

# Reasons and the “Motivated Numeracy Effect”

**Cristina Ballarini (cristina\_ballarini@brown.edu)**

Department of Cognitive, Linguistic, and Psychological Sciences, Brown University  
190 Thayer St, Providence, RI 02912 USA

**Steven Sloman (steven\_sloman@brown.edu)**

Department of Cognitive, Linguistic, and Psychological Sciences, Brown University  
190 Thayer St, Providence, RI 02912 USA

## Abstract

Does the ability to reason well make one less likely to engage in motivated reasoning? Following a paradigm used by Kahan, Peters, Dawson, and Slovic (2013), this study aims to replicate, extend, and explain the surprising finding that those most likely to process politicized data in a biased manner are those who score highest on a measure of numerical proficiency. Although our study found general effects of motivated reasoning, we failed to replicate Kahan et al.’s “motivated numeracy effect”. However, our study *did* find that, when forced to consider competing statistical interpretations of the data before responding, highly numerate participants were more likely than less numerate ones to choose a correct but belief-contradicting interpretation of data. These results suggest that while numerate participants were biased when generating responses, they were not when evaluating reasons to justify their responses.

**Keywords:** reasoning; motivated reasoning; decision making, science communication; inference; intelligence; rationality

## Introduction

When science becomes politicized, who can we trust to maintain objectivity? Conventional wisdom tells us that when it comes to assessing politically-charged data, those most capable of seeing past their biases and recognizing “the facts” are those most proficient in quantitative reasoning. If that’s the case, then people high in *numeracy*—a measure of the disposition and capacity to engage with quantitative information—ought to process information more objectively and therefore exhibit less bias in assessing it. However, a body of research suggests not only that polarization *increases* with numeracy, but also that highly numerate people process politicized data in a more biased manner (Kahan, Peters, Wittlin, Slovic, Ouellette, Braman, & Mandel, 2012; Kahan, Peters, Dawson, & Slovic, 2013; Kahan, Jenkins-Smith, & Braman, 2010).

In one of these studies (Kahan et al., 2013), participants were faced with a problem that tested their “ability to draw valid causal inferences from empirical data.” Participants all saw the same two-by-two data table, but the data were framed either as the results of a pharmaceutical study of a new rash cream *or* the results of a study of the effects of gun control on crime rates. Correctly interpreting the results required participants to detect covariance between the relevant intervention and two outcomes, and the numbers in each cell of the table were chosen such that a conclusion drawn using

one of two known “heuristic strategies” (comparing either the absolute value of positive outcomes *or* the difference in positive and negative outcomes between the two groups) did not agree with a conclusion correctly drawn using the “covariance strategy” (comparing the *ratio* of positive to negative outcomes between the two groups). Unsurprisingly, the authors found that participants highest in numeracy were most likely to give the correct response in the politically-neutral *rash cream* version of the problem. But in the politically-sensitive *gun control* version, highly numerate participants only performed better than their low numeracy counterparts in cases where using the covariance strategy would lead to a conclusion that aligned with their political beliefs.

This “motivated numeracy effect” fits with a large body of literature on motivated reasoning. In her comprehensive review, Kunda (1990) writes that while “people are more likely to arrive at conclusions that they want to arrive at... their ability to do so is constrained by their ability to construct seemingly reasonable justifications for these conclusions”. In many cases, such a “reasonable justification” will take the form of a processing strategy biased towards a favorable outcome (Giner-Sorolla & Chaiken, 1997). If the information is quantitative in nature, this processing strategy may be a statistical heuristic (Ginossar & Trope, 1987; Petty & Cacioppo, 1986). Taken together, these studies predict that if a sophisticated strategy like covariance detection is required to arrive at a favorable conclusion, then people low in numeracy will be limited in their ability to engage in motivated reasoning. On the flip side, just because highly numerate people possess both a heuristic and normative strategy does not mean that their application of these strategies will be unbiased. To this point, Stanovich and West (2008) found that studies in which correlations *were* found between cognitive ability and unbiased processing supplied cues which signaled that unbiased processing was required—most often, these studies employed a within-subjects design where the conflict between a normative and heuristic rule was transparent. When this conflict was obscured, Stanovich, West, and Toplak (2013) found the degree of “myside bias”—the tendency to evaluate evidence, generate evidence, and test hypotheses in a manner biased towards one’s priors—to be uncorrelated with measures of cognitive ability.

