to mind quickly. We were worried that some subjects on MTurk would be familiar with the CRT-items and thus (with the help of Ellen Peters) developed three new questions of the same type and approximately same difficulty as the three original CRT questions. For each participant,

we coded all responses as correct (=1) or incorrect (=0) and added them together, which gave a final variable (*numeracy*) measured on a 0–9 scale.

We followed the original study and constructed a composite measure of political orientation, which was a standardized sum of the two underlying variables measuring ideology and party self-identification, respectively. Both these variables were measured on a seven-point Likert scale. For example, on the question about party self-identification, the alternatives were “strong Democrat, Democrat, Independent lean Democrat, Independent, Independent lean Republican, Republican, strong Republican.” Each variable was standardized and then they were added together and then this sum was standardized. This yielded a final variable (*political orientation*) standardized around zero and where positive values indicate more conservative/Republican orientation, and negative values indicate more liberal/Democrat orientation.

### 2.3. Statistical analysis

In our main confirmatory analyses, we conducted a separate logistic regression for each scenario, where numeracy, political orientation, and their interaction were regressed on an indicator variable for correctly interpreting the fictitious data presented in the specific scenario. We then calculated predicted values for each level of numeracy (0–9) and for conservative/Republicans (political orientation evaluated at +1 SD) and liberal/Democrats (political orientation evaluated at –1 SD), respectively. These predictions are plotted in four figures below, one for each scenario, and they are the main basis for assessing the replicability of the motivated-numeracy effect. As specified in the preregistration, we also conducted logistic regressions without the interaction between numeracy and political orientation, to capture the overall effect of motivated reasoning, i.e. whether individuals in general are better at interpreting data that is congruent with their political orientation, but without taking into account that this effect may vary with numeracy.

### 2.4. Evaluation criteria for replication

We evaluated the replicability of motivated numeracy based on how closely we could reproduce the specific pattern found in the original study. The key component of that pattern is that participants should be better at interpreting data that is congruent with their political orientation, and, in particular, that this effect should be more pronounced among subjects with a high numeric ability. This means that conservative/Republicans should be better at interpreting the data in the crime-increase scenario, which is congruent with a conservative view on the probable effects of a gun ban (easier for criminals when law-abiding

citizens cannot defend themselves). Conversely, liberal/Democrats should be better at interpreting the data in the crime-decrease scenario, which is congruent with a liberal view on the probable effects of a gun ban. Our main point of evaluation was based on the same levels of low and high numeracy as in the original study, three and seven correct answers, respectively; but we also looked for a consistent pattern over the full numeracy scale. Our approach was clearly specified in the pre-registration in the form of ten specific criteria (section C.2), which, if fulfilled, would reproduce the specific pattern (for the crime scenarios) found in the original study. For ease of comparison we therefore present our main assessment in graphical form here in the main text (Fig. 4, below), and a point-by-point assessment of each criterion can be found in Supplementary material Table S2.

### 3. Results

Overall, 43% of subjects supplied the correct answer when interpreting the scenario data. This is very close to the original study, where 41% of answers were correct. As expected and in line with the original study, there was a positive effect of numeracy, such that subjects higher in numeracy were more likely to supply the correct answer, both in the rash scenarios (logistic regression; OR = 1.26, SE = 0.030,  $p < .001$ ,  $n = 1549$ ) and in the crime scenarios (OR = 1.15, SE = 0.027,  $p < .001$ ,  $n = 1524$ ).

#### 3.1. Confirmatory analyses

We first investigated whether subjects were better overall at interpreting data congruent with their political orientation, without taking into account that this effect potentially varies with numeracy (Table 2). The expected pattern was observed in the crime-increase scenario, where the odds of correctly interpreting the data was higher for subjects with a more conservative/Republican political orientation, but, surprisingly, there was no similar effect in the crime-decrease scenario. In that scenario, the estimated odds-ratio for political orientation was close to one and insignificant (OR = 0.98, SE = 0.079, 95% CI, 0.84, 1.15), meaning that the odds of correctly interpreting the data, when a ban on concealed handguns was linked to a decreased crime rate, appeared not to be affected by political orientation.

