THINKING & REASONING
2019, VOL. 25, NO. 4, 477–505
https://doi.org/10.1080/13546783.2018.1548379

Routledge
Taylor & Francis Group

Does "putting on your thinking cap" reduce myside bias in evaluation of scientific evidence?

Caitlin Drummond[a,b] and Baruch Fischhoff[c]

[a]Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA, USA; [b]Erb Institute for Global Sustainable Enterprise, University of Michigan, Ann Arbor, MI, USA; [c]Department of Engineering and Public Policy, Institute for Politics and Strategy, Carnegie Mellon University, Pittsburgh, PA, USA

ABSTRACT

The desire to maintain current beliefs can lead individuals to evaluate contrary evidence more critically than consistent evidence. We test whether priming individuals' scientific reasoning skills reduces this often-observed *myside bias,* when people evaluate scientific evidence about which they have prior positions. We conducted three experiments in which participants read a news-style article about a study that either supported or opposed their attitudes regarding the Affordable Care Act. We manipulated whether participants completed a test posing scientific reasoning problems before or after reading the article and evaluating the evidence that it reported. Consistent with previous research, we found that participants were biased in favor of evidence consistent with their prior attitudes regarding the Affordable Care Act. Priming individuals' scientific reasoning skills reduced myside bias only when accompanied by direct instructions to apply those skills to the task at hand. We discuss the processes contributing to biased evaluation of scientific evidence.

ARTICLE HISTORY Received 9 January 2018; Accepted 6 November 2018

KEYWORDS Judgment; reasoning; scientific communication; myside bias

Introduction

Americans have deeply held and diverse views on many issues where scientific evidence could play a central role, from the impact of international trade agreements to the existence of climate change. Their views are sometimes at odds with the scientific evidence (Funk & Rainie, 2015). In response, many observers have lamented a lack of scientific literacy (see Bauer, 2009; Bauer, Allum & Miller, 2007; Miller, 2004). However, such "deficit model" interpretations cannot readily account for the finding that,

CONTACT Caitlin Drummond drummonc@umich.edu

once issues have become controversial, polarisation is often greater among individuals with more general or science education and with higher scores on science literacy tests (Drummond & Fischhoff, 2017a; McCright, Marquart-Pyatt, Shwom, Brechin, & Allen, 2016). One possible explanation is that people with greater cognitive skills and knowledge are more adept at motivated cognition, finding ways to maintain currently favored beliefs through biased evaluation, interpretation, and recall (Kunda, 1990). Here, we ask how priming reasoning capabilities affects evaluations of scientific evidence that supports or contradicts favored beliefs.

The fullest accounts of such motivated cognition are found in studies of myside bias, whereby individuals tend to evaluate evidence consistent with their prior beliefs more positively than belief-inconsistent evidence (Toplak & Stanovich, 2003). For example, studies have found that people rate evidence supporting their beliefs as more convincing and of higher quality (Lord, Ross & Lepper, 1979), generate more arguments favoring their positions than opposing them (Toplak & Stanovich, 2003), counter-argue inconsistent information more than consistent information (Jain & Maheswaran, 2000), and require less information to reach desired conclusions (Ditto & Lopez, 1992). In contrast to the more general confirmation bias (Fischhoff & Beyth-Marom, 1983; Nickerson, 1998), which refers to seeking and interpreting information in ways that are biased in favor of existing beliefs, expectations, or hypotheses (e.g., Mynatt, Doherty & Tweney, 1977), myside bias refers to generating and evaluating evidence in ways that favor existing positions that one wants to be true (Stanovich, West & Toplak, 2013; Toplak & Stanovich, 2003). A strong version of myside bias requires ignoring unwelcome evidence (e.g., Wolfe & Britt, 2008; Wolfe, Britt & Butler, 2009). We will adopt the more general version, which encompasses any of these processes.

We examine how three ways of priming analytical information processing affect the evaluation of scientific evidence regarding a topic that is controversial for many people: the Affordable Care Act (ACA, or Obamacare). We consider the effects of priming on individuals' understanding and evaluations of the evidence, relative to their position on the ACA. In Study 1, the priming manipulation asks participants to complete a validated measure of individual differences of relevant abilities, the *Scientific Reasoning Scale* (SRS) (Drummond & Fischhoff, 2017b), which asks participants to solve 11 short scientific scenario problems. Study 2 compares the effects of taking the SRS with the effects of taking a numeracy test (Lipkus, Samsa & Rimer, 2001), designed to get people thinking, but not specifically about the quality of evidence. Study 3 strengthens the SRS manipulation by explicitly asking participants to apply those analytical thinking skills to evaluating the ACA-related evidence.

All participants completed the SRS, thereby allowing us to assess the extent to which the effects of priming depend on the extent to which individuals possess the scientific reasoning skills that priming could evoke. Each of the 11 SRS items presents a short reasoning problem, followed by a true/false statement that participants must evaluate. Previous studies have found considerable variance in SRS performance, with people who score higher on it also scoring higher on tests of numeracy, scientific knowledge, and cognitive reflection (Drummond & Fischhoff, 2017b). People with higher SRS scores also had more success in using the scientific information in a drug facts box and report more years of education. No gender differences have been observed.

We predicted that completing the SRS would prime participants to draw on their latent scientific reasoning skills when completing subsequent tasks involving the evaluation of scientific evidence. Unlike most priming manipulations, which seek to activate latent knowledge structures without conscious awareness (Bargh, 2006), our manipulation sought to evoke thought processes, namely, the higher-order reasoning skills relevant to evaluating scientific evidence. In effect, we treat the SRS as a warm-up task that implicitly asks participants to "put on their thinking caps," by activating their scientific reasoning skills.

With respect to the effects of the SRS manipulations on myside bias, we had competing predictions. On one hand, taking the SRS could reduce bias by increasing the availability of participants' analytical reasoning skills. Dual-process theories of cognition involve the interplay of two distinct ways of thinking: rapid, automatic, and associative thinking, and higher-order reasoning, which requires greater working memory and hypothetical thinking (Evans, 2008; Evans & Stanovich, 2013; Kahneman, 2011). Scientific evidence related to personally relevant topics might evoke affective and associative processes that swamp analytical reasoning, thereby increasing myside bias. Taking the SRS might provide a counterweight. Attridge & Inglis (2015) offered such an account for their finding that completing a cognitively difficult prior task (Raven's matrices) improved performance on the Cognitive Reflection Test (Frederick, 2005).

On the other hand, the SRS manipulation might prompt participants to use their latent scientific reasoning skills in the service of biased evidence evaluation. As Kunda (1990) noted in her seminal account of motivated reasoning, "people can process information in depth and be differentially sensitive to its strengths and weaknesses and yet be biased at the same time" (p. 490). Thus, priming analytical scientific thinking may lead to deeper motivated reasoning about the evidence, a finding that would be consistent with the finding that individuals with more general and science education and higher scores on science literacy tests often hold more polarised

views on controversial science topics (Drummond & Fischhoff, 2017a; McCright et al., 2016).

Plausibly, the strength of these effects, in either direction, will depend upon the extent to which participants have scientific reasoning skills that could be primed – for which the SRS serves as a measure. Thus, while our primary concern was the effects of priming scientific reasoning on evidence evaluation, we also conducted exploratory analyses of the extent to which participants with greater scientific reasoning ability were more (or less) biased and more (or less) affected by the experimental manipulation.

The three studies presented here are as follows: Study 1 tests the effects of the SRS priming manipulation. Study 2 provides a direct replication of the SRS priming manipulation from Study 1, as well as a comparison with an analogous numeracy priming manipulation, which we expected to be less effective, by virtue of not evoking specifically relevant skills. Study 3 sought to strengthen the effect of the SRS manipulation, by explicitly relating the test to the subsequent evidence-evaluation task. Studies 2 and 3 further tested the effects of the SRS manipulation by assessing whether people who received the SRS prime were more likely to correctly answer a technical question about the scientific evidence.