If the motivated numeracy effect depends crucially on both the salience and availability of heuristic and normative statistical strategies, what would happen if participants were to respond to Kahan et al.’s problems with both strategies at

hand? If motivated participants are inexorably biased in their selection of statistical strategies (as Kahan et al. suggest in their discussion), then providing “reasons” that describe each strategy before participants make their final judgement should infect the responses of those low in numeracy with the same bias exhibited by those high in numeracy. Alternatively, if motivation only obscures the need for heuristic override, then providing reasons may reduce bias in one of two ways. Assuming both low and highly numerate participants are equally cued by the reasons, performance should improve for all participants, regardless of whether the data they encounter are belief-affirming or belief-contradicting. However, if those high in numeracy are better able to recognize and respond to a conflict between normative and heuristic rules, then providing all participants with reasons should only benefit those high in numeracy.

## Study

This study examined the effect of reasons on the motivated numeracy effect, and attempted to replicate and generalize Kahan et al.’s (2013) results.

## Method

**Participants** Seventy-six undergraduates at Brown University participated in the study for course credit. Each participant attended one of four identical group sessions (excluding one who completed the survey during an individual session). All participants completed the study on personal computers.

**Procedure** Each participant was randomly assigned either to the “no reasons” or “reasons” condition. In both conditions, nine fictional experiments were described and the observed results were presented in the form of either a table, a scatterplot, or a histogram (three of each in total). Six of these data analysis problems dealt with politicized issues and three of the problems dealt with non-politicized issues.

In the “no reasons” condition, participants saw the results of the experiment and, on the same page, were asked to indicate which conclusion the described experiment supported. To respond, participants could select either between one of two opposing conclusions or could select “Other” and fill in a response. In the “reasons” condition, participants saw the results and, on the same page, were asked to select “the interpretation that best explains the data”. Then, on a separate page, participants saw the data again and were asked to respond just as participants in the “no reasons” condition did.

Immediately following the study, participants completed a nine-item numeracy scale and were asked to report their political outlook (5-point Likert scale), political affiliation (7-point Likert scale), and prior beliefs about each of the politicized issues presented in the study (7-point Likert scales). In a final series of questions, participants shared their

suspicious, confusions, and any other thoughts concerning the study.

**Stimuli** The six motivated problems looked at (1) the effect of gun control measures on crime rates, (2) the effect of mandatory anti-bias training on the number of minority civilians shot by police, (3) the effect of affirmative action on company profitability, (4) the effect of undocumented immigrant populations on violent crime rates, (5) the effect of stop-and-frisk practices on crime rates, and (6) the effect of taxing coal on unemployment rates. The three neutral problems looked at (1) the effectiveness of a rash cream, (2) the effectiveness of a fertilizer, and (3) the relationship between a property’s distance from a city and the real estate commission earned on that property. For each problem, the conclusion that the data supported was randomized for each participant by switching column or axis labels.

Of the nine problems, three presented results in table form (*gun control*, *affirmative action*, and *rash cream*), three in the form of a scatterplot (*stop-and-frisk*, *immigration*, and *property*), and three in the form of a histogram (*anti-bias*, *coal*, and *fertilizer*). A random sequence of three blocks of table-scatterplot-histogram was generated, and the order in which the nine problems appeared was then counterbalanced using a Latin square. Selecting the correct answer in case of a table or a histogram required participants to detect covariance, just as in Kahan et al.’s (2013) original study. Selecting the correct answer in case of a scatterplot required participants to notice an overall positive (about 0.40) or overall negative (about -0.40) correlation rather than extreme outliers.

