

# Motivated numeracy and enlightened self-government

DAN M. KAHAN\*

*Yale University, USA*

ELLEN PETERS

*The Ohio State University, USA*

ERICA CANTRELL DAWSON

*Cornell University, USA*

PAUL SLOVIC

*Decision Research & University of Oregon, USA*

**Abstract:** Why does public conflict over societal risks persist in the face of compelling and widely accessible scientific evidence? We conducted an experiment to probe two alternative answers: the ‘science comprehension thesis’ (SCT), which identifies defects in the public’s knowledge and reasoning capacities as the source of such controversies; and the ‘identity-protective cognition thesis’ (ICT), which treats cultural conflict as *disabling* the faculties that members of the public use to make sense of decision-relevant science. In our experiment, we presented subjects with a difficult problem that turned on their ability to draw valid causal inferences from empirical data. As expected, subjects highest in numeracy – a measure of the ability and disposition to make use of quantitative information – did substantially better than less numerate ones when the data were presented as results from a study of a new skin rash treatment. Also as expected, subjects’ responses became politically polarized – and even less accurate – when the same data were presented as results from the study of a gun control ban. But contrary to the prediction of SCT, such polarization did not abate among subjects highest in numeracy; instead, it *increased*. This outcome supported ICT, which predicted that more numerate subjects would use their quantitative-reasoning capacity selectively to conform their interpretation of the data to the result most consistent with their political outlooks. We discuss the theoretical and practical significance of these findings.

Submitted 18 June 2016; accepted 11 October 2016

\* Correspondence to: Yale Law School, PO Box 208215, New Haven, CT 06520, USA. Email: [dan.kahan@yale.edu](mailto:dan.kahan@yale.edu)

## Introduction

Disputed empirical issues – ones that admit of investigation by scientific inquiry – occupy a conspicuous place in US political life. Does consumption of fossil fuels generate global warming? Can nuclear waste be safely stored in deep geologic depositories? Will vaccinating adolescent girls against the human papilloma virus – the dominant cause of cervical cancer – lull them into engaging in unprotected sex, thereby increasing their exposure to other sexually transmitted diseases? Does banning the possession of concealed weapons increase crime – or decrease it? Will aggressive public spending limit the duration and severity of an economic recession – or compound them? Intense and often rancorous conflict on these issues persists, despite the availability of compelling and widely accessible empirical evidence (Kahan, 2010).

Such conflict casts doubt on the prospects for enlightened self-government. Collective welfare demands *empirically informed* collective action. To be sure, decision-relevant science rarely generates a unique solution to any policy dispute: even after the basic facts have been established, *what to do* will involve judgments of value that will vary across citizens who hold competing understandings of the public good. But unless citizens and their representatives possess empirically sound understandings of the dangers they face and the likely effects of policies to abate them, they will not even be able to identify, much less secure enactment of, policies that advance their ends.

Regardless of their political outlooks, then, all democratic citizens have a stake in dispelling persistent public conflict over decision-relevant science. Satisfying this common interest itself demands empirical study aimed at enlarging knowledge of why citizens of diverse outlooks disagree not just about what to do, but also about what is known to science.

In this paper, we present the results of a study aimed at testing rival accounts of public controversy over decision-relevant science. We begin with a brief overview of these two positions. We then describe the study and report the results. Finally, we offer an assessment of what the findings imply for future study of political conflict over policy-relevant facts and practical steps to dispel it.

## Polarization over decision-relevant science: two accounts

As is the case with most interesting social phenomena (Watts, 2011), the number of plausible explanations for persistent public controversy over risks and other policy-relevant facts exceeds the number that can actually be true. We identify two of the possibilities.

The first one attributes such conflict to deficits in the public's capacity to comprehend scientific evidence. The public, on this account, has a weak understanding of science (Miller & Pardo, 2000). Ordinary citizens are thus liable to misunderstand what scientists are telling them and vulnerable to being misled by those trying to deceive them for private advantage (McCaffrey & Buhr, 2008; Rosenau, 2012).

The public's limited knowledge is aggravated by psychological dynamics. Popular risk perceptions, it is thought, tend to originate in a rapid, heuristic-driven form of information processing – what decision scientists refer to as 'System 1' reasoning (Stanovich & West, 2000; Kahneman, 2003). Over-reliance on System 1 heuristics is the root of myriad cognitive biases. By fixing attention on emotionally gripping instances of harm, or by inducing selective attention to evidence that confirms rather than disappoints moral predispositions, System 1 information processing induces members of the public variously to overestimate some risks and underestimate others relative to the best available evidence, the proper evaluation of which requires exercise of more deliberate and reflective 'System 2' forms of information processing (Loewenstein *et al.*, 2001; Sunstein, 2003; Weber, 2006; Marx *et al.*, 2007; Weber & Stern, 2011).

We will call this account the 'science comprehension thesis' (SCT). SCT is an extremely popular explanation for conflicts over climate change and various other disputed risks, particularly among commentators who construct secondary interpretive accounts by synthesizing diverse findings from decision science (Sunstein, 2005, 2006, 2007).

We will call the second account of persistent controversy over policy-relevant facts the 'identity-protective cognition thesis' (ICT). ICT, in an important sense, stands SCT on its head. Whereas SCT attributes conflicts over decision-relevant science to deficits in science comprehension, ICT sees the public's otherwise intact capacity to comprehend decision-relevant science as disabled by cultural and political conflict.

Individuals, on this account, have a large stake – psychically as well as materially – in maintaining the status of, and their personal standing in, affinity groups whose members are bound by their commitment to shared moral understandings. If opposing positions on a policy-relevant fact (e.g. whether human activity is generating dangerous global warming) come to be seen as symbols of membership in and loyalty to competing groups of this kind, individuals can be expected to display a strong tendency to conform their understanding of whatever evidence they encounter to the position that prevails in theirs (Kahan *et al.*, 2011; McCright & Dunlap, 2013). Being a form of motivated reasoning (Kunda, 1990), identity-protective cognition can be viewed as a psychic self-defense mechanism that steers individuals

away from beliefs that could alienate them from others on whose support they depend in myriad domains of everyday life (Giner-Sorolla & Chaiken, 1997; Sherman & Cohen, 2006).

The conditions that trigger this sort of response are rare. Very few facts amenable to empirical investigation ever become symbols of group identity. The US public is not polarized over the utility of antibiotics in treating bacterial infections, the health risk of cell phone radiation, the efficiencies associated with governmental provision of public safety services such as fire and police protection, the disadvantages of relying on a system of bartering or privately guaranteed commercial paper rather than a common currency issued by and guaranteed by the government, etc. In the absence of divisive cultural conflict, citizens of all levels of science comprehension generally form positions that are consistent with the best available evidence. Moreover, there is every reason to believe in such circumstances that those who enjoy higher than average capacities for science comprehension use those capacities to make even better science-informed decisions.

But when a policy-relevant fact *does* become suffused with culturally divisive meanings, the pressure to form group-congruent beliefs will often dominate whatever incentives individuals have to ‘get the right answer’ from an empirical standpoint. On matters like climate change, nuclear waste disposal, the financing of economic stimulus programs, and the like, an ordinary citizen pays no price for forming a perception of fact that is contrary to the best available empirical evidence: that individual’s personal beliefs and related actions – as consumer, voter, or public discussant – are too inconsequential to affect the level of risk that person or anyone else faces or the outcome of any public policy debate. However, if that person gets the ‘wrong answer’ in relation to the one that is expected of members of his or her affinity group, the impact could be devastating: loss of trust among peers, stigmatization within his community, and even loss of economic opportunities (Kahan, 2012).

Thus, at one level – a very individualistic one – it will make perfect sense in this situation for individuals to attend to information, including evidence of what is known to science, that promotes the formation of identity-congruent beliefs. Again, even citizens of modest science literacy and critical reasoning skills will likely be able to form such beliefs without difficulty, because figuring out what view prevails among those with whom one shares one’s most important connections depends on a basic kind of cultural competence, not on an understanding of or a facility with empirical evidence. But those citizens who enjoy above-average science comprehension will not face any less incentive to form such beliefs; indeed, they will face pressure to *use* their intelligence and reasoning skills to find evidentiary support for identity-congruent

beliefs the comprehension of which would likely exceed the capacity of most of their peers (Kahan *et al.*, 2012).

At a collective level, of course, this style of engaging decision-relevant science can be disastrous. If all individuals follow it at the same time, it will impede a democratic society from converging – or at least converging as quickly as it otherwise would – on understandings of facts consistent with the best available evidence on matters that affect their common welfare. This outcome, however, will not change the incentive of any individual – who, despite the harm he or she suffers as a result of unaddressed risks or ill-considered policies, cannot change the course of public policymaking by changing his or her personal stances, which, if contrary to the ones that prevail in that person's group, will continue to expose him or her to considerable social disadvantage.

ICT thus sees a particular species of group competition as subversive of the disposition of democratic citizens to use their capacities to comprehend decision-relevant science – or at least to use them in ways that ordinarily steer them toward convergence on the best available evidence. Persistent conflict over risks and other policy-relevant facts reflects a ‘tragedy of the science communications commons’: a misalignment between the individual interests that culturally diverse citizens have in forming beliefs that connect them to others who share their distinctive understanding of the best life and the collective interest that members of all such groups share in the enactment of public policies that enable them to pursue their ends free from threats to their health and prosperity (Kahan *et al.*, 2012; Kahan, 2013).

## Study design and hypotheses

### *Overview*

We undertook a study to test SCT and ICT. The study involved drawing inferences from data generated by a fictional experiment. Both the focus of the experiment – the effectiveness of a new skin rash treatment and the effectiveness of a ban on carrying concealed weapons in public – and its outcome were manipulated. By enabling us to examine whether the subjects' ability to correctly interpret the data would vary only in relation to their proficiency in quantitative reasoning or vary as well in relation to the ideological congeniality of the experiment results, this design effectively pitted SCT and ICT against one another.

### *Sample*

The subjects for the study consisted of a nationally diverse sample of 1111 US adults. The subjects were recruited to participate by Polimetrix/YouGov, which

administered the study via that firm's online testing facilities. The sample was 52% female, and the mean age of the subjects was 48 years. Seventy-three percent of the subjects were white, and 11% were African-American. Twenty-eight percent of the sample self-identified as Republican, 36% as Democrat, and 30% as independent. Twenty-six percent identified themselves as either 'Liberal' or 'Very Liberal'; 38% as 'Conservative' or 'Very Conservative'; and 27% as 'Moderate'. The mean education level was 'some college'; the mean annual income was \$40,000 to \$49,000. The study was administered over a period from late April to mid-May, 2013.