Study 1

Study 1 examines how completing a scientific reasoning task affects participants' evaluations of scientific evidence that favors or opposes their existing attitudes about the Patient Protection and Affordable Care Act (ACA), a United States federal statute signed into law by President Barack Obama on March 23, 2010. The ACA expanded health insurance coverage for Americans in a variety of ways, including expanding eligibility for government-backed health insurance programs for low-income individuals, developing marketplaces in which individuals could purchase health insurance, and establishing an individual mandate that requires most individuals to have a minimum level of health insurance or pay a fine. The ACA was politically controversial, particularly amongst conservative voters.

In a 2 x 2 between-subjects design, we assigned participants to (a) complete the Scientific Reasoning Scale (SRS; Drummond & Fischhoff, 2017b), before or after reading a news article about the ACA (priming vs. control conditions, respectively); and (b) read an article about a study that found positive effects of the ACA or about a study that failed to find positive effects (inconsistent vs. consistent evidence conditions, depending on participants' attitudes toward the ACA). The two studies used similar methodologies. Although the news articles were written for the experiment, the

studies that they summarised were real. The full text of the news articles appears in the Appendix, Table A1.

Method

Recruiting

We recruited participants using Qualtrics' Online Sample service, which provides participants for surveys in partnership with online panel providers. In order to recruit roughly equal numbers of Affordable Care Act supporters and opponents, we first asked potential participants to answer an unpaid five-item screening test asking about contemporary political topics, promising that some would be invited to take a full survey, based upon their answers. One item stated, "The Affordable Care Act (ACA) changed the United States healthcare system. Some aspects of the ACA were controversial, including the creation of marketplaces where Americans can shop for insurance plans. The ACA has been challenged in the Supreme Court." Participants were then asked, "Do you think the Affordable Care Act should be repealed?" on a five-point scale (1 = Yes, definitely; 5 = No, definitely not). Those who answered "Yes, definitely" or "Yes, probably" were treated as Affordable Care Act opponents; those who answered "No, probably not" or "No, definitely not" were treated as Affordable Care Act supporters. Those who answered, "I'm not sure" were not invited to take the full survey. The other four screening items dealt with unrelated topics: the minimum wage, global warming, immigration, and the sale of firearms.

We set a target quota of roughly 300 supporters and 300 opponents, the largest possible sample size under the terms of the first author's Qualtrics Behavioral Research Grant. An estimated 150 participants per cell provides 80% power to detect a small effect size ($d = 0.11$) for the interaction between Priming Condition (SRS prime vs. Control) and Evidence Condition (Inconsistent vs. Consistent) for the planned two-way ANOVA.

We recruited an initial sample of 1017 participants. We excluded 121 who were neither supporters nor opponents of the Affordable Care Act and 14 who did not give informed consent. Among the remaining 882, Affordable Care Act opponents were more common than supporters. As a result, we turned away 234 opponents, while continuing to recruit supporters, until we reached our target. We oversampled roughly 25 in each group, anticipating attrition from participants not answering two data quality assurance questions correctly (Meade & Craig, 2012). Of 648 initial participants, 42 were excluded on those grounds. We also excluded one participant whose browser was configured to prevent our survey software from collecting response time information. The final sample had 605 participants,

with 303 Affordable Care Act opponents and 302 Affordable Care Act supporters.

Participants

We asked Qualtrics to recruit American adults to take our survey. Based on self-reports, 67% of the participants in the final sample were female, with a mean reported age of 47 ($SD = 13$). One percent reported not graduating high school, 25% attaining a high school degree or GED, 38% completing at least some college or an associate's degree, and 36% having a bachelor's degree or higher. Forty-one percent reported being politically moderate, 23% liberal or very liberal, and 36% conservative or very conservative.

Experimental design and measures

The experiment had a 2×2 between-subjects design. Data collection took place online and remotely, via a survey hosted on the Qualtrics platform.

Evidence condition. All participants read a news-style article, written by the authors, summarising a recent study on the effects of the Affordable Care Act. Participants were randomly selected to receive one of two versions of the article, describing evidence that was either consistent or inconsistent with their attitudes toward the Affordable Care Act. Table A1 in the Appendix has the full text of the articles. The supporting article described a study (Sommers, Buchmueller, Decker, Carey & Kronick, 2013) that found an increase in access to health care for 18- to -25 year-olds, interpreted as reflecting a provision that allowed young adults to stay on their parents' health insurance plans until age 26. The non-supporting article described a study (Kotagal, Carle, Kessler & Flum, 2014) that used the same methodology (difference in differences) and a similar dataset, but found no increase in access to health care for young adults. The two stimulus articles were written to be identical except for the results of the study that each reported. Before reading their article, participants were told that it reported a recent scientific study. After the survey, they were debriefed about the purpose of our research and given links to the abstracts of both ACA studies.

Priming condition. Participants were randomly assigned to the SRS prime or Control condition. In the SRS prime condition, participants completed the Scientific Reasoning Scale (SRS, Drummond & Fischhoff, 2017b) before reading the news article and answering questions about it. In the Control condition, participants first read the article and answered questions about it, and then completed the SRS. In this sample, the scale had a Cronbach's alpha of 0.62, lower than previously reported values (Cronbach's

alpha of 0.70, Drummond & Fischhoff, 2017b). Table A2 reports summary statistics.

Dependent measures. Immediately after reading their article, all participants rated their agreement with six statements about it and the evidence that it presented, on a seven-point Likert scale (1 = strongly disagree, 7 = strongly agree): The article's description of the scientific study was clear; The scientific evidence was convincing; The study's conclusions made sense; The way the study was conducted made sense; The study was conducted carefully; This study reflects the true impact that the ACA has had on healthcare for young adults. As responses to these six items were closely related (Cronbach's $\alpha = 0.91$), we formed an evidence evaluation index variable, equal to their average, which we treat as measuring participants' subjective evaluations of the evidence's quality.

Additional measures. In the final section of the survey, all participants completed two scales testing exploratory research questions unrelated to the present study and not reported here, and questions about four covariates: gender, age, highest level of education, and political beliefs (on a 5-point scale from 1 = very liberal to 5 = very conservative). Table A2 reports summary statistics for these measures.

Results

Assignment to condition

For 288 participants, the article presented evidence consistent with their ACA attitudes; for 317, it presented inconsistent evidence. Using *t*-tests, we found that SRS scores, age, and education were similar for participants who read consistent and inconsistent evidence (all *p*'s > 0.25). Those who read consistent evidence were more conservative politically ($M = 3.29$ vs. $M = 3.09$, $t(603) = 2.10$, $p = 0.036$) and more likely to be male (38% vs. 29%, $t(603) = 2.38$, $p = 0.017$). We found no differences between the SRS prime ($N = 300$) and Control ($N = 305$) conditions in participants' SRS score, gender, age, education, or political conservatism (all *p*'s > 0.10).

Myside bias

Table 1 presents group means on the evidence-evaluation index for the four conditions. These scores were first analysed using a 2 (SRS prime or Control condition) $\times$ 2 (Consistent or Inconsistent Evidence condition) between-subjects ANOVA, with no covariates. It revealed no interaction between Priming condition and Evidence condition, $F(1, 601) = 0.37$, MSE $=1.63$. We reran the model without the interaction term and observed significant main effects for both Priming condition, $F(1, 602) = 8.13$, MSE

Table 1. Group means for evaluations of the evidence in Study 1.

	Inconsistent evidence			Consistent evidence		
Priming condition	*M*	*SD*	*n*	*M*	*SD*	*n*
SRS prime	4.30	1.29	158	4.61	1.26	142
Control	4.66	1.28	159	4.84	1.28	146

Note: Evaluations of the evidence (index measure) reported on a seven-point scale, with higher scores reflecting more positive ratings. SRS = Scientific Reasoning Scale.

=1.63, $p = 0.005$, and Evidence condition, $F(1, 602) = 5.82$, MSE =1.63, $p = 0.016$. As predicted, participants rated evidence consistent with their ACA attitude more highly than inconsistent evidence, consistent with myside bias. Additionally, participants in the SRS prime condition rated evidence more critically than did those in the Control condition, consistent with the priming manipulation having evoked latent skills. The sizes of the two main effects were similar, Cohen's $d = 0.20$ and 0.23, respectively. Table 2 reports linear regressions indicating that these results are robust to the inclusion of covariates (Table 2, Models 1 and 2).