For tables and histograms, the interpretations presented in the “reasons” condition appealed to one of the two known “heuristic” responses first described by Wasserman, Dornier, and Kao (1990) or to covariance (a total of three reasons plus a fill-in-the-blank, “Other” response option). For scatterplots, the reasons drew attention to either overall correlation or outliers (a total of two reasons plus a fill-in-the-blank, “Other” response option).

Table 1: Reasons Provided for *Gun Control* Problems

Heuristic A-C	Heuristic A-B	Covariance
<i>The group that <u>did</u> ban carrying concealed handguns in public has more cities who saw an <u>[increase/decrease]</u> in crime than does the group that did <u>not</u> ban carrying concealed handguns in public.</i>	<i>Comparing the number of cities that saw crime <u>[increase/decrease]</u> to the number of cities that saw crime <u>[decrease/increase]</u>, there is a greater difference for the group that <u>did</u> ban carrying concealed handguns in public.</i>	<i>The ratio of cities that saw crime <u>[decrease/increase]</u> to cities that saw crime <u>[increase/decrease]</u> is larger for the group that <u>did</u> ban carrying concealed handguns in public than for the group that did <u>not</u>.</i>

The numeracy scale used included five questions adapted from Weller et al. (2013) and four CRT questions—three

from Toplak, West, and Stanovich (2014) and one from Frederick (2005).

## Results

Participants performed at ceiling on problems in which the results were presented in the form of a scatterplot ( $M_{\text{stop-and-frisk}} = 0.95$ ,  $M_{\text{immigration}} = 0.91$ ,  $M_{\text{property}} = 0.99$ ), so responses to these problems were excluded from the analysis. Responses to the *coal* problem were also excluded, as the mean extremity of priors reported at the end of the study indicated that, for this group of participants, the issue was not a motivated one. On the *coal* issue, participants reported priors that averaged 0.64 points from “No opinion”, compared to 1.6, 2.1, and 2.3 points on affirmative action, gun control, and bias, respectively. That left responses to three motivated problems (*gun control*, *anti-bias*, *affirmative action*) and two neutral problems (*rash cream*, *fertilizer*) from each participant for analysis—228 responses in total.

**Bias and Priors** Participants were predominantly liberal-democratic. In terms of outlook, 1% identified as “conservative” or “very conservative on a five-point scale. In terms of affiliation, 5% identified as Republican leaning, Republican, or strong Republican on a seven-point scale. On all three motivated issues, participants reported substantial bias consistent with their liberal leanings—on a seven-point scale, 5% did not support gun control, 1% did not support mandatory anti-bias training in police departments, and 5% did not support affirmative action. No significant difference in reported priors was found between HN participants and LN participants on gun-control ( $t(72.99) = -0.26$ ,  $p = 0.80$ ), anti-bias training ( $t(64.47) = 0.49$ ,  $p = 0.62$ ), or affirmative action ( $t(72.39) = -0.14$ ,  $p = 0.89$ ). To simplify the analysis, problems that supported the conclusion that gun control lead to a *decrease* in crime, that anti-bias training lead to a *decrease* in the number of minority civilians shot by police, or that affirmative action lead to an *increase* in company profit are labeled as “motivated affirming”. Problems that supported the conclusion that gun control lead to an *increase* in crime, that anti-bias training lead to an *increase* in the number of minority civilians shot by police, or that affirmative action lead to a *decrease* in company profit are labeled as “motivated contradicting”.