We then went on to investigate whether the extent to which the difference between conservative/Republicans and liberal/Democrats was greatest among subjects highest in numeracy, which is the main focus of the original paper by Kahan et al. There was no indication of such an effect in the crime-increase scenario (Table 3). The estimated odds-ratio for the interaction between political orientation and numeracy was close to one and insignificant (OR = 0.98, SE = 0.031,

**Table 2**  
Analyses for motivated reasoning.

	(1) Crime increase	(2) Crime decrease	(3) Rash got worse	(4) Rash got better
Numeracy	1.15 (0.038)	1.20 (0.043)	1.41 (0.051)	1.14 (0.038)
Political orientation	1.37 (0.105)	0.98 (0.079)	0.94 (0.074)	1.10 (0.082)
	[ $P < .001$ ]	[ $P = .79$ ]	[ $P = .46$ ]	[ $P = .19$ ]
Constant	0.38 (0.076)	0.19 (0.041)	0.12 (0.026)	0.43 (0.084)
Control variables	No	No	No	No
N	756	768	769	780

*Notes:* Logistic regression, estimates given as odds ratios, standard errors in parentheses. Dependent variable is binary for correct interpretation of the data presented in the scenario. Political orientation is a composite measure of two underlying variables measuring ideology and party self-identification, with positive values indicating more conservative/Republican orientation and negative values indicating more liberal/Democrat orientation. Numeracy (0–9) is the number of correct answers on the nine numeracy questions. Subjects who failed the attention check ( $n = 81$ ) were excluded from analysis.

**Table 3**  
Analyses for motivated numeracy.

	(1) Crime increase	(2) Crime decrease	(3) Rash got worse	(4) Rash got better
Numeracy	1.16 (0.039)	1.23 (0.047)	1.41 (0.051)	1.15 (0.039)
Political orientation	1.54 (0.297)	2.61 (0.570)	1.08 (0.241)	1.80 (0.354)
Political orientation × numeracy	0.98 (0.031)	0.84 (0.030)	0.98 (0.035)	0.92 (0.030)
	[ $P = .49$ ]	[ $P < .001$ ]	[ $P = .53$ ]	[ $P = .007$ ]
Constant	0.37 (0.076)	0.15 (0.036)	0.11 (0.026)	0.39 (0.080)
Control variables	No	No	No	No
N	756	768	769	780

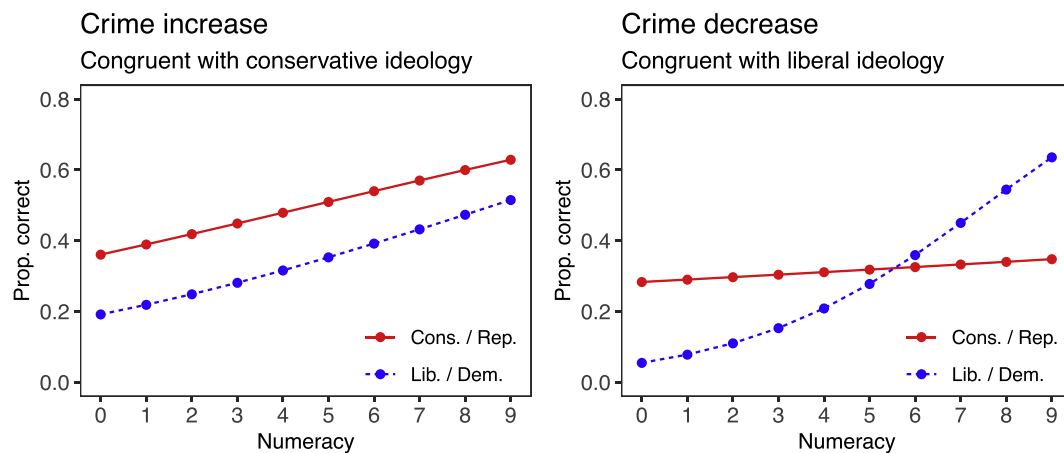
*Notes:* Logistic regression, estimates given as odds ratios, standard errors in parentheses. Dependent variable is binary for correct interpretation of the data presented in the scenario. Political orientation is a composite measure of two underlying variables measuring ideology and party self-identification, with positive values indicating more conservative/Republican orientation and negative values indicating more liberal/Democrat orientation. Numeracy (0–9) is the number of correct answers on the nine numeracy questions. Subjects who failed the attention check ( $n = 81$ ) were excluded from analysis.