Exploratory analyses: Myside bias and scientific reasoning

We next tested whether participants with greater scientific reasoning ability were more (or less) biased and affected by the manipulations. Specifically, in exploratory analyses, we assessed the extent to which myside bias, and the effects of the SRS prime, varied by SRS score. Table 2 shows that individuals with higher SRS scores rated the evidence more negatively. We found no significant interaction between Evidence condition and SRS, indicating that myside bias did not vary with SRS score (Table 2, Model 3). We also found no significant three-way interaction between SRS Priming condition, Evidence condition, and SRS score, indicating that the impact of the SRS Priming condition on myside bias did not differ by SRS score (Table 2, Model 4).

Discussion

Study 1 examined the effects of administering a priming task intended to evoke latent scientific reasoning skills. We tested two competing predictions regarding its effects on myside bias: it could reduce bias, by prompting participants to think harder about the evidence, or it could increase bias, by prompting them to use their reasoning skills to support their prior attitudes.

We observed the predicted and familiar myside bias; people were more critical of evidence inconsistent with their prior attitudes. We also observed that people were more critical of specific evidence (regarding the ACA) after answering general questions about evidence (the SRS test). However,

Table 2. Linear regressions predicting evaluations of the evidence in Study 1.

Variable	Model 1 β	Model 2 β	Model 3 β	Model 4 β
Consistent evidence	0.26*	0.17	0.08	–0.08
SRS prime condition	–0.36***	–0.44**	–0.36***	–0.61
SRS score	–0.12***	–0.12***	–0.14***	–0.16***
Education	0.02	0.02	0.02	0.02
Age	–0.01	–0.01	–0.01	–0.01
Male	0.39***	0.40***	0.39***	0.40***
Political conservatism	–0.11*	–0.11*	–0.11*	–0.11*
SRS prime condition × consistent evidence		0.18		0.27
Consistent evidence × SRS score			0.03	0.04
SRS prime condition × SRS score				0.03
SRS prime condition × consistent evidence × SRS score				–0.01
Constant	5.74***	5.77***	5.85***	5.98***
RSE	1.2	1.2	1.2	1.2
Multiple R^2	0.11	0.11	0.11	0.11
Adjusted R^2	0.10	0.10	0.10	0.10
F	10.68***	9.44***	9.41***	6.94***

Note: Evaluations of the evidence (index measure) reported on a seven-point scale, with higher scores reflecting more positive ratings. $N = 605$. SRS = Scientific Reasoning Scale.
*$p < 0.05$.
**$p < 0.01$.
***$p < 0.001$.

we found no interaction: completing the priming task had no effect on the degree of myside bias. Exploratory analyses found that participants with greater scientific reasoning skills, as measured by SRS scores, were more critical evaluators, but just as prone to myside bias and no more influenced by the priming task, compared to those with lower SRS scores.

Although our evidence evaluation index revealed the expected effect of the priming manipulation, its six items were so strongly correlated that they provided little insight into how the priming manipulation affected participants' reasoning processes, beyond making them generally more critical. For example, completing the SRS test might just have encouraged participants to read the stimulus articles more carefully, thereby discovering potential problems, without affecting what skills they used. The priming manipulation might also have triggered a demand effect, suggesting that the experimenters valued critical evaluators.

Study 2 addresses these concerns in two ways. One is to add a manipulation designed to prime critical thinking, but not specifically critical *scientific* thinking. The second is to add tasks designed to measure changes in actual and not just reported critical evidence evaluation. The new manipulation was a validated individual difference measure of numeracy (Lipkus et al., 2001), whose effects were compared to those of a replication of the SRS manipulation. The new tasks were (a) four questions asking participants to recall details from the article, to assess whether the primes induced more careful reading; and (b) a question asking participants to identify a limitation of the research, to assess whether the priming manipulations induced

more negative ratings, rather than actually engaging latent scientific reasoning skills.

Study 2

Study 2 replicated and extended Study 1. First, we added a new condition, the goal of which was to prime analytical thinking, but not scientific reasoning skills. In the Numeracy prime condition, participants completed a numeracy test before reading and evaluating the evidence, using a measure (Lipkus et al., 2001) that has been found to correlate positively with individuals' ability to read and evaluate numerical healthcare information (Peters, Dieckmann, Dixon, Hibbard, & Mertz, 2007).

We also added two tasks assessing the manipulations' effects on participants' information processing: (a) a four-item test asking participants to recall details of the evidence that they had just read and (b) a multiple-choice question asking them to choose the best criticism of the study, from a set of options. The recall test sought to assess whether participants in the priming conditions read the article more closely than did those in the control condition. The multiple-choice question sought to assess whether the priming manipulations evoked latent scientific reasoning skills or just induced a more negative mindset regarding the quality of scientific studies.

Method

Recruiting

We recruited participants from Amazon.com's Mechanical Turk, an online marketplace that posts tasks for workers to complete in exchange for compensation. We used the same screening procedure as in Study 1: potential participants were asked to take the (unpaid) 5-item screening test about their attitudes toward current political topics, promising that some would be invited to take the full survey, based upon their answers.

We also used the same quota system to recruit roughly equal numbers of Affordable Care Act supporters and opponents, this time seeking 900 total participants. A power analysis indicated that recruiting 800 participants would provide 80% power to detect a small effect size ($f = 0.11$) for the interaction between Priming Condition (SRS prime vs. Numeracy prime vs. Control) and Evidence Condition (Inconsistent vs. Consistent) in a two-way ANOVA. In keeping with the sample size of Study 1, and accommodating the attrition of participants who failed data quality assurance items, we raised that target to 900.

We recruited an initial sample of 1191 participants, 6 of whom did not give informed consent and were not asked to continue. Out of the 1185

participants who completed the 5 screening questions, 198 did not identify as either supporters or opponents of the Affordable Care Act, hence were not asked to take the full survey. 72 supporters were not asked to take the full survey once we were over quota. In contrast to Study 1, the quota for supporters filled more quickly, suggesting differences in the subject populations. A survey software error lost data for 8 of the remaining 915 participants. Thirty-seven additional participants did not answer both data quality assurance questions correctly (Meade & Craig, 2012). The final sample had 870 participants, 425 Affordable Care Act opponents and 445 supporters.

Participants

We recruited American MTurk workers to take our study. Based on self-reports, 49% of those in the final sample were male, with mean age of 37 ($SD = 11$). One percent reported less than high school education; 12% had graduated high school or received a GED; 38% had completed at least some college or received an associate's degree; 48% percent reported having a bachelor's degree or higher. Forty-nine percent reported being liberal or very liberal, 26% moderate, and 26% conservative or very conservative.

Experimental design and measures

The experiment had a 3×2 between-subjects design. Data collection took place online and remotely, via a survey hosted on the Qualtrics platform. As in Study 1, participants were randomly assigned to read one of two news articles, describing a study whose evidence that was either consistent or inconsistent with their position on the Affordable Care Act [full articles appear in Table A1]. Participants were also randomly assigned to one of three experimental conditions: the SRS prime of Study 1, the Control condition of Study 1, or a new Numeracy prime condition, in which participants completed an 11-item numeracy test (Lipkus et al., 2001) before reading the article and answering questions about it. Participants in the Control and Numeracy prime conditions took the SRS after reading the article. Participants in the SRS prime and Control conditions took the numeracy test after reading the article. Reliabilities for the SRS and the numeracy test were Cronbach's $\alpha = 0.72$ and 0.71, respectively; Table A3 reports summary statistics for these scales. Responses to the SRS and the Numeracy scale were correlated, $r = 0.47$, $p < 0.001$, consistent with prior work (Drummond & Fischhoff, 2017b).

Procedure. The stimuli and instructions were the same as in Study 1.

Dependent measures. Immediately after reading the article, all participants completed the same six-item evidence evaluation measure as in Study 1. Once again, ratings on the items were highly correlated

(Cronbach's $\alpha = 0.89$), leading us to create an index variable equal to their average.