### *Individual characteristic measures*

Subjects furnished standard demographic data, including political affiliations and outlooks. Party self-identification ('dem\_repub') was measured with a seven-point Likert item ('Strong Democrat'; 'Democrat'; 'Independent Lean Democrat'; 'Independent'; 'Independent Lean Republican'; 'Republican'; 'Strong Republican'). Political ideology ('libcon') was measured with a five-point Likert item ('Very Liberal'; 'Liberal'; 'Moderate'; 'Conservative'; 'Very Conservative'). Responses to these two items formed a reliable aggregate Likert scale ( $\alpha = 0.83$ ), which was labeled 'Conserv\_Repub' and transformed into a  $z$ -score to facilitate interpretation (Smith, 2000).

The right-left political orientation measured by Conserv\_Repub is a familiar representation of the basis on which the public is divided on many issues turning on decision-relevant science. More importantly, it is one indicator of the latent group affinities of the sort that can be expected to generate motivated reasoning.<sup>1</sup>

<sup>1</sup> It is not the only measure of these dispositions, of course, and not necessarily the most discerning one (Kahan *et al.*, 2012). An alternative that we have used in previous studies of the 'cultural cognition of risk' (Kahan, 2012) features 'cultural worldviews' assessed along two orthogonal dimensions ('hierarchy-egalitarianism' and 'individualism-communitarianism'). 'Cultural cognition', in this work, is simply the term used to designate the species of identity-protective cognition that accounts for a wide variety of risk controversies (Kahan, 2011). Most of the controversies involving environmental and technological risks (but not many others that feature public health risks [e.g., Kahan *et al.*, 2010]) feature controversy between individuals whose values are more hierarchical and individualistic on the one hand, and others who are more egalitarian and communitarian on the other. So does the controversy over gun risks (Kahan *et al.*, 2007; Gastil *et al.*, 2011). The former tend to be more 'conservative' and 'Republican', the latter more 'liberal' and 'Democratic', although individuals with strong worldviews or orientations of this sort are not in fact highly partisan (Kahan *et al.*, 2012). For accessibility and to promote commensurability between our results and those of researchers who tend to use 'right-left' political outlook measures, we use liberal-conservative ideology and partisan self-identification measures in this paper. As is true in other studies examining conflicts (Kahan *et al.*, 2012, 2017), an analysis of the data using the cultural worldview measures generates results that are the same in nature but even more dramatic in their effects.

We also measured the subjects' numeracy. A well-established and highly studied construct, numeracy encompasses not just mathematical ability, but also a disposition to engage quantitative information in a reflective and systematic way and to use it to support valid inferences (Peters *et al.*, 2006; Liberali *et al.*, 2012).

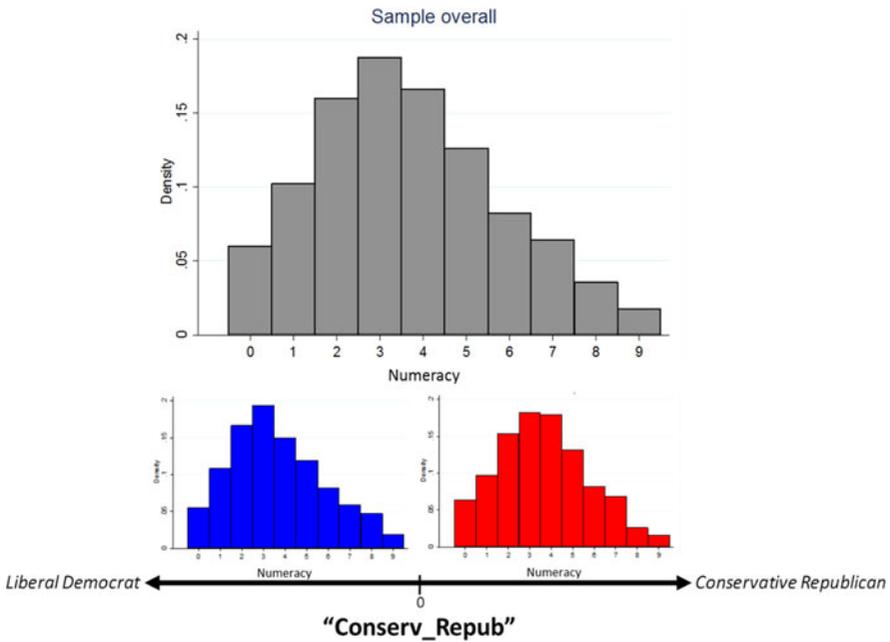
We assessed the numeracy of our subjects with a battery of word problems conventionally used for this purpose (Weller *et al.*, 2012). The mean number of correct responses was 3.7 ( $SD = 2.1$ ). The distribution of scores was essentially normal (kurtosis = 2.6), a result consistent with previous studies aimed at constructing a scale that could be expected to measure variation across the entire range of the latent capacity measured by the numeracy scale ( $\alpha = 0.74$ ). Subjects who scored above the mean on Conserv\_Repub scored slightly higher than those who scored below the mean ( $\Delta M = 0.3$ ,  $t = 2.44$ ,  $p = 0.02$ ) (Figure 1).

### *Stimulus*

The stimulus consisted of four versions of a problem involving interpretation of data and causal inference. The problem described an experiment and the observed results (Figure 2). Those results were reported in a  $2 \times 2$  contingency table, the columns of which specified the number of cases that reflected positive and negative results, respectively, and the rows of which reflected the experimental treatment.

Two of the versions of the experiment involved a skin rash treatment. In these versions, subjects were advised that “[m]edical researchers have developed a new cream for treating skin rashes.” They were also advised that “[n]ew treatments often work but sometimes make rashes worse,” and “skin rashes sometimes get better and sometimes get worse on their own,” whether or not a person is treated. To determine the effect of the new treatment, experimenters (the subjects were told) had divided a number of patients suffering skin rashes into two groups – one that was administered the skin cream and another that was not; after a 2-week period, experimenters totaled up how many of the patients got better and how many got worse in each group. Based on the results, as reflected in the  $2 \times 2$  contingency table (Figure 2), subjects were instructed to indicate whether the “[p]eople who used the skin cream were more likely to get better than those who didn’t,” or instead “[p]eople who used the skin cream were more likely to get worse than those who didn’t.”

The two versions of the skin treatment problem differed only with respect to which result the experiment supported. The numbers in the  $2 \times 2$  contingency table were kept the same, but the labels at the tops of the columns – “Rash got better” and “Rash got worse” – were manipulated (Figure 3).



**Figure 1.** Numeracy scores. Bars denote density distribution of scores on the numeracy scale

Correctly interpreting the data was expected to be difficult. Doing so requires assessing not just the absolute number of subjects who experienced positive outcomes (“rash better”) and negative ones (“rash worse”) in either or both conditions, but rather comparing the *ratio* of those who experienced a positive outcome to those who experienced a negative one in each condition. Comparing these ratios is essential to detecting *covariance* between the treatment and the two outcomes, a necessary element of causal inference that confounds even many intelligent people (Stanovich & West, 1998; Stanovich, 2009).

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

Each of these heuristic strategies generates a recognizable species of invalid causal inference. ‘1 vs 2’ amounts to assessing a treatment without considering information from a control. ‘1 vs 3’ compares outcomes in the treatment and control, but in a manner that neglects to consult information necessary to

Medical researchers have developed a new cream for treating skin rashes. New treatments often work but sometimes make rashes worse. Even when treatments don't work, skin rashes sometimes get better and sometimes get worse on their own. As a result, it is necessary to test any new treatment in an experiment to see whether it makes the skin condition of those who use it better or worse than if they had not used it.

Researchers have conducted an experiment on patients with skin rashes. In the experiment, one group of patients used the new cream for two weeks, and a second group did not use the new cream.

In each group, the number of people whose skin condition got better and the number whose condition got worse are recorded in the table below. Because patients do not always complete studies, the total number of patients in each two groups is not exactly the same, but this does not prevent assessment of the results.

Please indicate whether the experiment shows that using the new cream is likely to make the skin condition better or worse.

	Result	
	Rash Got Better	Rash Got Worse
Patients who <u>did</u> use the new skin cream	223	75
Patients who did <u>not</u> use the new skin cream	107	21

What result does the study support?

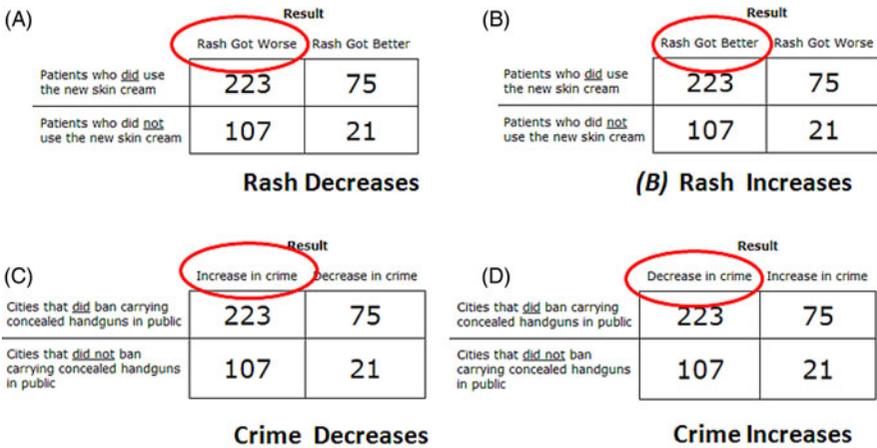
- People who used the skin cream were more likely to get better than those who didn't.
- People who used the skin cream were more likely to get worse than those who didn't.

**Figure 2.** Stimulus from ‘rash increases’ condition. Subjects were advised that researchers had conducted an experiment to determine whether a new skin treatment was effective or instead had adverse effects. The results were summarized in a  $2 \times 2$  contingency table, and the subjects were then instructed to determine whether the experiment showed that the skin conditions of people treated with the cream were more likely to ‘get better’ or ‘get worse’

disentangle the impact of the treatment from other influences at work in both conditions.