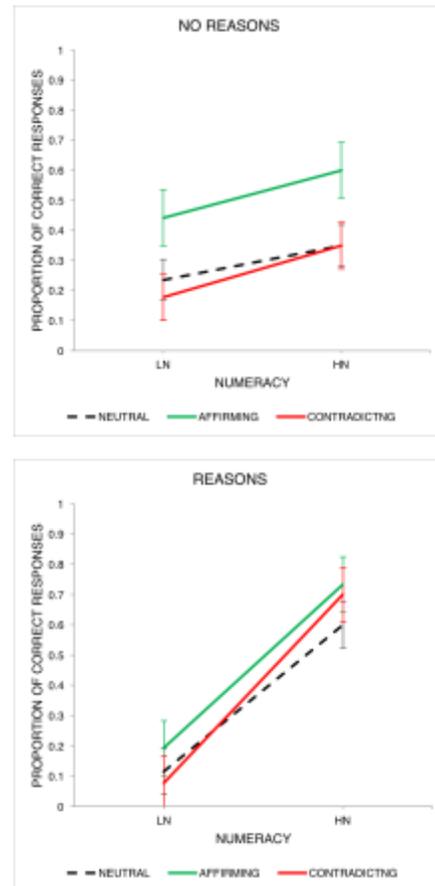
**Numeracy** Average numeracy was 6.25 out of nine. Numeracy classes were assigned by a median split, with “high numeracy” (HN) referring to participants with numeracy scores of 7 or above and “low numeracy” (LN) referring to participants with numeracy scores of 6 or below. The sizes of the resulting groups were 36 and 40, respectively. Numeracy scores were higher in the “reasons” condition ( $M = 6.58$ ,  $SD = 2.39$ ) than in the “no reasons” condition ( $M = 5.92$ ,  $SD = 1.75$ ), though not significantly so;  $t(67.74)$ ,  $p = 0.175$ . The difference can be attributed to CRT items: Though not significant, the two groups differed in performance on the four CRT items,  $t(71.24) = -1.50$ ,  $p =$

0.14, while performance on the other five numeracy items was the same,  $t(71.31) = -0.79$ ,  $p = 0.43$ .

**Motivated Reasoning and the Effect of Numeracy** After excluding 21 participants who did not encounter at least one problem of each valence (i.e. neutral, motivated affirming, and motivated contradicting), a repeated measures analysis of variance (ANOVA) revealed a significant effect of problem valence on performance,  $F(2,102) = 4.99$ ,  $p = 0.008$ ,  $\eta_G^2 = 0.041$ . This analysis also revealed a significant effect of numeracy on performance,  $F(1,51) = 20.01$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.18$ , with highly numerate participants more likely to respond correctly than their low numeracy counterparts.

Unlike Kahan et al. (2013), in our “no reasons” condition, we found no significant difference in performance between HN and LN participants on neutral ( $t(17.75) = 1.59$ ,  $p = 0.13$ ) motivated affirming ( $t(21.90) = 1.50$ ,  $p = 0.15$ ), and motivated contradicting problems ( $t(22.36) = 0.64$ ,  $p = 0.53$ ). After excluding the 21 participants, we also failed to find a significant interaction between numeracy and problem valence,  $F(2,50) = 0.058$ ,  $p = 0.944$ ,  $\eta_G^2 < 0.001$ . While participants clearly exhibit motivated *reasoning*, we failed to find a significant effect of motivated *numeracy*.

Figure 1: Performance in “No Reasons” and “Reasons” Conditions



**The Effect of Reasons** The analysis *did* reveal a significant interaction between numeracy and reasons,  $F(1,51) = 6.20$ ,  $p = 0.016$ ,  $\eta_c^2 = 0.064$ . HN participants performed significantly better in the “reasons” condition than in the “no reasons” condition, ( $M_{\text{reasons}} = 0.68$ ,  $M_{\text{no reasons}} = 0.43$ ,  $t(55.52) = 2.34$ ,  $p = 0.023$ ), while LN participants performed significantly *worse* with reasons ( $M_{\text{reasons}} = 0.13$ ,  $M_{\text{no reasons}} = 0.28$ ,  $t(88) = 2.15$ ,  $p = 0.034$ ). Note, however, that for both groups of participants, the “reasons” manipulation served to reduce bias, as measured by the difference in performance on motivated affirming and motivated contradicting problems (Figure 1). For HN participants,  $Bias_{\text{no reasons}} = 0.25$ , while  $Bias_{\text{reasons}} = 0.03$ . For LN participants  $Bias_{\text{no reasons}} = 0.26$ , while  $Bias_{\text{reasons}} = 0.11$ .

## General Discussion

While reasons served to reduce bias in all participants, those high in numeracy were better able to make use of those reasons to improve their performance. These data suggest that, in the presence of motivation, reasoners are not inexorably biased in their selection of statistical strategies—comparatively evaluating reasons can serve to block the effects of motivation, but only if one is able to understand those reasons. Motivation may encourage less reflective reasoning, but this effect is not irreparable.