Participants then answered a *reasoning* question about the study that they had just read about, "Which of the following is the best criticism that can be made of this study?" They selected among three reasons, one of which was consistent with the article: "The two age groups of study participants were different in ways other than their being affected by the ACA provision." (Both studies used a difference-in-differences analysis that compared 19- to -25 year-olds to 26- to -34 year-olds.) The other two reasons were not consistent with the article: "There were not enough people in the study" and "The young adults in the study were not representative of the American population." (The news article does not mention the study's sample size, but does indicate that it recruited participants as part of a nationally representative survey.)

Additionally, participants answered a four-item *recall* measure asking them to recall (a) the name of the national survey that provided the data for the study, (b) the outcomes that were examined, (c) the age groups of participants in the study, and (d) the study's findings. Participants' answers were scored for accuracy, based on the article that they had read, with scores ranging from 0 (none correct) to 4 (all correct).

Additional measures. Participants answered three exploratory items not reported here. In the final section of the survey, all participants answered the same demographic questions as in Study 1, asking for their gender, age, education, and political beliefs. Table A3 reports summary statistics for these measures.

After the survey, participants were debriefed about the purpose of the study and provided with links to the abstracts of both scientific papers.

Results

Assignment to condition

Random assignment resulted in 437 participants (223 supporters) reading the version of the article consistent with their ACA attitudes and 433 (222 supporters) reading the inconsistent one. Participants who read the versions that were consistent and inconsistent with their attitudes did not differ significantly in SRS scores, numeracy scores, age, gender, education, or political conservatism (all p's > 0.25). One-way ANOVAs found no significant differences across the three priming conditions in participants' SRS scores, age, education, or political conservatism (all p's > 0.20). Numeracy scores were somewhat higher in the Numeracy Priming condition ($M = 9.3$) than in the SRS Priming and Control conditions, ($M = 8.9$, in both), $F(2, 867) = 4.48$, MSE $= 3.63$, $p = 0.012$. There were more women in the Control condition

Table 3. Group means for evaluations of the evidence in Study 2.

	Inconsistent evidence			Consistent evidence		
Priming condition	*M*	*SD*	*n*	*M*	*SD*	*n*
SRS prime	4.32	1.22	136	4.97	1.11	155
Numeracy prime	4.33	1.14	148	4.83	1.15	138
Control	4.40	1.21	149	5.06	1.07	144

Note: Evaluations of the evidence (index measure) reported on a seven-point scale, with higher scores reflecting more positive ratings.
SRS = Scientific Reasoning Scale.

(56%) than in the Numeracy Priming and the SRS Priming conditions (49% and 47%, respectively), $F(2, 867) = 3.07$, MSE =0.25, $p = 0.047$.

Evidence-evaluation index scores

Group means for the evidence evaluation index appear in Table 3. Looking at the 2×2 subset of Study 2 that replicated Study 1 with the new sample ($N = 584$), we ran a two-way ANOVA, which revealed no interaction between SRS prime condition and Evidence condition, $F(1, 580) = 0.04$, MSE =1.33. We reran the model without the interaction term and observed a significant main effect for Evidence condition, $F(1, 581) = 46.5$, MSE =1.33, $p < 0.001$, but not for SRS prime condition. Participants who received Consistent evidence evaluated it less critically than did those who received Inconsistent evidence.

Looking at the 2×2 subset of Study 2 that compares the Numeracy prime and Control condition ($N = 579$), we find the same pattern: a main effect of Evidence condition and no interaction, $F(1, 576) = 0.77$, MSE =1.32. We reran the model without the interaction term and observed a significant main effect for Evidence condition, $F(1, 575) = 36.8$, MSE= 1.32, $p < 0.001$, but not for the Numeracy prime condition. Again, participants who received Consistent evidence evaluated it less critically than did those who received Inconsistent evidence.

Table 4 presents regression results. Model 0 displays the main effects of condition: echoing the ANOVA results above, responses did not differ between the SRS prime condition and the Control condition, and the Numeracy prime condition and the Control condition, but participants did rate consistent evidence more positively. Model 1 shows that these results are robust to the inclusion of covariates. Model 2 shows non-significant interactions between the SRS Priming and Evidence conditions and between the Numeracy Priming and Evidence conditions. Thus, neither prime appeared to affect the degree of myside bias, as seen in the ANOVAs.

Table 4. Linear regressions predicting evaluations of the evidence in Study 2.

Variable	Model 0 β	Model 1 β	Model 2 β	Model 3 β	Model 4 β
Constant	4.42***	5.80***	5.78***	5.99***	6.00***
Consistent evidence	0.60***	0.59***	0.63***	0.21	0.06
SRS prime condition	−0.08	−0.11	−0.12	−0.11	−0.09
Numeracy prime condition	−0.15	−0.14	−0.08	−0.15	−0.15
SRS score		−0.10***	−0.10***	−0.13***	−0.12***
SRS prime condition × consistent evidence			0.02		0.38
Numeracy prime condition × consistent evidence			−0.12		
Consistent evidence × SRS score				0.06	0.07*
SRS prime condition × SRS score					−0.01
SRS prime condition × consistent evidence × SRS score					−0.04
Numeracy score		−0.01	−0.01	−0.01	−0.01
Education		−0.04	−0.04	−0.04	−0.04
Age		−0.01*	−0.01*	−0.01*	−0.01*
Male		0.15	0.15	0.16*	0.16*
Political conservatism		−0.06	−0.06	−0.06	−0.06
RSE	1.15	1.11	1.12	1.11	1.11
Multiple R^2	0.07	0.13	0.13	0.14	0.14
Adjusted R^2	0.06	0.12	0.12	0.13	0.13
F	20.57***	14.59***	11.98***	13.58***	10.57***

Note: Evaluations of the evidence (index measure) reported on a seven-point scale, with higher scores reflecting more positive ratings. $N = 870$. SRS = Scientific Reasoning Scale.
*$p < 0.05$.
**$p < 0.01$.
***$p < 0.001$.

Myside bias and scientific reasoning

We next tested whether participants with greater scientific reasoning ability were more biased, and more affected by the manipulations. As in Study 1, participants with higher SRS scores evaluated the evidence more harshly. We found that individuals with higher SRS scores showed marginally more myside bias, $\beta = 0.06$, $p = 0.053$ [Table 4, column 3]. We also examined whether the effect of the SRS prime varied by SRS score, but did not find a three-way interaction involving Priming condition, Evidence condition, and SRS score [Table 4, column 4].

Recall

The number of correctly answered recall questions was similar in all three conditions, as seen in two-way ANOVAs and linear regressions (not reported) predicting the number of recall questions answered correctly as a function of condition and covariates. Thus, participants appeared to pay similar attention to article details in all conditions. On average, they answered 2.25 recall questions correctly out of 4 ($SD = 1.17$). Those with higher SRS scores and numeracy scores recalled more details correctly, $\beta = 0.05$, $p = 0.003$ and $\beta = 0.11$, $p < 0.001$, respectively.

Table 5. Logistic regressions predicting the probability of correctly identifying a limitation of the evidence in Study 2.

Variable	Model 0 β	Model 1 β	Model 2 β
Constant	0.90***	−1.31*	−1.34*
Consistent evidence	0.20	0.19	0.24
SRS prime condition	0.45*	0.51*	0.58*
Numeracy prime condition	−0.01	−0.04	−0.04
SRS prime condition x consistent evidence			−0.14
Numeracy prime condition x consistent evidence			−0.01
SRS score		0.07*	0.07*
Numeracy score		0.11*	0.11*
Education		0.07	0.07
Age		0.01	0.01
Male		−0.32	−0.32
Political conservatism		0.11	0.11
AIC	965	952	956

Note: $N = 870$. SRS = Scientific Reasoning Scale.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

We repeated these analyses, using logistic regressions to examine performance only on the recall question that asked participants to recall the findings of the study that they read. Fifty-nine percent of participants chose the correct answer among the four options. That percentage did not differ by condition. Thus, the ability to recall the finding was unrelated to whether it was consistent with participants' attitudes or if they had been primed in either way. Participants with higher SRS scores and numeracy scores were more likely to answer this item correctly, $\beta = 0.12$, $p < 0.001$ and $\beta = 0.16$, $p < 0.001$, respectively, while political conservatives were less likely to answer correctly, $\beta = -0.14$, $p = 0.03$.