In the real world, of course, use of either of these defective strategies – both of which amount to failing to use all of the information necessary to make a valid causal inference – might still generate the correct answer. But for our study stimulus, the numbers in the cells of the contingency table were deliberately selected so that use of either heuristic strategy would generate an incorrect interpretation of the results of the fictional skin treatment experiment.

The second two versions of the experiment involved a gun control measure (Figure 3). Subjects were instructed that a “city government was trying to decide whether to pass a law banning private citizens from carrying concealed handguns in public.” Government officials, subjects were told, were “unsure whether the law will be more likely to decrease crime by reducing the



**Figure 3.** Experimental conditions. Subjects were assigned to one of four conditions. The conditions are identified by labels (A) to (D) in a manner that indicates the result or outcome of the experiment that is most supported by the data contained in the relevant table. The correct interpretation of the data was manipulated by varying the results specified by the headings above the columns

number of people carrying weapons or increase crime by making it harder for law-abiding citizens to defend themselves from violent criminals.” To address this question, researchers had divided cities into two groups: one consisting of cities that had recently enacted bans on concealed weapons and another that had no such bans. They then observed the numbers of cities that experienced “decreases in crime” and those that experienced “increases in crime” in the next year. Again supplied this information in a 2 × 2 contingency table, subjects were instructed to indicate whether “cities that enacted a ban on carrying concealed handguns were more likely to have a decrease in crime,” or instead were “more likely to have an increase in crime than cities without bans.” The column headings on the 2 × 2 table were again manipulated, generating one version in which the data, properly interpreted, supported the conclusion that cities banning guns were more likely to experience increased crime relative to those that had not, and another version in which cities banning guns were more likely to experience decreased crime.

Overall, then, there were four experimental conditions, reflecting opposite experiment results for both the skin treatment version of the problem and the gun ban version. The design was a between-subjects ones, in which individuals were assigned to only one of these conditions. For the sake of expository convenience, we will refer to the conditions as ‘rash decrease’, ‘rash increase’,

‘crime decrease’, and ‘crime increase’, with the label describing the result that a correct interpretation of the  $2 \times 2$  contingency table would most support.

### *Hypotheses*

We formed three hypotheses. The first was that *subjects scoring high in numeracy would be more likely to get the right result in both skin treatment conditions.*

This hypothesis reflected results in previous studies. As indicated, such studies show that the covariance-detection problem featured in this experiment is very difficult for most people to answer correctly (Stanovich, 2009).

One recent study, however, shows that the likelihood of answering the problem correctly is predicted by an individual’s score on the cognitive reflection test (CRT) (Toplak *et al.*, 2011). The CRT features a set of problems, each of which is designed to prompt an immediate and intuitively compelling response that is in fact incorrect. Because supplying the correct answer requires consciously stifling this intuition and logically deducing the right response, the CRT is understood to measure the disposition to use the slower, deliberate form of information processing associated with System 2, as opposed to the rapid, heuristic-driven form associated with System 1.

The CRT requires elementary mathematical skills, but is not a numeracy test per se (Liberali *et al.*, 2012). However, insofar as making valid causal inferences in the covariance-detection problem likewise demands suppressing the heuristic tendency to give decisive significance to suggestive but incomplete portions of the information reflected in the  $2 \times 2$  contingency table, it is not surprising that individuals who score higher on the CRT are more likely to correctly interpret the data that the table contains.

We would expect the numeracy scale to be an even stronger predictor of how likely a person is to select the correct response in the skin treatment versions of this problem. Like the CRT, the numeracy scale measures a disposition to subject intuition to critical interrogation in light of all available information, and thus to avoid mistakes characteristic of overreliance on heuristic, System 1 information processing (Liberali *et al.*, 2012). Indeed, two CRT items are conventionally included in the numeracy scale (Weller *et al.*, 2012), and we added the third in this study in order to reinforce its sensitivity to the disposition to preempt reliance on unverified intuition. However, whereas the CRT measures the disposition to use System 2 information processing generally, the numeracy scale measures how proficient individuals are in using it to reason with quantitative information in particular, a capacity specifically relevant to the covariance-detection problem featured in the stimulus.

The hypothesis that performance in the skin treatment conditions would be positively correlated with numeracy was common to SCT and ICT. The second and third hypotheses reflect opposing SCT and ICT predictions relating to results in the gun ban conditions.

Whereas the issue in the skin treatment versions of the covariance-detection problem – does a new skin cream improve or aggravate a commonplace and non-serious medical condition? – is devoid of partisan significance, the question whether a gun ban increases or instead decreases crime is a high-profile political one that provokes intense debate. Consistent with the growing literature on culturally or ideologically motivated reasoning (Jost *et al.*, 2014), we anticipated that subjects in the gun ban conditions would be more likely to construe the data as consistent with the position that prevails among persons who share their political outlooks – regardless of which version of the problem (‘crime increases’ or ‘crime decreases’) they were assigned. Specifically, we surmised that gratification of the interest subjects would have in confirmation of their predispositions would reinforce their tendency to engage in heuristic reasoning when subjects were assigned to the conditions in which ‘1 vs 2’ or ‘1 vs 3’ generated a *mistaken* answer that was nonetheless congenial to their political outlooks. That ideologically motivated reasoning would compound heuristic reasoning in this way was specifically supported by studies showing that an existing position on a contested nonpolitical issue (Dawson & Gilovich, 2000), aversion to threatening information (Dawson *et al.*, 2002), and prior beliefs (Stanovich & West, 1998) can all magnify the sorts of reasoning errors that are frequently encountered in covariance-detection problems identical or closely related to the one featured in our stimulus.

But in whom should motivated cognition interfere with reasoning in this way and by how much? SCT blames a deficit in public science comprehension for consistent controversy over risks and related facts. Numeracy furnishes one important measure of the capacity to make sense of scientific evidence. Based on SCT, then, it seems reasonable to predict that the degree of ideological polarization expected to be observed in the gun ban conditions would *abate* as subjects become higher in numeracy. Such a prediction would be consistent, too, with the position, advanced by many scholars, that ideologically motivated cognition is itself best understood as a form of the heuristic-driven information processing that is characteristic of System 1 reasoning (Sunstein, 2006; Westen *et al.*, 2006; Marx *et al.*, 2007; Weber & Stern, 2011; Lodge & Taber, 2013). If so, one might expect numeracy, as a capacity that demands the form of conscious reflection associated with System 2, to preempt the motivated reasoning that most people bring to bear on the covariance-detection problem. The second hypothesis, then, was that *subjects scoring higher in numeracy*

would be more likely to construe the data correctly not only when it was consistent with their ideological predispositions, but also when it was inconsistent with them, and thus they were likely to display less ideological polarization than subjects lower in numeracy.

ICT generates a different prediction based on a different understanding of the source of public conflict over decision-relevant science. ICT views ideologically motivated reasoning as a form of identity self-defense that reliably protects individuals' interests by guiding them to construe evidence in a manner that enables them to persist in culturally congenial or identity-affirming beliefs. Persistent political polarization over policy-relevant facts that admit of empirical study is not a consequence of any deficit in science comprehension, but rather a consequence of the disabling impact of symbolic status competition on the disposition of individuals to *use* their ability to comprehend science in a manner geared to producing evidence-congruent beliefs (Kahan *et al.*, 2006).

On this account, we should *not* expect the degree of ideological polarization that is likely to characterize subjects' responses in the gun ban conditions to diminish with numeracy. Indeed, the third hypothesis, one that is associated with ICT but not SCT, is that *ideological polarization in the gun ban conditions should be most extreme among those highest in numeracy.*

Individuals high in science comprehension have a special resource to engage evidence in a manner calculated to generate ideological congenial conclusions. Because the causal-inference problem featured in this study is genuinely difficult, it makes sense to expect subjects to resort to such reasoning when assigned to the condition ('crime decreases' for 'liberal Democrats' and 'crime increases' for 'conservative Republicans') in which less effortful heuristic engagement with the provided information would generate an identity-threatening conclusion for persons with their political outlooks. But those low in numeracy will be less likely to succeed in discerning the correct, identity-affirming conclusion that the data in that condition actually support because they are less likely to possess the sorts of science-comprehension capacities that doing so requires.

At the same time, ICT predicts that high-numerate subjects should not be substantially more likely to correctly interpret the data than low-numeracy ones when assigned to the condition that presents identity-threatening information ('crime increases' for 'liberal Democrats' and 'crime decreases' for 'conservative Republicans'). Engaging in the form of intensive reflection associated with System 2 is cognitively taxing; it is something we should expect individuals to resort to only when *they themselves* are likely to recognize (consciously or otherwise) some advantage in using it. Thus, where less

effortful, heuristic engagement with the information in the gun ban conditions generates an identity-affirming conclusion, ICT predicts that high-numeracy subjects will ‘settle’ for heuristic reasoning, and end up with the *wrong* interpretation of the data, just as their low-numeracy counterparts will (Kahan, 2013).

If high-numeracy subjects use their special cognitive advantage selectively – only when doing so generates an ideologically congenial answer but not otherwise – they will end up even more polarized than their low-numeracy counterparts. Such a result, while highly counterintuitive from the perspective of SCT, would be consistent with the view of a smaller group of scholars who take the view that identity-protective cognition operates on both heuristic and systematic – System 1 and System 2 – forms of information processing (Giner-Sorolla & Chaiken, 1997; Chen *et al.*, 1999; Cohen, 2003; Kahan, 2013). It would also be consistent with, and help to explain, results from observational studies showing that the most science-comprehending citizens are the most polarized on issues like climate change and nuclear power (Hamilton *et al.*, 2012; Kahan *et al.*, 2012).