Though we failed to find a significant effect of motivated numeracy, it is important to note that Kahan et al.’s (2013) original study analyzed 1111 observations from 1111 participants, while our study analyzed 206 observations from 55 participants (55 after the 21 participants who didn’t encounter problems of each valence were excluded). As Kahan et al. note, in this paradigm, the “strength of inferences drawn from ‘null’ findings depends heavily on statistical power”, and our sample may have been too small to detect the effect of motivated numeracy those researchers found. If Kahan et al.’s original findings are valid, these results support the hypothesis that the motivated numeracy effect results from belief bias obscuring the need for heuristic override. While motivation biased HN participants in their selection of an appropriate statistical strategy, they could appreciate the correct strategy when it was presented (and when the conflict between normative and heuristic strategies was apparent). But in any case, whether polarization increases with numeracy or whether it remains constant, our results suggest that evaluating reasons can reduce the effect of this polarization on reasoning.

If reasons served to block the effects of motivated reasoning, in virtue of what did they do so? The data suggest that evaluating reasons may have encouraged reflectiveness. To this point, not only were CRT scores higher in the “reasons” condition, but considered together, our four CRT items were the best predictor of a correct response on motivated contradicting, motivated affirming, and neutral problems. Recall that numeracy scales were completed after responding to the data analysis problems, suggesting that

evaluating reasons may have elicited a more analytic frame of mind.

Table 2: Correlations Between Numeracy Scale Items and Task Performance

Numeracy Scale Item	Performance on Motivated Tasks	Performance on Neutral Tasks
N1	-0.03	0.17
N2	0.11	0.11
N3	0.07	0.20
N4	0.07	0.17
N5	0.22	0.10
CRT1	0.43	0.31
CRT2	0.29	0.33
CRT3	0.32	0.32
CRT4	0.44	0.39
NUMERACY-CRT	0.17	0.24
NUMERACY (incl. CRT)	0.42	0.43
CRT	0.52	0.47

These results additionally suggest that CRT is not just a measure of numeracy, a position debated in the literature (Liberali, Reyna, Furlan, Stein, & Pardo, 2012); the CRT scale was always a better predictor of performance on the covariance detection task than the numeracy scale considered without the CRT items. While there was a ceiling effect for three of the numeracy items (N1, N2, N4), the two items for which this effect was absent still showed a lower correlation with performance compared to the CRT. Out of the non-CRT items, only N5 (a Bayes’s rule problem) showed a correlation comparable to any of the CRT items. However, unlike N1-N4, N5 may be more of a measure of reflectiveness than quantitative ability, *per se*—even with a frequency chart, the majority of people fail to attend to base-rates in problems like N5 (Bar-Hillel, 1980; Gigerenzer & Hoffrage, 1995), and this base-rate neglect is correlated with low CRT scores (Hoppe & Dusterer, 2011).

That said, considering that those low in numeracy generally performed *worse* with reasons, what appears to be a decrease in bias may not be the result of increased reflectiveness. To this point, unlike their highly numerate counterparts, LN participants in the “reasons” condition performed *worse* on CRT items (average scores of 0.94) than they did in the “no reasons” condition (average score of 1.25). Why might this have been the case? One hypothesis is that those low in numeracy had trouble understanding the reasons provided. But the number of LN participants who reported experiencing some confusion during the study (12%) was comparable to the number of HN participants who reported experiencing confusion (9%). What’s more likely is that those low in numeracy were unable to appreciate and make use of the covariance strategy when it was presented as a reason.

There are three alternative explanations for our results that are important to consider. First, the difference in CRT scores between the “reasons” and “no reasons” condition may

not have reflected an effect of reasons on reflectiveness, but only an unfortunate selection confound. However, it's not clear how the presence of such a confound would affect our conclusions. As our analyses conditioned on numeracy, selection bias would only affect sample sizes, not mean performance scores.