Reasoning about study limitations

Compared to participants in the Control condition, those who received the SRS prime were more likely to answer the reasoning question correctly (81% vs. 73%), whereas those who received the Numeracy prime were not (73%). As Table 5 shows, that success was unrelated to whether participants received attitude-consistent evidence. Thus, participants' responses to this reasoning question did not display myside bias. Table 5 presents logistic regressions predicting the likelihood of a correct answer. We specified a simple model without covariates (Model 0) and with covariates (Model 1), and a model adding interactions terms between the conditions (Model 2). Model 0 indicates that those in the SRS prime condition were more likely to answer the question correctly than were those in the control condition; no effect of Evidence condition was observed. Model 1 shows that these

results are robust to the inclusion of covariates: Participants in the SRS prime condition were still more likely to answer the question correctly, but not participants in the Numeracy prime condition. Participants in all conditions were equally able to identify a limitation of the study, regardless of whether they read evidence consistent or inconsistent with their beliefs. Model 2 revealed no significant interactions between Evidence condition and either the SRS or the Numeracy Priming condition, indicating that the primes had similar effects whether the evidence was consistent or inconsistent with participants' beliefs. Those with higher SRS and numeracy test scores were more likely to answer the question correctly.

Discussion

Study 2, like Study 1, revealed the familiar myside bias: participants rated evidence more critically when it was inconsistent with their prior beliefs (Cohen's $d = 0.52$). However, unlike in Study 1, taking the Scientific Reasoning Scale (SRS) before reading the evidence did not affect ratings. Nor did taking the numeracy test before reading the evidence. Neither taking the SRS nor taking the numeracy test had any effect on myside bias. While participants with higher SRS scores were marginally more biased in favor of belief-consistent evidence, the effect of the priming manipulation was unrelated to scientific reasoning ability.

Despite having no effect in this sample on either ratings or myside bias, the SRS priming manipulation did affect participants' reasoning processes, increasing their ability to identify a study limitation. That difference was not observed with the Numeracy test prime. The ability to identify a limitation was unrelated to the evidence that participants had read, indicating that they did not selectively identify the limitation when it was inconsistent with their attitude, as myside bias might encourage. The similarity of participants' ability to recall details from the article, in the different conditions, suggests that priming effects are not due to paying greater attention to the article.

Thus, taking the SRS appears to have activated latent scientific reasoning skills, and not simply to have evoked a more negative or critical attitude. However, the activation of those skills did not lead to more critical or less biased ratings of the evidence.

Reducing bias due to prior beliefs requires possessing and applying the skills needed to solve the problem at hand (Stanovich & West, 2008; Stanovich et al., 2013). Our SRS prime manipulation sought to activate participants' latent scientific reasoning skills; however, it did not send them an explicit signal to apply these skills to their evaluations of the new evidence. Study 3 makes that need explicit, by adding a new experimental condition,

SRS prime + Instructions. Before participants in that condition evaluate the evidence, they are instructed to focus on applying their scientific reasoning skills to the task. We continue to examine the relationship between the manipulation and SRS scores.

Study 3 compared the new SRS prime + Instructions condition to the Control and SRS prime conditions of the prior two studies.

Study 3

In both Study 1 and Study 2, taking the Scientific Reasoning Scale did not reduce myside bias. Study 3 sought to strengthen the SRS prime manipulation by adding explicit instructions to apply one's scientific reasoning skills when evaluating the stimulus study. Study 3 includes the SRS prime and Control conditions of Study 1, along with a new SRS prime + Instructions condition. All participants answered the full set of dependent measures from Study 2.

Method

Recruiting

As in Study 2, we recruited participants from Amazon.com's Mechanical Turk. We used the same screening procedure as in Studies 1 and 2. We also used the same quota system to recruit roughly equal numbers of Affordable Care Act supporters and opponents, this time seeking 900 total participants. As in Study 2, a power analysis indicated that recruiting 800 participants would provide 80% power to detect a small effect size ($f = 0.11$) for the interaction between Priming Condition (SRS prime vs. SRS prime + Instructions vs. Control) and Evidence Condition (Inconsistent vs. Consistent) in the planned two-way ANOVA. In keeping with Studies 1 and 2, which sought to recruit 150 participants per condition, we set our target *N* as 900, also accommodating the attrition of participants who failed data quality assurance items.

We recruited an initial sample of 1160 participants, 9 of whom did not give informed consent and were not asked to continue. Out of the 1151 participants who completed the 5 screening questions, 203 did not identify as either supporters or opponents of the Affordable Care Act, hence were not asked to take the full survey. The quota for ACA supporters filled more quickly, and 43 ACA supporters were not asked to take the full survey once we were over quota. 78 additional participants did not answer both data quality assurance questions correctly (Meade & Craig, 2012). A survey software error lost data for three additional participants. The final sample had 824 participants, 384 Affordable Care Act opponents and 440 supporters.

Participants

We recruited American MTurk workers for our study, excluding those who had taken Study 2. Data collection occurred remotely and online. Based on self-reports, 53% of participants in the final sample were male, with mean age of 37 ($SD = 11$). One percent reported less than high school education; 9% had graduated high school or received a GED; 33% had completed at least some college or received an associate's degree; 58% percent reported having a bachelor's degree or higher. Forty-seven percent reported being liberal or very liberal, 22% moderate, and 30% conservative or very conservative.

Experimental design and measures

The experiment had a 3×2 between-subjects design. Participants were randomly selected to receive one of two versions of the ACA article used in Studies 1 and 2. Additionally, participants were randomly assigned to the SRS priming or Control conditions, identical to those in Study 2, or to the new SRS priming + Instructions condition.

Participants in the new SRS priming + Instructions condition received additional instructions to apply their reasoning skills to reading the article and the evidence it contained. Before taking the SRS, these participants were told, "This section of the survey is intended to help you "warm up" your scientific thinking skills." Before reading the article, they were told: "In the first part of the survey, you had a chance to warm up your scientific thinking skills. Now, please apply your scientific thinking skills to evaluate the quality of the science and the strength of the evidence in the study that is described in the article." Before answering the dependent measures, participants were told, "As you answer these questions, please tell us what you think about the quality of the science and the strength of the evidence in this one study."

The stimuli and measures were otherwise the same as in Study 1. Table A4 reports summary statistics for covariates.

Results

Assignment to condition

Random assignment resulted in 424 participants (227 supporters) reading the version of the article consistent with their ACA attitudes and 400 (213 supporters) reading the inconsistent one. Participants who read the versions that were consistent and inconsistent with their attitudes did not differ significantly in SRS scores, age, gender, education, or political conservatism (all p's > 0.05). One-way ANOVAs found no significant differences across the three priming conditions in participants' SRS scores, age, gender, education, or political conservatism (all p's > 0.09).

Table 6. Group means for evaluations of the evidence in Study 3.

	Inconsistent Evidence			Consistent Evidence		
Priming Condition	*M*	*SD*	*n*	*M*	*SD*	*n*
SRS prime	4.59	1.33	135	5.08	1.22	135
SRS prime + Instructions	4.51	1.31	128	4.90	1.14	151
Control	4.35	1.19	137	5.18	0.99	138

Note: Evaluations of the evidence (index measure) reported on a seven-point scale, with higher scores reflecting more positive ratings. SRS = Scientific Reasoning Scale.

Evidence-evaluation index scores

Group means for the evidence evaluation index are displayed in Table 6. Cronbach's alpha for the evidence evaluation index was 0.9. Looking at the 2x2 subset of Study 3 that replicated Studies 1 and 2 with the new sample ($N = 545$), we ran a two-way ANOVA, which revealed no interaction between the SRS prime and Evidence conditions, $F(1, 541) = 2.75$, MSE =1.41. We reran the model without the interaction term and observed a significant main effect for Evidence condition, $F(1, 542) = 42.1$, MSE =1.41, $p < 0.001$, but not for the SRS prime condition. Participants who received Consistent evidence evaluated it less critically than did those who received Inconsistent evidence.