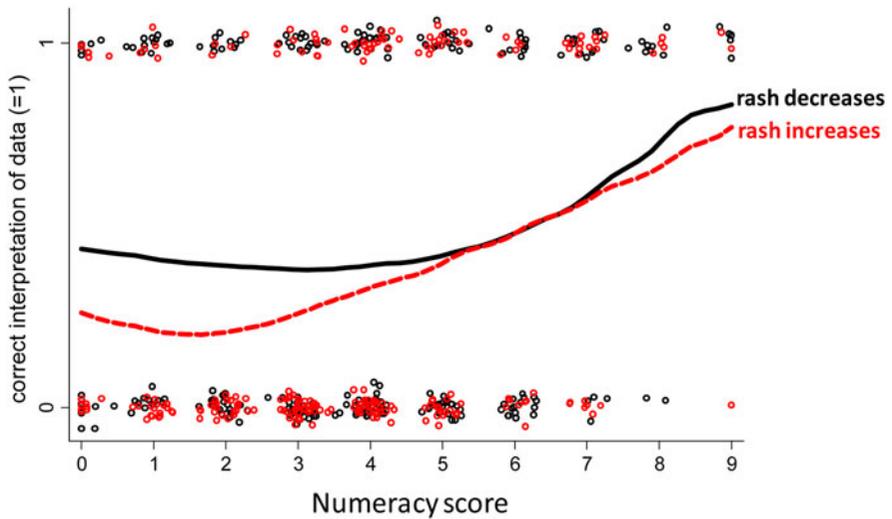
## Results

### *Preliminary analyses*

No matter what condition subjects were assigned to, they were highly likely to select the wrong response to the covariance-detection problem. Overall, 59% of the subjects supplied the incorrect answer.

Figure 4 presents a scatter plot of subject responses in the skin treatment conditions. It supplies strong support for the first hypothesis – that the likelihood of correctly interpreting the data in the skin treatment conditions would be conditional on numeracy. Reflecting the difficulty of the task, subjects of low and even moderate numeracy scores were more likely than not to select the wrong answer in both ‘rash decreases’ and ‘rash increases’ conditions. Even among those scoring in the top 50% on the numeracy scale (four or more answers correct), less than half (48%) supplied the correct answer. It was not until scores on the numeracy scale reached 90th percentile or above (seven to nine correct) that 75% of the subjects correctly identified the result that was most supported by the data in the  $2 \times 2$  contingency table.

Figure 5 presents a scatter plot of subject responses in the gun control condition. The pattern differs from that in the skin treatment conditions. The impact of numeracy on performance in the ‘crime increases’ condition is minimal. The proportion of subjects correctly interpreting the data did increase

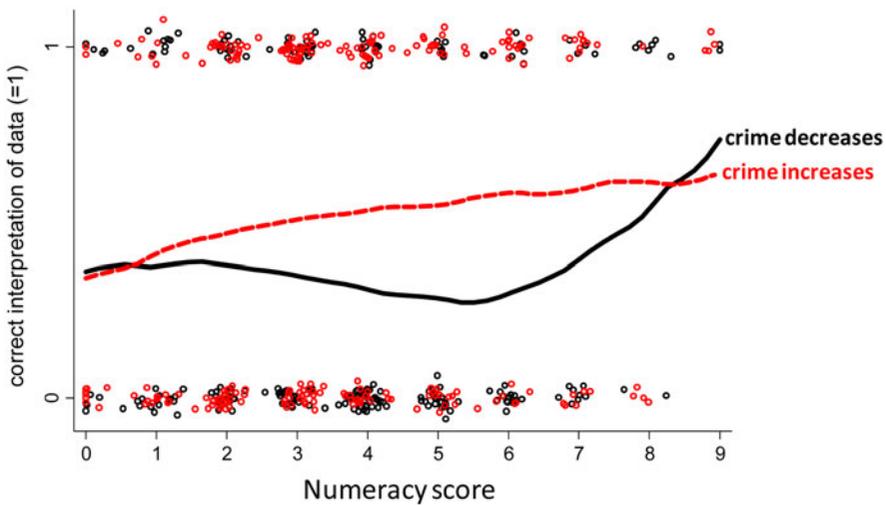


**Figure 4.** Scatter plot of responses in skin treatment conditions. Red circles identify subjects assigned to the ‘rash increases’ condition and black ones to the ‘rash decreases’ condition. Locally weighted regression lines (Cohen *et al.*, 2003) track the proportion of subjects answering the problem correctly in relation to numeracy levels in the two skin treatment conditions

as numeracy increased in the ‘crime decreases’ condition, but even at the highest levels of numeracy, the percentage of subjects who supplied the incorrect response in that condition was relatively high. Overall, even among subjects in the 90th percentile, only 57% of those assigned to one of the gun control conditions correctly identified the outcome most supported by the data. The discrepancy is consistent with the inference that a factor present in the gun control conditions but not in the skin treatment conditions inhibits the contribution that numeracy makes to identifying the correct answer.

Figure 6 plots responses for all four conditions among subjects of opposing political outlooks. Visual inspection demonstrates no meaningful variation among ‘Liberal Democrats’ (subjects scoring below the mean on *Conserv\_Repub*) and ‘Conservative Republicans’ (subjects scoring above the mean) in the skin rash conditions. For both groups, the relationship between numeracy and identifying the result genuinely supported by the data displays the same pattern as was observed in the sample as a whole.

Nevertheless, visual inspection suggests a clear interaction between numeracy and political outlooks in the gun ban conditions (Figure 6). Liberal Democrats become increasingly likely to correctly identify the result supported

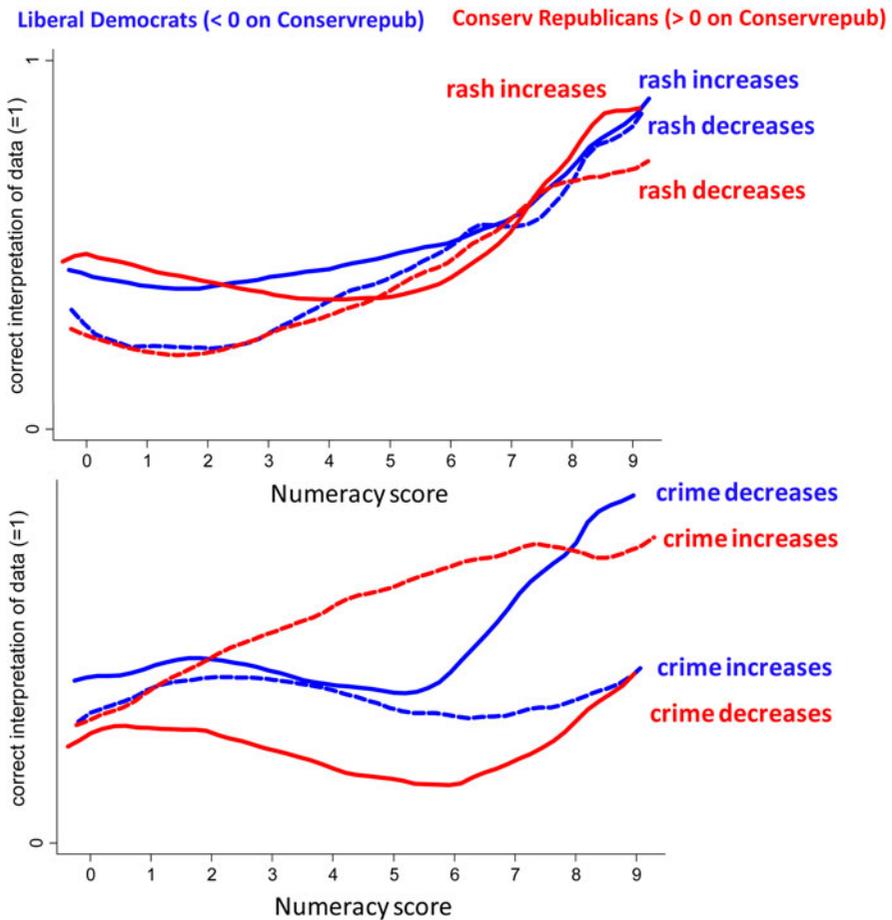


**Figure 5.** Scatter plot of responses in gun ban conditions. Red circles identify subjects assigned to the ‘crime increases’ condition and black ones to the ‘crime decreases’ condition. Locally weighted regression lines (Cohen *et al.*, 2003) track the proportion of subjects answering the problem correctly in relation to numeracy levels in the two gun control conditions

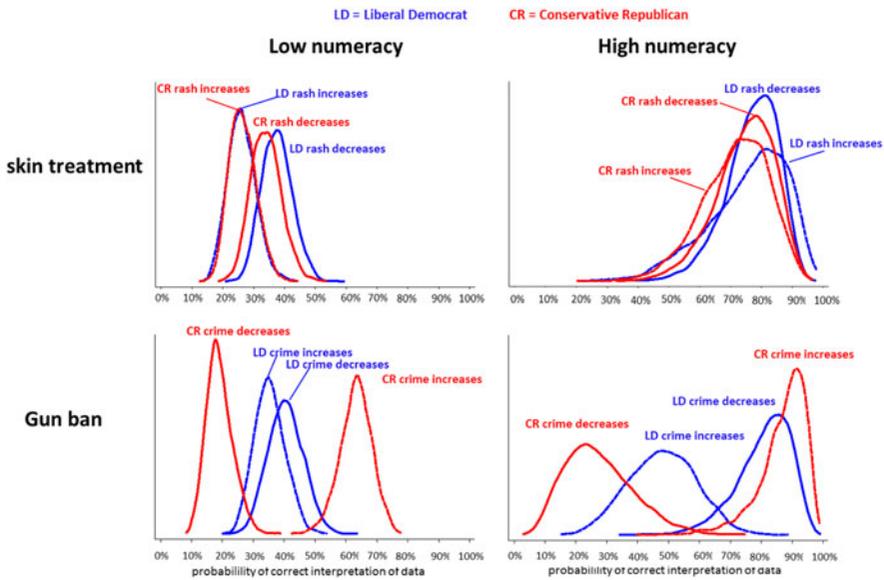
by the data as they become more numerate in the ‘crime decreases’ condition, but increasing numeracy had minimal impact for Liberal Democrats in the ‘crime increases’ condition. Among Conservative Republicans, the pattern was inverted: the impact of higher numeracy on subjects’ ability to supply the correct answer was substantially larger in the ‘crime increases’ condition than in the ‘crime decreases’ one.

In other words, higher numeracy improved subjects’ performance in detecting covariance only in the ‘gun control’ condition in which the correct response was congenial to the subjects’ political outlooks. This result is *inconsistent* with the second SCT hypothesis, which predicted that political polarization – of the form clearly apparent at low and middle levels of numeracy – would abate at higher levels.