95% CI, 0.92, 1.04), meaning that, on average, a fix change in political orientation is associated with a change in the odds of correctly interpreting the data of approximately the same degree irrespective of subjects' numeric ability. For example, the difference between a liberal/Democrat (at  $-1$  SD on political orientation) and a conservative/Republican (at  $+1$  SD on political orientation) in the predicted probability of correctly interpreting the data was 16.8 percentage points for subjects with low numeric ability and 13.8 percentage points for subjects with high numeric ability. On the contrary, in the crime-decrease scenario the estimate for the interaction was in the expected direction and highly significant (OR = 0.84, SE = 0.03, 95% CI, 0.79, 0.90). However, given that we found no overall effect of motivated reasoning in this scenario, this effect is difficult to interpret, and is not by itself in line with the hypothesis of motivated numeracy.

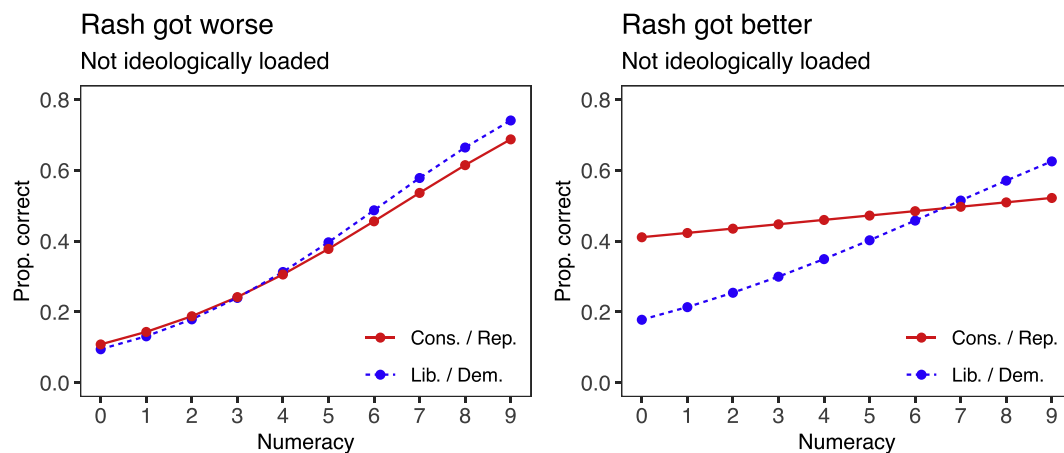
We then calculated predicted values for each level of numeracy (0–9) and for conservative/Republicans (political orientation evaluated at  $+1$  SD) and liberal/Democrats (political orientation evaluated at  $-1$  SD), respectively. These calculations were based on parameters from the regressions in Table 3, and the resulting predictions are plotted in Figs. 2–3, below. Here we can see clearly that there is no tendency for an increasing gap between conservative/Republicans and liberal/Democrats in the crime-increase scenario; the two lines in the figure are almost parallel, and, if anything, the gap seems to become smaller rather than larger at higher levels of numeracy. We can also see the pattern implied by the significant interaction effect in the crime-decrease scenario, where, contrary to what we expected ex-ante, liberal/Democrats performed worse than conservative/Republicans did up and until a numeracy score of six, and only beyond that level is the pattern reversed and thus consistent with motivated numeracy.