Using a two-way ANOVA to look at the 2×2 subset of Study 2 that compares the SRS prime + Instructions to the Control condition ($N = 554$), we found a significant interaction between Evidence condition and SRS condition, $F(1, 550) = 4.95$, MSE =1.34, $p = 0.03$.

We used linear regressions to investigate the nature of this interaction. Table 7 presents regression results. Model 0 displays the main effects of condition: responses did not differ between the SRS prime condition and the Control condition, or between the SRS prime + Instructions condition and the Control condition, but participants did rate consistent evidence more positively. Model 1 shows that these results are robust to the inclusion of covariates. Model 2 shows a non-significant interaction between the SRS prime and Evidence conditions, and a statistically significant interaction, $\beta = -0.46$, $p < 0.05$, between the SRS prime + Instructions and Evidence conditions. Taking the SRS, unaccompanied by the instructions, did not reduce the extent to which participants' responses displayed myside bias. However, taking the SRS accompanied by the extra instructions reduced myside bias, compared to the control condition.

Myside bias and scientific reasoning

Next, we examined whether participants with greater scientific reasoning ability were more biased in favor of belief-consistent evidence, and more affected by the manipulations. Table 7 shows that, as in Studies 1 and 2,

Table 7. Linear regressions predicting evaluations of the evidence in Study 3.

Variable	Model 0 β	Model 1 β	Model 2 β	Model 3 β	Model 4 β	Model 5 β
Constant	4.48***	6.04***	5.95***	6.37***	6.37***	6.24***
Consistent evidence	0.57***	0.58***	0.85***	−0.04	−0.18	0.30
SRS prime	0.08	0.004	0.18	−0.002	−0.04	−0.002
SRS prime + Instructions	−0.07	−0.13	0.11	−0.14	−0.14	0.31
SRS score		−0.12***	−0.12***	−0.17***	−0.17***	−0.15***
SRS prime x consistent evidence			−0.35		0.41	
SRS prime + Instructions x consistent evidence			−0.46*			−0.91*
Consistent evidence x SRS score				0.10**	0.12***	0.06
SRS prime x SRS score					0.01	
SRS prime x consistent evidence x SRS score					−0.08	
SRS prime + Instructions x SRS score						−0.05
SRS prime + Instructions x consistent evidence x SRS score						0.10
Education		−0.01	−0.02	−0.02	−0.02	−0.02
Age		−0.02***	−0.02***	−0.02***	−0.02***	−0.02***
Male		0.04	0.05	0.05	0.05	0.06
Political conservatism		−0.03	−0.04	−0.03	−0.03	−0.04
RSE	1.2	1.14	1.14	1.13	1.13	1.13
Multiple R^2	0.05	0.15	0.16	0.17	0.17	0.17
Adjusted R^2	0.05	0.15	0.15	0.16	0.16	0.16
F	15.89***	18.54***	15.52***	17.87***	13.61***	13.88***

Note: Evaluations of the evidence (index measure) reported on a seven-point scale, with higher scores reflecting more positive ratings. $N = 824$. SRS = Scientific Reasoning Scale.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

participants with higher SRS scores evaluated the evidence more harshly. Individuals with higher SRS scores again showed more myside bias, $\beta = 0.10$, $p < 0.01$ [Table 7, Model 3]. We also examined whether the effect of the SRS prime, and of the SRS prime + Instructions, varied by SRS score. We did not find a three-way interaction involving either Priming condition, Evidence condition, and SRS score (Table 7, Models 4 and 5).

Recall

We again used two-way ANOVAs and linear regressions (not reported) to examine whether the number of correctly answered recall questions was similar in all three conditions. As in Study 2, there were no statistically significant differences in recall by condition; thus, participants appeared to pay similar attention to article details in all conditions. On average, participants answered 2.33 recall questions correctly out of 4 ($SD = 1.16$). Those

Table 8. Logistic regressions predicting the probability of correctly identifying a limitation of the evidence in Study 3.

Variable	Model 0 β	Model 1 β	Model 2 β
Constant	0.89***	−0.55	−0.61
Consistent evidence	−0.13	−0.14	0.08
SRS prime	0.39*	0.47*	0.59*
SRS prime + Instructions	0.46*	0.54**	0.79**
SRS prime x consistent evidence			−0.24
SRS prime + Instructions x consistent evidence			−0.46
SRS score		0.10**	0.10**
Education		−0.004	−0.01
Age		0.02**	0.02**
Male		0.05	0.06
Political conservatism		−0.04	−0.05
AIC	930	916	919

Note: $N = 824$. SRS = Scientific Reasoning Scale.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

with higher SRS scores recalled more details correctly, $\beta = 0.10$, $p < 0.001$; those with higher education recalled fewer details correctly, $\beta = -0.09$, $p < 0.01$.

We next used logistic regressions to examine performance only on the question asking participants to recall the findings of the study that they read. Fifty-four percent of participants chose the correct answer among the four options, and that percentage did not differ by condition. Individuals who read consistent evidence were more likely to recall the study findings correctly, $\beta = 0.32$, $p < 0.05$. Participants with higher SRS scores were more likely to answer this item correctly, $\beta = 0.21$, $p < 0.001$, while more educated respondents were less likely to answer correctly, $\beta = -0.13$, $p = 0.03$.

Reasoning about study limitations

Compared to participants in the Control condition, those who received the SRS prime were more likely to answer the reasoning question correctly (77% vs. 69%), as were those who received the SRS prime + Instructions (78%). Table 8 presents logistic regressions predicting the likelihood of a correct answer. Model 0 indicates that those in the SRS prime and those in the SRS prime + Instructions conditions were more likely to answer the question correctly than were those in the control condition. Participants' answers to the reasoning question were again unrelated to whether they received attitude-consistent evidence. Model 1 shows that these results are robust to the inclusion of covariates, and Model 2 shows no significant interactions between Evidence condition and priming conditions, suggesting that the primes had similar effects whether the evidence was consistent

or inconsistent with participants' beliefs. Males and those with higher SRS scores were more likely to answer the question correctly.

Discussion

In Study 3, as in Studies 1 and 2, participants' subjective evaluations of scientific evidence on the efficacy of the Affordable Care Act were biased in favor of evidence consistent with their ACA beliefs (Cohen's $d = 0.47$). And, as in Studies 1 and 2, priming participants' scientific reasoning skills with a test of their scientific reasoning ability did not reduce the extent of this bias. However, a new experimental condition, which paired the SRS prime with explicit instructions to apply one's scientific thinking skills to the evidence evaluation, did reduce participants' myside bias. As in Study 2, recall of factual information did not differ across condition, nor did time spent reading the article ($F = 0.32$, $p = 0.73$), suggesting that differences across condition were not simply due to paying more attention to the article.

As in Study 2, participants who took the SRS before reading the article were more likely to identify a limitation of the study correctly; this was true for those in the SRS prime and SRS prime + Instructions conditions. These results suggest that taking the SRS activated participants' latent scientific reasoning skills; however, this activation only translated into reduced myside bias when participants were explicitly instructed to apply their reasoning skills.

General discussion

In three studies, we assessed whether priming scientific reasoning skills would affect the degree of myside bias in participants' evaluations of scientific evidence regarding the Affordable Care Act. Our prime was the 11-item Scientific Reasoning Scale (SRS; Drummond & Fischhoff, 2017b), a test of scientific reasoning skills. Across three studies, we found a small-to-medium-sized myside bias, with participants more critical of studies inconsistent with their prior beliefs. Study 1 found that taking the SRS before reading and evaluating scientific evidence did not affect myside bias. Study 2 replicated this result, but found that the SRS prime made participants better able to identify a study limitation, suggesting that the SRS activated participants' latent reasoning abilities. Study 2 also found that taking a numeracy test before reading and evaluating evidence had no effect on any measure, suggesting that priming critical non-scientific thinking did not affect judgments of scientific evidence. Finally, Study 3 added a new condition, supplementing the SRS prime with instructions to apply one's scientific thinking skills to the evidence evaluation task. As before, the SRS prime alone was not enough to reduce myside bias, but adding instructions to use those skills did.