Indeed, visual inspection suggests that polarization – as measured by the gap between subjects of opposing political outlooks assigned to the same experimental condition – was greatest among subjects highest in numeracy. Such a result would fit the third ICT hypothesis, which predicted that subjects capable of correctly interpreting the data would resort to the form of effortful, System 2 processing necessary to do so *only* when the less effortful, heuristic, or System 1 assessment of the data threatened their ideological identities.



**Figure 6.** Responses by subjects of opposing cultural outlooks. Locally weighted regression lines (Cohen *et al.*, 2003) track the proportion of subjects answering the problem correctly in relation to numeracy levels in the various conditions. Blue lines plot relationships for subjects who score below the mean and red ones are for subjects who score above the mean on *Conserv\_Repub*, the composite measure based on liberal–conservative ideology and identification with one or the other major party. Solid lines are used for subjects in the condition that the data, when properly interpreted, support the inference that either skin rashes or crime decreased; dashed lines are used for subjects in conditions in which the data, when properly interpreted, support the inference that either skin rashes or crime increased



**Figure 7.** Predicted probabilities of correctly interpreting the data. Density distributions derived via Monte Carlo simulation from regression Model 3, Table A1, when predictors for Conserv\_Repub set at  $-1 SD$  and  $+1 SD$  for ‘liberal Democrat’ and ‘conservative Republican’, respectively, and numeracy set at three and seven correct answers for ‘low numeracy’ and ‘high numeracy’, respectively (King *et al.*, 2000)

### Multivariate analyses

In order to perform a more exacting test of the study hypotheses, a multivariate regression model was fit to the data. Described in more detail in the technical appendix, the regression model contained predictors for subjects’ political outlooks and numeracy and for the experimental condition to which they were assigned.

Consistent with visual inspection of the raw data, the results of this analysis confirm that higher numeracy increases the probability that subjects will correctly interpret the results in the skin treatment conditions. The results also suggest that less numerate subjects are more likely to correctly interpret the data in the ‘rash decreases’ condition than in the ‘rash increases’ condition, but by an amount that is modest in size and nonsignificant ( $9\% \pm 14\%$ ,  $LC = 0.95$ ).

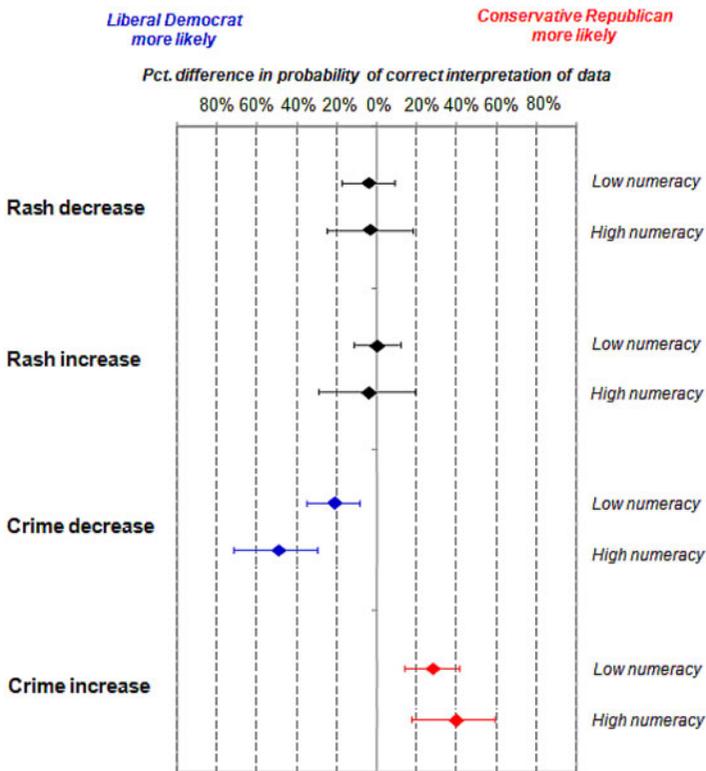
Such outcomes are presented graphically in Figure 7. Generated by Monte Carlo simulation, the density plots illustrate the estimated probability of correctly interpreting the data, and the precision of that estimate, for a

low-numeracy (three correct) and a high-numeracy (seven correct) ‘liberal Democrat’ ( $-1 SD$  on Conserv\_Repub) and for a low-numeracy and a high-numeracy ‘conservative Republican’ ( $+1 SD$ ) in each experimental condition (King *et al.*, 2000).

Figure 7 also strongly disconfirms the second SCT hypothesis. A low-numeracy liberal Democrat is *more* likely to correctly identify the outcome supported by the data than is a low-numeracy conservative Republican when the data, in fact, support the conclusion that a gun ban decreases crime, but is *less* likely to correctly identify the outcome when the data support the conclusion that a gun ban *increases* crime. This pattern of polarization, *contrary to SCT hypothesis*, does not abate among high-numeracy subjects. Indeed, it *increases*. On average, the high-numeracy partisans whose political outlooks were affirmed by the data, when properly interpreted, was 45% more likely ( $\pm 14\%$ ,  $LC = 0.95$ ) to identify the conclusion actually supported by the gun ban experiment than were the high numeracy partisans whose political outlooks were affirmed by selecting the incorrect response. The average difference in the case of low-numeracy partisans was 25% ( $\pm 10\%$ ) – a difference of 20% ( $\pm 16\%$ ). Corroborating the inference that this effect was attributable to ideologically motivated reasoning, there were no meaningful or significant partisan differences among high-numeracy subjects – or low-numeracy ones, for that matter – in the skin treatment conditions (Figure 8). These findings *support the third ITC hypothesis*.

The reason that numeracy amplified polarization, these analyses illustrate, was that high-numeracy partisans were more likely than low-numeracy ones to identify the correct response to the covariance-detection problem when doing so affirmed subjects’ political outlooks. A high-numeracy conservative Republican, the model predicted, was 21% ( $\pm 16\%$ ) more likely than a low-numeracy one to recognize the correct result in the ‘crime increases’ condition; in the ‘crime decreases’ condition, a high-numeracy liberal Democrat was 32% ( $\pm 20\%$ ) more likely than a low-numeracy one to identify the correct response. But when the data, correctly interpreted, threatened subjects’ outlooks, high-numeracy partisans enjoyed no meaningful advantage over their low-numeracy counterparts (3% [ $\pm 16\%$ ] for conservative Republicans in ‘crime decreases’; 11% [ $\pm 20\%$ ] for liberal Democrats in ‘crime increases’), all of whom were *unlikely* to identify the correct response (Figure 7).

This pattern is also consistent with ITC. ITC predicts that where an individual has an identity-protective stake in a particular outcome, he or she will resort to effortful, System 2 processing – of the sort needed to draw valid inferences from complex data – *only* when less effortful heuristic reasoning generates a conclusion that threatens his or her identity. Here, high-numeracy subjects in the gun ban conditions were likely to terminate their engagement with the evidence when heuristic assessment of it gratified their political predispositions – even



**Figure 8.** Predicted differences in probability that partisans will correctly interpret the data. Predicted differences in probabilities derived via Monte Carlo simulation from regression Model 3, Table A1. Predictors for Conserv\_Repub set at  $-1 SD$  and  $+1 SD$  for ‘liberal Democrat’ and ‘conservative Republican’, respectively, and numeracy set at three and seven correct answers for ‘low numeracy’ and ‘high numeracy’, respectively. Confidence intervals indicate the 0.95 levels of confidence

though the resulting inference that they drew about the result of the experiment was incorrect.

At the same time, the source of the contribution that numeracy makes to enlarging polarization in the gun control conditions also helps to address the question of whether subjects of all levels of numeracy were construing the data in a reflexively or automatically partisan fashion without making *any* effort to engage it. This interpretation is not consistent with the data. If this were happening, low-numeracy partisans would have done *just as well* as high-numeracy ones when assigned to the condition in which a correct response was affirming of their identities.

In this particular context, then, *accurately discerning* the identity-affirming outcome depended on a high degree of numeracy. It was the selective exercise of the special capacity that higher numeracy confers in this regard that aggravated partisan polarization among high-numeracy subjects.

The regression analysis also identified one additional main effect. Even after accounting for the effects of political outlooks and numeracy, being assigned to ‘crime increases’ as opposed to the ‘crime decreases’ condition substantially improved subject performance in the covariance-detection problem (Table A1, Model 3). The size of the effect ( $b = 1.07$ ,  $z = 4.02$ ,  $p < 0.01$ ) is equivalent to a 26% increase ( $\pm 12\%$ ), which can be interpreted as how much more likely an individual of mean political outlooks and mean numeracy would be to identify the correct result in the ‘crime increases’ condition than his or her counterpart in the ‘crime decreases’ condition.

This outcome was not anticipated. But insofar as previous research on the ability to detect covariance has shown that confirmation bias can magnify the tendency of subjects to rely decisively, and mistakenly, on a heuristic strategy (Stanovich & West, 1998), this result can plausibly be viewed as suggesting the presence of a strong expectation among a large proportion of subjects of diverse political outlooks that the gun ban would be ineffective.

## Discussion

### *Summary of principal results*

The experiment that was the subject of this paper was designed to test two opposing accounts of conflict over decision-relevant science. The first – the science comprehension thesis (SCT) – attributes such conflicts to the limited capacity of the public to understand the significance of valid empirical evidence. The second – the identity-protective cognition thesis (ICT) – sees a particular recurring form of group conflict as *disabling* the capacities that individuals have to make sense of decision-relevant science: when policy-relevant facts become identified as symbols of membership in and loyalty to affinity groups that figure in important ways in individuals’ lives, they will be motivated to engage empirical evidence and other information in a manner that more reliably connects their beliefs to the positions that predominate in their particular groups than to the positions that are best supported by the evidence.

Study subjects were assigned to analyze the results of an experiment. Correctly interpreting the data required subjects to engage in a form of quantitative analysis – identifying covariance between an experimental treatment and outcomes – that is essential to valid causal inference but that many

people have difficulty performing reliably and accurately. Not surprisingly, we found that when the experiment was styled as one involving a skin rash treatment, the subjects' probability of identifying the most supported outcome was highly sensitive to subjects' numeracy, a capacity to understand and make proper use of quantitative information in reasoning tasks.