We conducted the same type of analysis for the rash scenarios and we decided to include the plotted predictions (Fig. 3) here in the main analysis, although we did not specify this in the preregistration. The rash scenarios are relevant because they constitute a baseline where none of the scenarios (rash got better or rash got worse) is strongly linked to a specific political ideology. Thus, there should be less divergence between liberal/Democrats and conservative/Republicans in the rash scenarios, and we see this to some extent, but not fully, in our analysis. The pattern for rash got worse looks as expected, with no discernible difference due to political orientation but a clear positive effect of numeracy on the probability to correctly interpret the scenario data. The pattern for rash got better looks a bit different, with a non-negligible difference between conservative/Republicans and liberal/Democrats at low levels of numeracy. This effect was unexpected and is difficult to





**Fig. 2.** Proportion correct interpretation for crime scenarios as a function of numeracy, plotted by scenario and political orientation. Predicted probabilities based on the logistic regressions (which are displayed in Table 3), evaluated at  $\pm 1$  SD for political orientation (Cons./Rep. and Lib./Dem., respectively). The stimuli (scenario) in Crime increase is congruent with conservative/Republican ideology and the stimuli in Crime decrease is congruent with liberal/Democrat ideology. Sample sizes are  $n = 756$  for Crime increase and  $n = 768$  for Crime decrease.



**Fig. 3.** Proportion correct interpretation for rash scenarios as a function of numeracy, plotted by scenario and political orientation. Predicted probabilities based on logistic regressions (which are displayed in Table 3), evaluated at  $\pm 1$  SD for political orientation (Cons./Rep. and Lib./Dem., respectively). None of the stimuli (scenarios) is congruent with a specific political ideology. Sample sizes are  $n = 769$  for Rash got worse and  $n = 780$  for Rash got better. These two plots were not listed as confirmatory analyses in the preregistration.

explain from the perspective of motivated numeracy.

### 3.2. Evaluation of replication

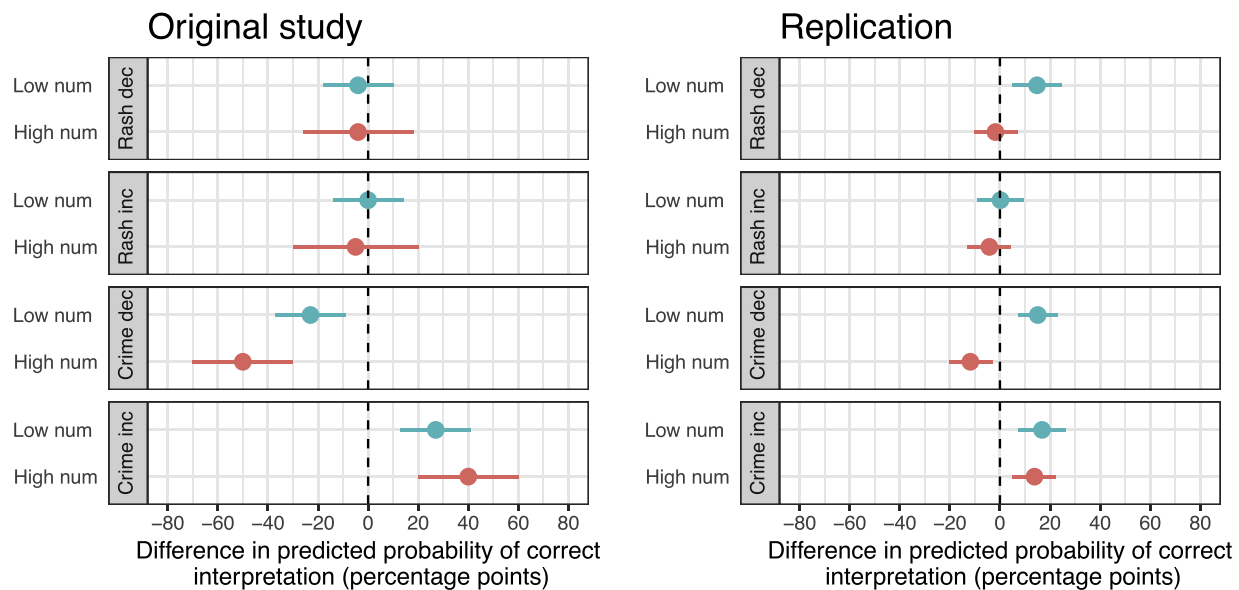
Our primary basis for evaluating the replicability of the motivated numeracy-effect in Kahan et al. (2017) is the predicted probabilities plotted in Fig. 2. Looking at that figure, we can see a pattern that is, at best, only partially consistent with motivated numeracy. Whereas conservative/Republicans are indeed better at interpreting data in the crime-increase scenario, as expected, there is no indication that the gap vis-à-vis liberal/Democrats increases with numeracy. And in the crime-decrease scenario, the overall effect of political orientation is of the wrong sign for more than half of the numeracy scale. Key components of motivated numeracy are thus missing in both scenarios.