Dual-process theories of reasoning suggest that biases such as myside bias occur due to automatic, intuitive judgments evoked by affective and associative processes. Such biases can be overcome by more effortful, analytical processing, for those who have those skills (Evans, 2008; Evans & Stanovich, 2013; Kahneman & Frederick, 2002). However, individuals who score higher on some measures of higher-order, analytical reasoning skills, such as intelligence and cognitive ability, have not proven less vulnerable to many judgmental biases (Stanovich et al., 2013; Stanovich & West, 2008). Such results suggest that explicit signals to activate one's analytical thinking skills, or task instructions to decouple prior beliefs, are needed to reduce myside bias (Evans, Handley, Neilens, & Over, 2010; Stanovich et al., 2013; Stanovich and West, 2008). Thus, the SRS priming manipulation may have been enough to make participants more critical (in Study 1) and more likely to identify a study limitation (Studies 2 and 3). However, it was not enough to reduce myside bias. Adding instructions to apply one's scientific reasoning skills did. We believe that our added instructions shifted participants' focus from using their analytical skills to support their existing positions to focusing on the evidence, in what might be considered a shift from motivated analytical reasoning to unmotivated analytical reasoning.

The absence of myside bias in respondents' identification of study limitations is consistent with this interpretation. Less directly, this interpretation is consistent with an emerging pattern of greater polarisation in scientific beliefs among individuals with more education and scientific reasoning skills, in situations without such instructions (Drummond & Fischhoff, 2017a; McCright et al., 2016).

The studies presented here have several limitations. First, Study 1 participants were drawn from a different population (Qualtrics) than Studies 2 and 3 (MTurk). The Qualtrics panelists in Study 1 had lower SRS scores ($M = 5.57$) than the Mechanical Turk workers in Study 2 ($M = 6.99$) and Study 3 ($M = 6.46$). However, it is not clear how that might have affected the results. A second limitation is that participants may have held beliefs regarding the Affordable Care Act of varying strength and substance. For example, opposition to the ACA has come from both conservatives who think that its regulations go too far and liberals who think it does not go far enough. We used just a general screening question to recruit supporters and opponents. However, here, too, it is uncertain how the differences may affect our results. In retrospect, it would have been good to assess participants' knowledge of the ACA. Future research might examine how myside bias varies with the reasons for beliefs about a topic, and knowledge of the topic. A fourth possible limitation, also without obvious potential effects, is using just one topic, the ACA. A final concern is the wording of the Instruction condition of Study 3. It was formulated to avoid demand effects, by asking participants to apply their scientific thinking skills without suggesting how, but might still have offered some hints.

In addition to their substantive message, our studies suggest that cognitively demanding individual difference measures, such the SRS, be deferred to the end of studies, lest they affect performance on other measures.

Fischhoff (1982) proposed classifying debiasing techniques into four categories: offering warnings, describing the bias, providing performance feedback, and changing how individuals approach a task. His review of two biases (hindsight bias, overconfidence), as well as a more recent review by Milkman, Chugh & Bazerman, (2009) found that warnings and descriptions had little effect, but task restructuring and performance feedback might provide some benefit (see also Morewedge et al., 2015; Moore et al., 2017). Our manipulations differ from conventional warnings, in providing a general direction (use your skills) with prompts for what that might mean (the SRS questions). Taken together, our findings imply that encouraging individuals to put on their "thinking caps" to evaluate scientific evidence may only help them to think in ways that reduce the myside bias when there is a clear signal to do so. Future work could develop practical and implementable interventions to provide both aspects of the needed debiasing.

Acknowledgements

Portions of the research have been presented at the 2016 Society for Judgment and Decision Making annual meeting, the associated Preconference on Debiasing Decision Makers, and the 2016 Society for Risk Analysis annual meeting. The authors thank Stephen Broomell and Andrew Parker for their comments on an earlier draft of this paper. Studies 1 and 2 were conducted while Caitlin Drummond was a graduate student at Carnegie Mellon University; Study 3 was conducted while she was a postdoctoral researcher at the University of Michigan.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported in part by the National Science Foundation [grant NCSE-1537364, to the authors; grant SES-0949710, to the Center for Climate and Energy Decision Making]; and by the Swedish Foundation for Humanities and Social Science. Caitlin Drummond was supported by a National Science Foundation Graduate Research Fellowship [grant DGE-1252522]. Study 1 was also partially supported by a Qualtrics Behavioral Research Grant awarded to Caitlin Drummond.

ORCID

Caitlin Drummond http://orcid.org/0000-0002-6480-170X
Baruch Fischhoff http://orcid.org/0000-0002-3030-6874

References

Attridge, N., & Inglis, M. (2015). Increasing cognitive inhibition with a difficult prior task: implications for mathematical thinking. *ZDM Mathematics Education*, *47* (5), 723–734. http://dx.doi.org/10.1007/s11858- 014-0656-1

Bargh, J. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology*, *36*(2), 147–168. http://doi.org/10.1002/ejsp.336.What

Bauer, M. W. (2009). The evolution of public understanding of science–discourse and comparative evidence. *Science Technology & Society*, *14*(2), 221–240. http://doi.org/10.1177/097172180901400202

Bauer, M. W., Allum, N., & Miller, S. (2007). What can we learn from 25 years of PUS survey research? Liberating and expanding the agenda. *Public Understanding of Science*, *16*(1), 79–95. http://doi.org/10.1177/0963662506071287

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, *63*(4), 568–584. http://doi.org/10.1037//0022-3514.63.4.568

Drummond, C., & Fischhoff, B. (2017a). Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(36), 9587–9592. doi:10.1073/pnas.1704882114

Drummond, C., & Fischhoff, B. (2017b). Development and validation of the scientific reasoning scale. *Journal of Behavioral Decision Making*, *30*(1), 26–38. doi:10.1002/bdm.1906

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology*, *59*(1), 255–278. doi:10.1146/annurev.psych.59.103006.093629

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. doi:10.1177/1745691612460685 doi:10.1177/1745691612460685

Evans, J. S. B. T., Handley, S. J., Neilens, H., & Over, D. (2010). The influence of cognitive ability and instructional set on causal conditional inference. *The Quarterly Journal of Experimental Psychology*, *63*(5), 892–909. doi:10.1080/17470210903111821

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). New York, NY: Cambridge University Press.

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, *90*(3), 239–260. doi:10.1037/0033-295X.90.3.239

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. http://doi.org/10.1257/089533005775196732 doi:10.1257/089533005775196732

Funk, C., & Rainie, L. (2015). Public and scientists' views on science and society. Washington, DC: Pew Research Center. Retrieved from the Pew Research Center website: http://www.pewinternet.org/2015/01/29/public-and-scientists-views-on-science-and-society/

Jain, S. P., & Maheswaran, D. (2000). Motivated reasoning: A depth of processing perspective. *Journal of Consumer Research*, *26*(4), 358–371. http://doi.org/10.1086/209568 doi:10.1086/209568

Kahneman, D. (2011). *Thinking fast and slow*. New York, NY: Farrar, Strauss & Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited□: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, and D. Kahneman (Eds.), *Heuristics of intuitive judgment: Extensions and applications*. New York, NY: Cambridge University Press. http://doi.org/10.1017/CBO9780511808098.004

Kotagal, M., Carle, A., Kessler, L., & Flum, D. (2014). Limited impact on health and access to care for 19- to 25-year- olds following the patient protection and affordable care act. *JAMA Pediatrics*, *168*(11), 1023–1029. doi:10.1001/jamapediatrics.2014.1208 doi:10.1001/jamapediatrics.2014.1208

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. doi:10.1037//0033-2909.108.3.480 doi:10.1037//0033-2909.108.3.480

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*(1), 37–44. http://doi.org/10.1177/0272989X0102100105 doi:10.1177/0272989X0102100105

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098–2109. doi:10.1037//0022-3514.37.11.2098 doi:10.1037//0022-3514.37.11.2098

McCright, A. M., Marquart-Pyatt, S. T., Shwom, R. L., Brechin, S. R., & Allen, S. (2016). Ideology, capitalism, and climate: Explaining public views about climate change in the United States. *Energy Research & Social Science*, *21*, 180–189. doi:10.1016/j.erss.2016.08.003

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. doi:10.1037/a0028085 doi:10.1037/a0028085

Milkman, K. L., Chugh, D., & Bazerman, M. H. (2009). How can decision making be improved? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *4*(4), 379–383.