Also not surprisingly – given the growing literature on ideologically motivated reasoning – subjects' likelihood of correctly identifying the correct response varied in relation to the subjects' political outlooks when the experiment was styled as one involving a gun control ban. Subjects were more likely to correctly identify the result most supported by the data when doing so affirmed the position one would expect them to be politically predisposed to accept – that the ban decreased crime, in the case of more liberal subjects who identify with the Democratic Party; and that it increased crime, in the case of more conservative ones who identify with the Republican Party – than when the correct interpretation of the data threatened or disappointed their predispositions.

SCT, however, predicted that polarization among high-numeracy partisans would be lower than among low-numeracy ones in the gun ban conditions, consistent with the premise that political conflict over decision-relevant science is fed by defects in the capacity of ordinary members of the public to make sense of empirical evidence. The data *did not support* this prediction. On the contrary, numeracy *magnified* political polarization among high-numeracy partisans. This result was consistent with ICT.

More numerate individuals are benefitted by forming identity-congruent beliefs just as much as less numerate individuals are, and harmed just as much from forming identity-noncongruent beliefs. But more numerate individuals have a cognitive *ability* that lower-numeracy ones do not. ICT predicts that more numerate individuals will use that ability opportunistically in a manner geared to promoting their interests in forming and persisting in identity-protective beliefs.

The results in the experiment suggest that high-numeracy partisans did exactly that in the gun ban conditions. Where reliance on low-effort heuristic reasoning suggested an inference that was affirming of their political outlooks, high-numeracy partisans selected the answer that reflected that mode of information processing – even though it generated the wrong answer. But where reliance on low-effort heuristic processes suggested an inference that was threatening to their outlooks, high-numeracy partisans used the ability that they but not their low-numeracy counterparts possessed to make proper use of all the quantitative information presented in a manner that generated a *correct*, identity-affirming conclusion. This selectivity of their use of their greater capacity to draw inferences from quantitative information is what generated greater polarization among high-numeracy partisans than low-numeracy ones.

*Ideologically motivated cognition and dual process reasoning generally*

The ICT hypothesis corroborated by the experiment in this paper conceptualizes numeracy as a capacity associated with the disposition to engage in deliberate, effortful System 2 reasoning as applied to quantitative information. The results of the experiment thus help to deepen insight into the ongoing exploration of how ideologically motivated reasoning interacts with System 2 information processing generally.

As suggested, dual process reasoning theories typically posit two forms of information processing: a “fast, associative” one “based on low-effort heuristics,” and a “slow, rule based” one that relies on “high-effort systematic reasoning” (Chaiken & Trope, 1999, p. ix). Some researchers have assumed (not unreasonably) that ideologically motivated cognition – the tendency selectively to credit or discredit information in patterns that gratify one’s political or cultural predispositions – reflects over-reliance on the heuristic forms of information processing associated with heuristic-driven, System 1 of information processing (e.g. Sunstein, 2006; Westen *et al.*, 2006; Marx *et al.*, 2007; Weber & Stern, 2011; Lodge & Taber, 2013).

There is mounting evidence that this assumption is incorrect. It includes observational studies that demonstrate that science literacy, numeracy, and education (Hamilton, 2011; Hamilton *et al.*, 2012; Kahan *et al.*, 2012; Kahan, 2015) – all of which it is plausible to see as elements or outgrowths of the critical reasoning capacities associated with System 2 information processing – are associated with more, not less, political division of the kind one would expect if individuals were engaged in motivated reasoning.

Experimental evidence points in the same direction. Individuals who score higher on the cognitive reflection test (CRT), for example, have shown an even stronger tendency than ones who score lower to credit evidence selectively in patterns that affirm their political outlooks (Kahan, 2013). The evidence being assessed in that study was non-quantitative, but involved a degree of complexity that was likely to obscure its ideological implications from subjects who were inclined to engage the information in a casual or heuristic fashion. The greater polarization of subjects who scored highest on the CRT was consistent with the inference that individuals who are more disposed to engage systematically with information would be more likely to discern the political significance of it and would use their critical reasoning capacities selectively to affirm or reject it conditional on its congeniality to their political outlooks.

The experimental results we report in this paper display the same interaction between motivated cognition and System 2 information processing. The numeracy scale predicts how likely individuals are to resort to more systematic as opposed to heuristic engagement with *quantitative* information essential to

valid causal inference. The results in the gun ban conditions suggest that high-numeracy subjects made *use* of this System 2 reasoning capacity selectively in a pattern consistent with their motivation to form a politically congenial interpretation of the results of the gun ban experiment. This outcome is consistent with that of scholars who see both systematic (or System 2) and heuristic (System 1) reasoning as vulnerable to motivated cognition (Giner-Sorolla & Chaiken, 1997; Chen *et al.*, 1999; Cohen, 2003).

These findings also bear on whether ideologically motivated cognition is usefully described as a manifestation of ‘bounded rationality’. Cognitive biases associated with System 1 reasoning are typically characterized that way on the grounds that they result from an overreliance on heuristic patterns of information processing that reflect generally adaptive but still demonstrably inferior substitutes for the more effortful and more reliable type of information processing associated with System 2 reasoning (e.g. Jolls *et al.*, 1998; Kahneman, 2003).

We submit that a form of information processing cannot reliably be identified as ‘irrational’, ‘subrational’, ‘boundedly rational’, or the like independent of what an individual’s aims are in making use of information (Baron, 2008, p. 61). It is perfectly rational, *from an individual welfare perspective*, for individuals to engage with decision-relevant science in a manner that promotes culturally or politically congenial beliefs. What any individual member of the public thinks about the reality of climate change, the hazards of nuclear waste disposal, or the efficacy of gun control is too inconsequential to influence the risk that that person or anyone he or she cares about faces. Nevertheless, given what positions on these issues signify about a person’s defining commitments, forming a belief at odds with the one that predominates on it within important affinity groups of which such a person is a member could expose him or her to an array of highly unpleasant consequences (Kahan, 2012). Forms of information processing that reliably promote the stakes that individuals have in conveying their commitment to identity-defining groups can thus be viewed as manifesting what Anderson (1993) and others (Lessig, 1995; Akerlof & Kranton, 2000; Cohen, 2003; Hillman, 2010; Stanovich, 2013) have described as *expressive rationality*.

If ideologically motivated reasoning is expressively rational, then we should expect those individuals who display the highest reasoning capacities to be the ones most powerfully impelled to engage in it (Kahan *et al.*, 2012). This study now joins the rank of a growing list of others that fit this expectation and that thus supports the interpretation that ideologically motivated reasoning is *not* a form of bounded rationality, but instead a sign of how it becomes rational for otherwise intelligent people to use their critical faculties when they find themselves in the unenviable situation of having to choose between crediting the best available evidence or simply *being* who they are.

### Conclusion: protecting the ‘science communication environment’

To conclude that ideologically motivated reasoning is expressively rational obviously does *not* imply that it is socially or morally desirable (Lessig, 1995). Indeed, the implicit conflation of individual rationality and collective wellbeing has long been recognized to be a recipe for confusion, one that not only distorts inquiry into the mechanisms of individual decision-making, but also impedes the identification of social institutions that remove any conflict between those mechanisms and attainment of the public good (Olson, 1965). Accounts that misunderstand the expressive rationality of ideologically motivated cognition are unlikely to generate reliable insights into strategies for counteracting the particular threat that persistent political conflict over decision-relevant science poses to enlightened democratic policy-making.

Commentators who subscribe to what we have called the science comprehension thesis typically propose one of two courses of action. The first is to strengthen science education and the teaching of critical reasoning skills, in order better to equip the public for the cognitive demands of democratic citizenship in a society where technological risk is becoming an increasingly important focus of public policy-making (Miller & Pardo, 2000). The second is to dramatically shrink the scope of the public’s role in government by transferring responsibility for risk regulation and other forms of science-informed policy-making to politically insulated expert regulators (Breyer, 1993). This is the program advocated by commentators who believe that the public’s over-reliance on heuristic-driven forms of reasoning is too elemental to human psychology to be corrected by any form of education (Sunstein, 2005).

Because it rejects the empirical premise of the science comprehension thesis, the identity-protective cognition thesis takes issue with both of these prescriptions. The reason that citizens remain divided over risks in the face of compelling and widely accessible scientific evidence, this account suggests, is not that that they are insufficiently rational; it is that they are *too* rational in extracting from information on these issues the evidence that matters most for them in their everyday lives. In an environment in which positions on particular policy-relevant facts become widely understood as symbols of people’s membership in and loyalty to opposing cultural groups, it will promote people’s individual interests to attend to evidence about those facts in a manner that reliably conforms their beliefs to the ones that predominate in the groups they are members of. Indeed, the tendency to process information in this fashion will be strongest among individuals who display the reasoning capacities most strongly associated with science comprehension.

Thus, improving public understanding of science and propagating critical reasoning skills – while immensely important, both intrinsically and practically

(Dewey, 1910) – cannot be expected to dissipate persistent public conflict over decision-relevant science. Only removing the source of the motivation to process scientific evidence in an identity-protective fashion can. The conditions that generate symbolic associations between positions on risk and associated facts on the one hand and cultural identities on the other must be neutralized *in order to ensure that citizens make use of their capacity for science comprehension.*<sup>2</sup>

In a deliberative environment protected from the entanglement of cultural meanings and policy-relevant facts, moreover, there is little reason to assume that ordinary citizens will be unable to make an intelligent contribution to public policy-making. The amount of decision-relevant science that individuals reliably make use of in their everyday lives far exceeds what any of them (even scientists, particularly when acting outside of the domain of their particular specialty) are capable of understanding on an expert level. They are able to accomplish this feat because they are experts at something else: identifying who knows what about what (Keil, 2010), a form of rational processing of information that features consulting others whose basic outlooks individuals share and whose knowledge and insights they can therefore reliably gauge (Kahan *et al.*, 2010).

These normal and normally reliable processes of knowledge transmission break down when risk or associated facts are transformed (whether through strategic calculation or misadventure and accident) into divisive symbols of cultural identity. The solution to this problem is not – or certainly not *necessarily* – to divest citizens of the power to contribute to the formation of public policy. Rather, it is to adopt measures that effectively shield decision-relevant science from the influences that generate this reason-disabling state (Kahan *et al.*, 2006).