As specified in the preregistration we can also compare our results directly with Kahan et al. (2017) by evaluating the differences in predicted probabilities at the same levels of low and high numeracy, namely three and seven correct answers, respectively (see Fig. 4, below). In crime increase, the replication difference between conservative/Republicans and liberal/Democrats is 17 percentage points for low numeracy and 14 for high numeracy, compared to approximately 27 and

40 percentage points, respectively, in the original study. In crime decrease, the replication difference between conservative/Republicans and liberal/Democrats is 15 percentage points for low numeracy and  $-12$  for high numeracy, compared to approximately  $-23$  and  $-50$  percentage points, respectively, in the original study. We can thus see clear points of divergence vis-à-vis the original study, most notably that the difference by political orientation is not greater for subjects with high numeric ability in the crime-increase scenario, and that the difference for subjects with low numeric ability is of the wrong sign in the crime-decrease scenario. Taken together, we find little support for motivated numeracy in our data.

### 3.3. Secondary analyses

We conducted three sets of robustness checks, preregistered as secondary analyses. They involved (i) including control variables in the underlying logistic regressions, (ii) including subjects that failed the attention check, and (iii) using linear instead of logistic regression. The results were very similar across all robustness checks and consistent with the conclusions drawn from the main confirmatory analysis, above. See Supplementary Figs. S1–S3 for details.



**Fig. 4.** Summary of main findings in original study (left panel) and replication (right panel). The left panel is based on results reported in Fig. 8 in Kahan et al. (2017). For the replication, the differences in predicted probabilities are calculated from the predictions shown in Figs. 2–3, at numeracy equal to three (low num.) and seven (high num.) correct answers, respectively. Results for the underlying logistic regressions can be found in Table 3.

#### 4. Discussion

We conducted a large-scale preregistered replication of Kahan et al. (2017), who found that motivated reasoning about the probable effects of a gun ban on violent crime was more pronounced in individuals with high numeric ability; a phenomenon that has been called *motivated numeracy*. Despite its importance, both from a theoretical and empirical as well as practical point of view, there is yet no consensus in the literature about the factual strength of motivated numeracy. Some follow-up studies have found clear support for the phenomenon but others have failed to do so, even suggesting that an effect in the opposite direction is more likely. Our study therefore fills an important gap in this literature.