Miller, J. D. (2004). Public understanding of, and attitudes toward, scientific research: What we know and what we need to know. *Public Understanding of Science*, *13*(3), 273–294. http://doi.org/10.1177/0963662504044908

Moore, D. A., Swift, S. A., Minster, A., Mellers, B., Ungar, L., Tetlock, P., … Tenney, E. R. (2017). Confidence calibration in a multiyear geopolitical forecasting competition. *Management Science*, *63*(11), 3552–3565. doi:10.1287/mnsc.2016.2525

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 129–140. http://doi.org/10.1177/2372732215600886 doi:10.1177/2372732215600886

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, *29*(1), 85–95. doi:10.1080/00335557743000053

Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. doi:10.1037/1089-2680.2.2.175

Peters, E., Dieckmann, N., Dixon, A., Hibbard, J. H., & Mertz, C. K. (2007). Less is more in presenting quality information to consumers. *Medical Care Research and Review*, *64*(2), 169–190. http://doi.org/10.1177/1077558707064002030 1 doi:10.1177/1077558707064002030 1

Sommers, B. D., Buchmueller, T., Decker, S. L., Carey, C., & Kronick, R. (2013). The affordable care act has led to significant gains in health insurance and access to care for young adults. *Health Affairs*, *32*(1), 165–174. doi:10.1377/hlthaff.2012.0552 doi:10.1377/hlthaff.2012.0552

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Personality Processes and Individual Differences*, *94*(4), 672–695. doi:10.1037/0022-3514.94.4.672

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. doi: 10.1177/0963721413480174

Toplak, M. E., & Stanovich, K. E. (2003). Associations between myside bias on an informal reasoning task and amount of post-secondary education. *Applied Cognitive Psychology*, *17*(7), 851–860. doi:10.1002/acp.915 doi:10.1002/acp.915

Wolfe, C. R., & Britt, M. A. (2008). The locus of the myside bias in written argumentation. *Thinking and Reasoning*, *14*(1), 1–27. http://doi.org/10.1080/13546780701527674 doi:10.1080/13546780701527674

Wolfe, C. R., Britt, M. A., & Butler, J. A. (2009). Argumentation schema and argumentation. *Written Communication*, *26*(2), 183–209. http://doi.org/10.1177/0741088309333019

Appendix A

Table A1. Experimental materials: News articles.

Study: Health Care *More Affordable* for Many Young Adults	Study: Health Care *Remains Unaffordable* for Many Young Adults
A new study has found that *fewer young adults are putting off or choosing not to get medical care* after the Affordable Care Act (ACA).	A new study has found that *young adults are no more likely to be able to afford prescription medication and physician visits* after the Affordable Care Act (ACA).
A provision in the ACA allows young adults to stay on their parents' health insurance until the age of 26. At the time that the ACA was passed, its supporters claimed that this provision would increase the number of young adults getting health care.	A provision in the ACA allows young adults to stay on their parents' health insurance until the age of 26. At the time that the ACA was passed, its supporters claimed that this provision would increase the number of young adults getting health care.
Now, the research is in to address these claims.	Now, the research is in to address these claims.
The study used data from *the nationally representative National Health Interview Survey,* run by the Centers for Disease Control and Prevention.	The study used data from *two nationally representative surveys: the Behavioral Risk Factor Surveillance System and the National Health Interview Survey, both of which are* run by the Centers for Disease Control and Prevention.
The study focused on two survey questions. The first asked respondents whether they *had delayed getting care in the past year because of its cost.* The second asked whether they *did not get care in the past year because of cost*.	The study focused on two survey questions. The first asked respondents whether they *were unable to afford prescription medication in the past year*. The second asked whether they *were unable to afford to see a physician in the past year*.
The study found that the percentage of young adults (aged 19 to 25) who reported *delaying getting care* dropped by 5.6 percentage points after the ACA came into effect.	The study found that the percentage of young adults (aged 19 to 25) who reported *being unable to afford prescription medication* dropped by *3.1* percentage points after the ACA came into effect.
But, the researchers couldn't tell whether this drop was caused by the ACA provision or by other possible factors, such as an improvement in the economy that increased wages and made health care more affordable.	But, the researchers couldn't tell whether this drop was caused by the ACA provision or by other possible factors, such as an improvement in the economy that increased wages and made health care more affordable.
To assess the effects of these other factors, the researchers repeated their analyses for respondents aged 26-34, who were not affected by the ACA provision.	To assess the effects of these other factors, the researchers repeated their analyses for respondents aged 26-34, who were not affected by the ACA provision.
The authors found that the proportion of the 26-to-34-year-olds who reported *delaying getting care* also dropped, by *1.6* percentage points. That drop was *significantly smaller* than the *5.6* percentage point drop for the young adults, indicating that the ACA was probably driving the increase in young adults getting health care.	The authors found that the proportion of the 26-to-34-year-olds who reported *being unable to afford prescription medication* also dropped, by *2.7* percentage points. That drop was *not significantly different* than the *3.1* percentage point drop for the young adults, indicating that the ACA was probably *not* driving the increase in young adults getting health care.
The story was the same when the authors looked at the respondents who *chose not to get medical care in the past year because of cost.* This research suggests that the Affordable Care Act may have helped make health care more affordable for young adults.	The story was the same when the authors looked at the respondents who *were unable to afford to see a physician*. This research suggests that the Affordable Care Act may *not* have helped make health care more affordable for young adults.

Note: The article in the left column was based on Sommers et al, 2013, and the article in the right column was based on Kotagal et al,2014. Words in Italicised text describe specific aspects of the two articles; words were not Italicised when presented to participants.

Table A2. Means, scale reliabilities, and correlations between covariates in Study 1.

Variable	M (SD)	α	1	2	3	4
1. SRS	5.57 (2.45)	0.62				
2. Education	4.65 (1.38)	__	0.21***			
3. Age	47.33 (13.41)	__	0.22***	−0.03		
4. Male	0.47	__	0.04	0.10*	0.06	
5. Political conservatism	3.18 (1.09)	__	0.04	0.06	0.12**	0.08*

Note: Pearson correlations between covariates in Study 1 ($N = 605$). SRS = Scientific Reasoning Scale.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

Table A3. Means, scale reliabilities, and correlations between covariates in Study 2.

Variable	M (SD)	α	1	2	3	4	5
1. SRS	6.99 (2.63)	0.72					
2. Education	5.06 (1.29)	__	0.23***				
3. Age	37.17 (10.89)	__	0.11**	0.01			
4. Male	0.49	__	0.15***	0.00	−0.06		
5. Political conservatism	2.66 (1.15)	__	−0.16***	−0.09*	0.01	0.02	
6. Numeracy	9.05 (1.91)	0.71	0.47***	0.17***	0.08*	0.19***	−0.11**

Note: Pearson correlations between covariates in Study 2 ($N = 870$). SRS = Scientific Reasoning Scale.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

Table A4. Means, scale reliabilities, and correlations between covariates in Study 3.

Variable	M (SD)	α	1	2	3	4
1. SRS	6.46 (2.70)	0.72				
2. Education	5.32 (1.25)	__	0.10**			
3. Age	36.7 (11)	__	0.13***	−0.08*		
4. Male	0.53	__	0.001	0.05	−0.18***	
5. Political conservatism	2.72 (1.27)	__	−0.18***	−0.03	0.03	0.09*

Note: Pearson correlations between covariates in Study 3 ($N = 824$). SRS = Scientific Reasoning Scale.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.