Just as individual wellbeing depends on the quality of the natural environment, so the collective welfare of democracy depends on the quality of a

<sup>2</sup> We would add, however, that we do not believe that the results of this or any other study we know of rule out the existence of cognitive dispositions that *do* effectively mitigate the tendency to display ideologically motivated reasoning. Research on the existence of such dispositions is ongoing and important (Baron, 1995; Lavine *et al.*, 2012). Existing research, however, suggests that the incidence of any such disposition in the general population is small and is distinct from the forms of critical reasoning disposition – those that are associated with constructs such as science literacy, cognitive reflection, and numeracy – that are otherwise indispensable to science comprehension. In addition, we submit that the best current understanding of the study of science communication indicates that the low incidence of this capacity, if it exists, is not the source of persistent conflict over decision-relevant science. Individuals who are endowed with perfectly ordinary capacities for comprehending science can be expected reliably to use them to identify the best available scientific evidence *so long as* risks and associated policy-relevant facts are shielded from antagonistic cultural meanings.

*science communication environment* that is hospitable to the exercise of the ordinarily reliable reasoning faculties that ordinary citizens use to discern what is collectively known. Identifying strategies for protecting the science communication environment from antagonistic cultural meanings – and for decontaminating it when such protective measures fail – is the most critical contribution that decision science can make to the practice of democratic government.

## Acknowledgements

We are indebted to Angie Johnston, Scott Johnson, Matthew Fisher, Andrew Meyer, and Maggie Wittlin for discussion and comments.

## Funding

Research for this paper was funded by the National Science Foundation, Grant SES 0922714, and the Cultural Cognition Lab at Yale Law School.

## References

- Ai, C. and E. C. Norton (2003), 'Interaction terms in logit and probit models', *Economics Letters*, 80(1): 123–129.
- Akerlof, G. A. and R. E. Kranton (2000), 'Economics and identity', *The Quarterly Journal of Economics*, 115: 715–753.
- Anderson, E. (1993), *Value in ethics and economics*, Cambridge, Mass: Harvard University Press.
- Baron, J. (1995), 'Myside bias in thinking about abortion', *Thinking & Reasoning*, 1: 221–235.
- Baron, J. (2008), *Thinking and deciding*, New York: Cambridge University Press.
- Breyer, S. G. (1993), *Breaking the Vicious Circle: Toward Effective Risk Regulation*, Cambridge, MA: Harvard University Press.
- Chaiken, S. and Y. Trope (1999), *Dual-process theories in social psychology*, New York, NY: Guilford Press.
- Chen, S., K. Duckworth and S. Chaiken (1999), 'Motivated Heuristic and Systematic Processing', *Psychological Inquiry*, 10: 44–49.
- Cohen, G. L. (2003), 'Party over Policy: The Dominating Impact of Group Influence on Political Beliefs', *Journal of Personality and Social Psychology*, 85: 808–822.
- Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences*, Hillsdale, NJ: Lawrence Erlbaum Assocs.
- Cohen, J., P. Cohen, S. G. West and L. S. Aiken (2003), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd Edn., Mahwah, N.J.: L. Erlbaum Associates.
- Dawson, E., T. Gilovich and D. T. Regan (2002), 'Motivated Reasoning and Performance on the was on Selection Task', *Personality and Social Psychology Bulletin*, 28: 1379–1387.
- Dawson, E., T. Gilovich and D. T. Regan (2000), Motivated Reasoning and Susceptibility to the "Cell A" Bias. Unpublished manuscript.
- Dewey, J. (1910), 'Science as Subject-Matter and as Method', *Science*, 31: 121–127.

- Gastil, J., D. Braman, D. Kahan and P. Slovic (2011), 'The Cultural Orientation of Mass Political Opinion', *PS: Political Science & Politics*, **44**: 711–714.
- Giner-Sorolla, R. and S. Chaiken (1997), 'Selective Use of Heuristic and Systematic Processing Under Defense Motivation', *Personality and Social Psychology Bulletin*, **23**: 84–97.
- Greene, W. (2010), 'Testing hypotheses about interaction terms in nonlinear models', *Economics Letters*, **107**(2): 291–296.
- Hamilton, L. C. (2011), 'Education, politics and opinions about climate change evidence for interaction effects', *Climatic Change*, **104**: 231–242.
- Hamilton, L. C., J. Matthew and A. Schaefer (2012), 'Public knowledge and concern about polar-region warming', *Polar Geography*, **35**: 155–168.
- Hillman, A. L. (2010), 'Expressive behavior in economics and politics', *European Journal of Political Economy*, **26**: 403–418.
- Jolls, C., C. R. Sunstein and R. Thaler (1998), 'A Behavioral Approach to Law & Economics', *Stanford Law Review*, **50**: 1471.
- Jost, J. T., E. P. Hennes, and H. Lavine (2014), "'Hot" political cognition: Its self-, group-, and system-serving purposes', in D. E. Carlson (ed.), *Oxford handbook of social cognition*, New York: Oxford University Press.
- Kahan, D. (2010), 'Fixing the Communications Failure', *Nature*, **463**: 296–297.
- Kahan, D. M. (2015), 'Climate-Science Communication and the Measurement Problem', *Advances in Political Psychology*, **36**: 1–43.
- Kahan, D. (2012), 'Why we are poles apart on climate change', *Nature*, **488**: 255.
- Kahan, D. M., K. H. Jamieson, A. Landrum and K. Winneg (2017), 'Culturally antagonistic memes and the Zika virus: an experimental test', *Journal of Risk Research*, **20**(1): 1–40.
- Kahan, D., D. Braman, G. Cohen, J. Gastil and P. Slovic (2010), 'Who Fears the HPV Vaccine, Who Doesn't, and Why? An Experimental Study of the Mechanisms of Cultural Cognition', *Law and Human Behavior*, **34**: 501–516.
- Kahan, D. M. (2013), 'Ideology, Motivated Reasoning, and Cognitive Reflection', *Judgment and Decision Making*, **8**: 407–424.
- Kahan, D. M. (2012), 'Cultural Cognition as a Conception of the Cultural Theory of Risk', in R. Hillerbrand, P. Sandin, S. Roeser & M. Peterson (eds.) *Handbook of Risk Theory: Epistemology, Decision Theory, Ethics and Social Implications of Risk*, London: Springer Ltd, 725–760.
- Kahan, D. M. (2011), 'The Supreme Court 2010 Term—Foreword: Neutral Principles, Motivated Cognition, and Some Problems for Constitutional Law', *Harvard Law Review*, **126**: 1.
- Kahan, D. M., D. Braman, J. Gastil, P. Slovic and C. K. Mertz (2007), 'Culture and Identity-Protective Cognition: Explaining the White-Male Effect in Risk Perception', *Journal of Empirical Legal Studies*, **4**: 465–505.
- Kahan, D. M., D. Braman, J. Monahan, L. Callahan and E. Peters (2010), 'Cultural Cognition and Public Policy: The Case of Outpatient Commitment Laws', *Law and Human Behavior*, **34**: 118–140.
- Kahan, D. M., H. Jenkins-Smith and D. Braman (2011), 'Cultural Cognition of Scientific Consensus', *Journal of Risk Research*, **14**: 147–174.
- Kahan, D. M., E. Peters, M. Wittlin, P. Slovic, L. L. Ouellette, D. Braman and G. Mandel (2012), 'The polarizing impact of science literacy and numeracy on perceived climate change risks', *Nature Climate Change*, **2**: 732–735.
- Kahan, D. M., P. Slovic, D. Braman and J. Gastil (2006), 'Fear of Democracy: A Cultural Evaluation of Sunstein on Risk', *Harvard Law Review*, **119**: 1071–1109.
- Kahneman, D. (2003), 'Maps of Bounded Rationality: Psychology for Behavioral Economics', *American Economic Review*, **93**: 1449–1475.
- Keil, F. C. (2010), 'The Feasibility of Folk Science', *Cognitive Science*, **34**: 826–862.

- King, G., J. Honaker, A. Joseph and K. Scheve (2001), 'Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation', *American Political Science Review*, 95: 49–69.
- King, G., M. Tom and J. Wittenberg (2000), 'Making the Most of Statistical Analyses: Improving Interpretation and Presentation', *American Journal of Political Science*, 44: 347–361.
- Kunda, Z. (1990), 'The Case for Motivated Reasoning', *Psychological Bulletin*, 108: 480–498.
- Lavine, H., C. D. Johnston and M. R. Steenbergen (2012), *The ambivalent partisan: how critical loyalty promotes democracy*, New York, NY: Oxford University Press.
- Lessig, L. (1995), 'The Regulation of Social Meaning', *University of Chicago Law Review*, 62: 943–1045.
- Liberali, J. M., V. F. Reyna, S. Furlan, L. M. Stein and S. T. Pardo (2012), 'Individual Differences in Numeracy and Cognitive Reflection, with Implications for Biases and Fallacies in Probability Judgment', *Journal of Behavioral Decision Making*, 25: 361–381.
- Lodge, M. and C. S. Taber (2013), *The rationalizing voter*, Cambridge, New York: Cambridge University Press.
- Loewenstein, G. F., E. U. Weber, C. K. Hsee and N. Welch (2001), 'Risk as Feelings', *Psychological Bulletin*, 127, 267–287.
- Marx, S. M., E. U. Weber, B. S. Orlove, A. Leiserowitz, D. H. Krantz, C. Roncoli and J. Phillips (2007), 'Communication and mental processes: Experiential and analytic processing of uncertain climate information', *Global Environmental Change – Human and Policy Dimensions*, 17 (1): 47–58.
- McCaffrey, M. S. and S. M. Buhr (2008), 'Clarifying climate confusion: addressing systemic holes, cognitive gaps, and misconceptions through climate literacy', *Physical Geography*, 29(6): 512–528.
- McCright, A. M. and R. E. Dunlap (2013), 'Bringing ideology in: the conservative white male effect on worry about environmental problems in the USA', *Journal of Risk Research*, 16: 211–226.
- Miller, J. D. and R. Pardo (2000), in M. Dierkes & C.v. Grote (eds.) *Between understanding and trust: The public, science and technology*, Australia: Harwood Academic, 131–156.
- Olson, M. (1965), *The logic of collective action; public goods and the theory of groups*, Cambridge, MA: Harvard University Press.
- Pampel, F. C. (2000), *Logistic regression: a primer*, Thousand Oaks, CA: Sage Publications.
- Peters, E., D. Västfjäll, P. Slovic, C. K. Mert, K. Maocco and S. Dickert (2006), 'Numeracy and Decision Making', *Psychological Science*, 17: 407–413.
- Powers, E. A. (2005), 'Interpreting logit regressions with interaction terms: an application to the management turnover literature', *Journal of Corporate Finance*, 11(3): 504–522.
- Rosenau, J. (2012), 'Science denial: a guide for scientists', *Trends in microbiology*, 20: 567–569.
- Rubin, D. (2004), *Multiple imputation for nonresponse in surveys*, Hoboken, NJ: Wiley-Interscience.
- Rubin, D. B. (1987), *Multiple imputation for nonresponse in surveys*, New York, NY: Wiley.
- Sherman, D. K. and G. L. Cohen (2006), *Advances in Experimental Social Psychology*, Vol. 38, Academic Press, 183–242.
- Smith, E. R. (2000), in H. T. Reis and C. M. Judd (eds.) *Handbook of Research Methods in Social and Personality Psychology*, Cambridge: Cambridge University Press, 17–39.
- Stanovich, K. E. and R. F. West (1998), 'Who uses base rates and P (DIH)? An analysis of individual differences', *Memory & Cognition*, 26: 161–179.
- Stanovich, K. E. and R. F. West (2000), 'Individual differences in reasoning: Implications for the rationality debate?' *Behavioral and Brain Sciences*, 23(5): 645–665.
- Stanovich, K. E. (2009), *What Intelligence Tests Miss: The Psychology of Rational Thought*, New Haven, CT: Yale University Press.