It might also be suggested that the effect of reasons resulted from a task demand. Because in the "reasons" condition, participants saw similar reasons presented with both neutral and motivated versions of the problem, they may have come to suspect that the study was testing their bias. This may have been the case with HN participants in the "reasons" condition, 32% of whom reported suspicions about the study (e.g. "I thought that this study was probably testing how our beliefs influence our abilities to analyze the data"). Fewer LN participants (19%) reported suspicions about the experiment, suggesting that if such a task demand was present, HN participants were better at picking up on (as well as responding to) it. The crucial point, though, is that even if HN participants were responding to a task demand, they could only supply responses that they thought experimenters wanted to hear if they could determine what those responses were. The fact that their responses were so often correct is consistent with our conclusion that HN participants were better at recognizing a correct response.

Third and most importantly, it could be argued that our results support an alternative explanation of Kahan et al.'s motivated numeracy effect: namely, that HN participants were more motivated because they had stronger priors. Here, we found no difference in the extremity of priors between low and highly numerate participants, and we also failed to find a significant motivated numeracy effect. Ultimately, while it is not clear how the alternative explanation could explain the effect of reasons, this is a pressing question for future research.

The implications of these results for science communication complicate Kahan et al.'s (2013) conclusions. While Kahan et al. concluded that "improving public understanding of science and propagating critical reason skills... cannot be expected to dissipate persistent public conflict over decision-relevant science", our study indicates that understanding and being able to make use of normatively correct interpretational strategies can make people more responsive to debiasing efforts, at least when those efforts encourage reflective processing.

## Conclusion

These data suggest that providing reasons can block the effects of motivated reasoning, and that such intervention is most successful for those high in numeracy. Though highly numerate people are more able to recognize when a sophisticated statistical strategy is appropriate, this recognition is impaired when a more immediate, heuristic strategy points to their desired conclusion. Making the need for heuristic override salient improves performance for those

high in numeracy, but is not enough to affect those low in numeracy.

## Acknowledgements

We thank Babak Hemmatian and Marc Lluís for their assistance with statistical analysis, and Dan Kahan for his helpful comments. This work was supported by the Varieties of Understanding Project at Fordham University and the John Templeton Foundation. The opinions expressed are the authors' and do not necessarily reflect the views of the Varieties of Understanding Project, Fordham University, or the John Templeton Foundation.

## References

- Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44(3), 211-233.
- Frederick S. (2005). Cognitive reflection and decision making. *J. Econ. Perspect.* 19, 25-42
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological review*, 102(4), 684.
- Giner-Sorolila, R., & Chaiken, S. (1997). Selective use of heuristic and systematic processing under defense motivation. *Personality and Social Psychology Bulletin*, 23(1), 84-97.
- Ginossar, Z., & Trope, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology*, 52(3), 464.
- Hoppe, E. I., & Kusterer, D. J. (2011). Behavioral biases and cognitive reflection. *Economics Letters*, 110(2), 97-100.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, 80(4), 237.
- Kahan, D. M., Jenkins-Smith, H., & Braman, D. (2011). Cultural cognition of scientific consensus. *Journal of Risk Research*, 14(2), 147-174.
- Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2013). Motivated numeracy and enlightened self-government. *Yale Law School, Public Law Working Paper*, (307).
- Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature climate change*, 2(10), 732-735.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, 108(3), 480.
- Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual Differences in Numeracy and Cognitive Reflection, with Implications for Biases and Fallacies in Probability Judgment. *Journal of Behavioral Decision Making*, 25(4).
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In *Communication and persuasion* (pp. 1-24). Springer New York.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of personality and social psychology*, 94(4), 672.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science*, 22(4), 259-264.

Toplak M. E., West R. F., Stanovich K. E. (2014). Assessing miserly information processing: an expansion of the Cognitive Reflection Test. *Think. Reason.* 20, 147–168.

Weller J. A., Dieckmann N. F., Tusler M., Mertz C. K., Burns W. J., Peters E. (2013). Development and testing of an abbreviated numeracy scale: a Rasch analysis approach. *J. Behav. Decis. Mak.* 26, 198–212.