As expected, we found a positive correlation between numeracy and the likelihood to correctly interpret the fictitious data (in any scenario), and there was a general effect of motivated reasoning, at least in one of the two polarizing scenarios. However, we could not replicate the main finding from the original study, i.e., that motivated reasoning increases with numeracy. In the crime-increase scenario, we found that conservative/Republicans were indeed better at interpreting the data, but the estimated gap vis-à-vis liberal/Democrats did not increase with numeracy; and in the crime-decrease scenario, the overall effect of political orientation was of the wrong sign for more than half of the numeracy scale. Key components of motivated numeracy are thus missing in our data.

Although we fail to replicate the motivated numeracy effect, our findings are in line with a recent literature that examines how cognitive sophistication and analytic ability influence information processing for a wider set of topics and settings. An emerging conclusion in this literature is that motivated numeracy, or the reasoning account of identity-protective cognition more broadly, seems unlikely to generalize beyond a relatively narrow set of conditions (McPhetres & Pennycook, 2019; Pennycook, Cheyne, Koehler, & Fugelsang, 2020; Pennycook, McPhetres, Bago, & Rand, 2020; Pennycook & Rand, 2019; Roozenbeek et al., 2020; Strömbäck et al., 2021; Tappin, Pennycook, & Rand, 2020a). Moreover, even when confined to a narrow set of special topics (gun control, climate change, immigration), follow-up studies of Kahan et al. (2017) have not consistently shown that numeracy exacerbates motivated reasoning; the effects vary substantially across existing

studies. One possible explanation could be that the paradigmatic approach used in much of the literature on motivated reasoning (our study included) seems to conflate prior factual beliefs with political group formation (Bago, Rand, & Pennycook, 2020; Tappin, Pennycook, & Rand, 2020b). Under this account, an alternative explanation for the results in the original paper is that people who are higher in numeracy also have stronger prior beliefs about the impacts of gun control, and thus appear to respond with lower accuracy when interpreting the fictitious data. This could also help explain why our results differ from the original findings, because baseline attitudes toward the effects of gun control may change over time and across samples. At the same time we should acknowledge that our data-contingent analyses show indicative support for a weaker form of identity-protective cognition, since motivated reasoning seems to be taking place primarily at higher levels of numeracy in our sample. More research is clearly needed to identify the limiting conditions under which numeracy should be seen as plausibly harmful for accurate belief formation about scientific issues.

Our study has several strengths, including a large sample and a detailed analysis plan, with stringent evaluation criteria for replicability, but there are also some limitations, most notably that we used a convenience sample from MTurk rather than a more representative sample from YouGov, as in the original study. Generalizations of our results should thus be made with caution, keeping this limitation in mind. Still, several studies have shown high correspondence between results from population based samples and convenience samples from platforms like MTurk (Coppock, 2019; Mullinix, Leeper, Druckman, & Freese, 2015), including conditional average treatment effects, i.e. effects of some experimental stimulus on responses within different subgroups (Coppock, Leeper, & Mullinix, 2018). This is reassuring, and goes some way to alleviate concerns that in particular conservatives on MTurk are special in some (unobserved) sense and that this would invalidate inference based on samples from this platform (Clifford, Jewell, & Waggoner, 2015). Convenience sampling is of course not a perfect replacement for population-based sampling, but the available evidence suggests that it can be useful for advancing a theoretical debate. A theory that applies to the US national population should arguably apply to a subset of that population as well, including a sample of adult Americans drawn from MTurk (Coppock & McClellan, 2019). Thus, the fact that we did not find good evidence for motivated

numeracy in our sample should, at the very least, be taken as incremental evidence against the applicability and generalizability of that theory.

We are not in the business of rejecting the idea of motivated numeracy altogether, but rather see it as a valid hypothesis that merits more research. However, our findings suggest that the cumulative evidence for motivated numeracy is weaker than previously thought, and that caution is warranted when it comes to forming policy conclusions about preferred type and content of scientific communication, also keeping in mind that polarization does not always mean that classical interventions cannot work as intended (van der Linden et al., 2018). Taken together, our replication together with the existing literature suggests that motivated numeracy should not be viewed as a typical feature of human cognition that will impede straight communication of scientific facts, even for highly polarized topics such as gun control or climate change.