- Stanovich, K. E. (2013), 'Why humans are (sometimes) less rational than other animals: Cognitive complexity and the axioms of rational choice', *Thinking & Reasoning*, 19: 1–26.
- Stanovich, K. E., R. F. West and M. E. Toplak (2011), in J. S. B. T. Evans, K. I. Manktelow, D. E. Over and S. Elqayam eds.) *The science of reason: a Festschrift for Jonathan St. B.T Evans*, New York, NY: Psychology Press, 355–396.
- Streiner, D. L. (2003), 'Unicorns Do Exist: A Tutorial on "Proving" the Null Hypothesis', *Canadian Journal of Psychiatry*, 48: 756–761.
- Sunstein, C. R. (2005), *Laws of Fear: Beyond the Precautionary Principle*, Cambridge, UK ; New York, NY: Cambridge University Press.
- Sunstein, C. R. (2006), 'Misfearing: A Reply', *Harvard Law Review*, 119: 1110–1125.
- Sunstein, C. R. (2007), 'On the Divergent American Reactions to Terrorism and Climate Change', *Columbia Law Review*, 107: 503–557.
- Sunstein, C. R. (2003), 'What's Available? Social Influences and Behavioral Economics', *Northwestern University Law Review*, 97.
- Toplak, M., R. West and K. Stanovich (2011), 'The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks', *Memory & Cognition*, 39(7): 1275–1289.
- Wasserman, E. A., W. W. Dornier and S. F. Kao (1990), 'Contributions of Specific Cell Information to Judgments of Intervent Contingency', *Journal of Experimental Psychology, Learning, Memory, and Cognition*, 16(3): 509–521.
- Watts, D. J. (2011), *Everything is Obvious: Once You Know the Answer: How Common Sense Fails*, Atlantic Books.
- Weber, E. (2006), 'Experience-Based and Description-Based Perceptions of Long-Term Risk: Why Global Warming Does Not Scare Us (yet)', *Climatic Change*, 77: 103–120.
- Weber, E. U. and P. C. Stern (2011), 'Public Understanding of Climate Change in the United States', *American Psychologist*, 66: 315–328.
- Weller, J. A., N. F. Dieckmann, M. Tusler, C. Mertz, W. J. Burns and E. Peters (2012), 'Development and testing of an abbreviated numeracy scale: A Rasch analysis approach', *Journal of Behavioral Decision Making*, 26: 198–212.
- Westen, D., P. Blagov, K. Harenski, C. Kilts and S. Hamann (2006), 'Neural Bases of Motivated Reasoning: An fMRI Study of Emotional Constraints on Partisan Political Judgment in the 2004 U.S. Presidential Election', *Journal of Cognitive Neuroscience*, 18: 1947–1958.

## Statistical appendix

This appendix furnishes information on the multivariate regression model used to test the study hypotheses.

### Statistical power and missing data

Because it pits opposing ICT and SCT hypotheses against one another, the study design contemplated the possibility of drawing inferences from the absence of an effect (decreased or increased polarization among high-numeracy subjects in the gun ban conditions). The strength of inferences drawn from ‘null’ findings depends heavily on statistical power. The large size of the sample furnished adequate power to detect even small effect sizes (e.g.  $r = 0.10$ ) with a probability well over 0.80 at  $p \leq 0.05$  (Cohen, 1988). As a result, findings of non-significance could be equated with a lack of effect and with a low risk of a Type II error (Streiner, 2003). To ensure full exploitation of the power associated with the large sample size, missing data were replaced by multiple imputation (King *et al.*, 2001; Rubin, 2004).

### The model

The regression model is a logistic regression model, which is appropriate for assessing the probability of a binary outcome (Cohen *et al.*, 2003) – here, correctly responding to the covariance detection problem. The goal of the model is to determine whether and to what extent numeracy influences the probability that a subject will answer the problem correctly conditional on his or her level of numeracy, his or her political outlook, and the experimental condition to which he or she was assigned. Model variables thus included ones for those predictors, plus cross-product terms suited for assessing their interactions with one another (Cohen *et al.*, 2003, p. 494).

We entered the relevant variables in stages. The aim of the initial two stages was to determine what portion of the variance in the probability of answering correctly could be attributed solely to numeracy and the experimental assignments independently of subjects’ political outlooks. The scatter plots (Figures 4–6) suggested that the impact of numeracy on subjects’ ability to identify the correct response in the covariance-detection problem was not linear, but rather was triggered at a threshold between the 75th and 90th percentiles (five and seven answers correct) on the numeracy scale. The scatter plots also suggested that the impact of numeracy on improving the performance of subjects was uneven across the skin treatment and gun control conditions, a result consistent with the hypothesis that ideologically motivated reasoning would

inhibit effortful processing of information in conditions in which heuristic strategies for assessing the data affirmed subjects' political outlooks. Consistent with these patterns, we found that a quadratic model – one that assumed that numeracy's impact on identifying the data would be curvilinear and vary across each condition (Pampel, 2000, p. 20) – fit the data better than a model that assumed that numeracy's contribution would be linear and invariant across each condition (Table A1, Stages 1–2).

After identifying the best-fitting model based only on subject numeracy and political outlooks, we added terms designed to test whether the impact of numeracy on subject performance was conditional on their political outlooks. Two-way interaction terms that reflected the impact of political outlooks in each condition and three-way ones that reflected how the impact of numeracy varied in each condition in relation to subjects' political outlooks also improved the fit of the model (Table A1, Stage 3). Adding three-way interaction terms in order to reflect the impact of numeracy<sup>2</sup> and political outlook did not improve model fit.

The model cannot be easily explained based on the regression output alone. Models containing multiple two- and three-way interaction terms frequently defy straightforward interpretation, particularly in logistic regression, where the impact that such terms have cannot be determined simply by examining their sign and significance (Ai & Norton, 2003). The only reliable way to determine the effect of such predictors is to calculate their marginal impact on the outcome variable as one varies the values of the predictors in a manner that corresponds to the hypotheses of interest (Powers, 2005; Greene, 2010). Such an analysis appears in the 'Multivariate analyses' section of the text.

**Table A1.** Multivariate regression analysis ( $N = 1111$ ). Outcome variable is ‘correct’, a binary variable coded ‘1’ for correctly interpreting the data and ‘0’ for incorrectly interpreting them. Predictor estimates are logit coefficients with the  $z$ -score test statistic indicated parenthetically. ‘Rash\_decreases’, ‘rash\_increases’, and ‘crime\_increases’ are dummy variables reflecting experimental condition assignment (0 = unassigned, 1 = assigned); the reference assignment is to ‘crime decreases’. Both Conserv\_Repub and numeracy are centered at ‘0’ for ease of interpretation. Bold typeface indicates that a predictor coefficient, model  $F$ -test, or incremental change in model  $F$ -test is significant at  $p < 0.05$ .

	Stage 1	Stage 2	Stage 3
Numeracy	0.11 (0.78)	0.01 (0.07)	-0.01 (-0.05)
rash_decreases	0.36 (1.91)	0.36 (1.42)	0.40 (1.57)
rash_increases	-0.12 (-0.63)	0.01 (0.02)	0.06 (0.22)
crime_increases	<b>0.64</b> (3.46)	<b>1.03</b> (4.06)	<b>1.07</b> (4.02)
numeracy × rash_decreases	0.19 (0.95)	0.21 (1.00)	0.23 (1.05)
numeracy × rash_increases	<b>0.50</b> (2.17)	<b>0.51</b> (2.29)	0.55 (2.29)
numeracy × crime_increases	0.26 (1.26)	0.38 (1.79)	<b>0.46</b> (2.01)
numeracy <sup>2</sup>		<b>0.32</b> (2.77)	<b>0.31</b> (2.46)
numeracy <sup>2</sup> × rash_decreases		0.01 (0.08)	0.02 (0.14)
numeracy <sup>2</sup> × rash_increases		-0.09 (-0.49)	-0.07 (-0.39)
numeracy <sup>2</sup> × crime_increases		<b>-0.39</b> (-2.39)	-0.31 (-1.75)
Conserv_Repub			<b>-0.64</b> (-3.95)
Conserv_Repub × rash_decreases			0.56 (2.64)
Conserv_Repub × rash_increases			<b>0.63</b> (2.82)
Conserv_Repub × crime_increases			<b>1.28</b> (6.02)
numeracy × Conserv_Repub			-0.33 (-1.89)
numeracy × Conserv_Repub × rash_decreases			0.33 (1.40)
numeracy × Conserv_Repub × rash_increases			0.26 (1.08)
numeracy × Conserv_Repub × crime_increases			0.54 (2.17)
constant	<b>-0.58</b> (-4.06)	<b>-0.91</b> (-4.64)	<b>-0.96</b> (-4.70)
$F$ -test	(7, 1102) <b>6.46</b>	(11, 1098) <b>5.78</b>	(19, 1090) <b>5.06</b>
$\Delta F$ -test		(4, 1105) <b>5.20</b>	(8, 1101) <b>5.21</b>