### Open data and materials

The preregistration together with data & analysis codes as well as a transcript of the survey instrument can be found at the project's OSF repository, <https://osf.io/pzwta/> (Persson, Andersson, Koppel, Västfjäll, & Tinghög, 2021).

### Funding

This work was supported by the Swedish Research Council [2018-01755] and the Royal Swedish Academy of Sciences [SO2019-0040]. Funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

### Author contributions

E.P., D.A., L.K., D.V., and G.T. designed the study. E.P. analyzed the data and drafted the manuscript. All authors revised the manuscript and approved the final manuscript for submission.

### Declaration of conflicting interests

All authors declare no competing interests that could have appeared to influence the submitted work.

### Acknowledgments

We are grateful to Dan Kahan and Ellen Peters for providing valuable comments and discussion and for responding to any question we had about the original material. Our work has benefited from resources provided via the mass replication project led by Gilad Feldman.

### Appendix A. Supplementary Information

Supplementary information to this article can be found's OSF repository, <https://osf.io/pzwta/> and at <https://doi.org/10.1016/j.cognition.2021.104768>.

### References

- Bago, B., Rand, D. G., & Pennycook, G. (2020). Reasoning about climate change. *PsyArXiv Preprint*. <https://doi.org/10.31234/osf.io/vcpkb>. available at.
- Baker, S. G., Patel, N., von Gunten, C., Valentine, K. D., & Scherer, L. D. (2020). Interpreting politically-charged numerical information: The influence of numeracy and problem difficulty on response accuracy. *Judgment and Decision making*, 15(2), 203–213.
- Ballarín, C., & Sloman, S. A. (2017). Reasons and the “motivated numeracy effect”. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, 1580–1585.
- van Boven, L., Ramos, J., Montal-Rosenberg, R., Kogut, T., Sherman, D. K., & Slovic, P. (2019). It depends: Partisan evaluation of conditional probability importance. *Cognition*, 188, 51–63. <https://doi.org/10.1016/j.cognition.2019.01.020>.

- Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from Mechanical Turk valid for research on political ideology? *Research & Politics*, 2(4), 1–9. <https://doi.org/10.1177/2053168015622072>.
- Connor, P., Sullivan, E., Alfano, M., & Tintarev, N. (2020). Motivated numeracy and active reasoning in a Western European sample. *Behavioural Public Policy*, 1–23. <https://doi.org/10.1017/bpp.2020.32>.
- Coppock, A. (2019). Generalizing from survey experiments conducted on Mechanical Turk: A replication approach. *Political Science Research and Methods*, 7(3), 613–628. <https://doi.org/10.1017/psrm.2018.10>.
- Coppock, A., Leeper, T. J., & Mullinix, K. J. (2018). Generalizability of heterogeneous treatment effect estimates across samples. *Proceedings of the National Academy of Sciences*, 115(49), 12441–12446. <https://doi.org/10.1073/pnas.1808083115>.
- Coppock, A., & McClellan, O. A. (2019). Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. *Research & Politics*, 6(1), 1–14. <https://doi.org/10.1177/2053168018822174>.
- Drummond, C., & Fischhoff, B. (2017). Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. *Proceedings of the National Academy of Sciences*, 114(36), 9587–9592. <https://doi.org/10.1073/pnas.1704882114>.
- Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic Perspectives*, 30(3), 133–140. <https://doi.org/10.1257/jep.30.3.133>.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42. <https://doi.org/10.1257/089533005775196732>.
- Guay, B., & Johnston, C. (2021). Ideological asymmetries and the determinants of politically motivated reasoning. *American Journal of Political Science*. forthcoming.
- Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision making*, 8(4), 407–424.
- Kahan, D. M., Jenkins-Smith, H., & Braman, D. (2011). Cultural cognition of scientific consensus. *Journal of Risk Research*, 14(2), 147–174. <https://doi.org/10.1080/13669877.2010.511246>.
- Kahan, D. M., & Peters, E. (2017). Rumors of the “nonreplication” of the “motivated numeracy effect” are greatly exaggerated. In *Yale law & economics research paper no 584*.
- Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, 1(1), 54–86. <https://doi.org/10.1017/bpp.2016.2>.
- Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, 2, 732. <https://doi.org/10.1038/nclimate1547>.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>.
- Kuru, O., Pasek, J., & Traugott, M. W. (2017). Motivated reasoning in the perceived credibility of public opinion polls. *Public Opinion Quarterly*, 81(2), 422–446. <https://doi.org/10.1093/poq/nfx018>.
- Landrum, A. R., Lull, R. B., Akin, H., Hasell, A., & Jamieson, K. H. (2017). Processing the papal encyclical through perceptual filters: Pope Francis, identity-protective cognition, and climate change concern. *Cognition*, 166, 1–12. <https://doi.org/10.1016/j.cognition.2017.05.015>.
- Lind, T., Erlandsson, A., Västfjäll, D., & Tinghög, G. (2018). Motivated reasoning when assessing the effects of refugee intake. *Behavioural Public Policy*, 1–24. <https://doi.org/10.1017/bpp.2018.41>.
- van der Linden, S., Leiserowitz, A., & Maibach, E. (2018). Scientific agreement can neutralize politicization of facts. *Nature Human Behaviour*, 2(1), 2–3. <https://doi.org/10.1038/s41562-017-0259-2>.
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098–2109. <https://doi.org/10.1037/0022-3514.37.11.2098>.
- McPhetres, J., & Pennycook, G. (2019). Science beliefs, political ideology, and cognitive sophistication. *OSF Preprints*. <https://doi.org/10.31234/osf.io/ad9v7>. available at.
- Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, 2(2), 109–138. <https://doi.org/10.1017/XPS.2015.19>.
- Nurse, M. S., & Grant, W. J. (2020). I'll see it when I believe it: Motivated numeracy in perceptions of climate change risk. *Environmental Communication*, 14(2), 184–201. <https://doi.org/10.1080/17524032.2019.1618364>.
- Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2020). On the belief that beliefs should change according to evidence: Implications for conspiratorial, moral, paranormal, political, religious, and science beliefs. *Judgment and Decision making*, 15(4), 476–498.
- Pennycook, G., McPhetres, J., Bago, B., & Rand, D. G. (2020). Attitudes about COVID-19 in Canada, the U.K., and the U.S.A.: a novel test of political polarization and motivated reasoning. *PsyArXiv Preprint*. <https://doi.org/10.31234/osf.io/zjhjkp>.
- Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50. <https://doi.org/10.1016/j.cognition.2018.06.011>.
- Persson, E., Andersson, D., Koppel, L., Västfjäll, D., & Tinghög, G. A preregistered replication of motivated numeracy. (2021). Dataset available on Open Science Framework at <https://osf.io/pzwta/>.
- Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., ... van der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around

- the world. *Royal Society Open Science*, 7(10), 201199. <https://doi.org/10.1098/rsos.201199>.
- Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, 26(5), 559–569. <https://doi.org/10.1177/0956797614567341>.
- Strömbäck, C., Västfjäll, D., & Tinghög, G. (2021). *Motivated reasoning fast and slow*. Unpublished manuscript.
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755–769. <https://doi.org/10.1111/j.1540-5907.2006.00214.x>.
- Tappin, B. M., Pennycook, G., & Rand, D. G. (2020a). Bayesian or biased? Analytic thinking and political belief updating. *Cognition*, 204, 104375. <https://doi.org/10.1016/j.cognition.2020.104375>.
- Tappin, B. M., Pennycook, G., & Rand, D. G. (2020b). Rethinking the link between cognitive sophistication and politically motivated reasoning. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000974>. forthcoming.