Belief Updating and Argument Evaluation

Megan D Bardolph (mbardolph@ucsd.edu)
Department of Cognitive Science (0515), 9500 Gilman Drive
La Jolla, CA 92093 USA

Seana Coulson (scoulson@ucsd.edu)
Department of Cognitive Science (0515), 9500 Gilman Drive
La Jolla, CA 92093 USA

Abstract

Studies of how evidence affects beliefs sometimes show belief polarization in response to mixed evidence. However, the nature of the mental processes leading to change in opinion is up for debate. Different accounts of how people process evidence and then update their beliefs make different predictions, especially about one-sided evidence, which is rarely examined. We presented subjects with multiple text arguments regarding socio-political topics as one-sided or mixed evidence. Participants rated arguments differently according to their extant beliefs, which is consistent with accounts of motivated reasoning. They did not polarize afterward, instead showing evidence of belief updating according to Bayesian principles: belief change is sensitive to prior opinions and to the direction and quality of the evidence presented. These data support rethinking some of the mental processes underlying incorporation of evidence into a personal belief structure.

Keywords: cognitive science; decision making; reasoning; language and thought; psychology; motivated reasoning; rationality

Introduction

As people navigate a world filled with information, they must make decisions in order to accomplish various goals. The beliefs that individuals hold provide a structure in which new information is evaluated and potentially integrated with existing beliefs. Different models explain how people evaluate information and use that information to update their beliefs, leading to potentially different implications about human rationality (Oaksford & Chater, 2007; Klayman & Ha, 1987).

Information from the world can be thought of as data, or evidence, supporting or disconfirming a hypothesis. The evidence may be accepted without examination or it may be judged before it is used to update one's beliefs, or a hypothesis (Nisbett & Ross, 1980; Pyszczynski & Greenberg, 1987). From a Bayesian perspective, evidence is judged according to existing hypotheses about the world along with data that has already been observed. Bayes' rule provides a general model for constructing a posterior probability, $P(hypothesis|data)$, as a function of prior beliefs and observed evidence: $P(hypothesis) * P(data|hypothesis)$.

People's prior hypotheses about the world may differ depending on the data they have observed. Furthermore, the nature of people's hypothesis space for a given topic is not always easy to define (Tenenbaum, Kemp, Griffiths, & Goodman, 2011; Jern, Chang, & Kemp, 2014). Accounts of Bayesian updating explicitly allow for evidence to be treated differently depending on whether it is in agreement with one's prior beliefs (Gerber & Green, 1999). However, unless there is reason to suspect the source or validity of the evidence, Bayesian normative accounts still require that people update their prior beliefs in the direction of the evidence. This updating can be small, but it cannot be in the opposite direction. If such a shift occurs, it should be viewed as a violation of normative updating under this model.

Alternatively, differential rating of information (evidence) may be due to motivated, or hot cognition processes (Kunda, 1990; Ditto & Lopez, 1992). Under motivated accounts, evidence compatible with an extant opinion is accepted, while incompatible evidence creates negative emotions and is therefore critically examined and judged more negatively because of its incompatibility. This difference in judgement can lead to attitude polarization, or belief polarization.

Lord, Ross, and Lepper (1979) suggested that attitude polarization occurs because people with opposing views can come to opposite conclusions from the very same set of evidence. In a classic study, the authors queried participants about their views on capital punishment, and then revealed the results of two studies, one which suggested the death penalty deters crime, and one which suggested the opposite conclusion. Participants were asked to rate the quality of each study, and then to recharacterize their views on the death penalty. The authors found that proponents of capital punishment rated the study showing the deterrent effect of the death penalty to be superior to that showing that the death penalty did not affect crime levels, and subsequently adjusted their beliefs to more strongly favor capital punishment. By contrast, opponents of capital punishment favored the study that showed the death penalty had little effect on crime, and subsequently adjusted their beliefs to more strongly oppose capital punishment. So-called biased assimilation is the phenomenon by which participants' prior beliefs impact the way they evaluate novel evidence, and it would seem to undermine the possibility of achieving consensus (Lord, Ross, & Lepper, 1979).

Taber and Lodge (2006) suggest that "primacy and automaticity of affect kick-start the processes that spark motivated biases when citizens encounter attitudinally contrary information." Taber and colleagues (2009) found evidence of an attitude congruency bias, where people evaluate arguments and evidence that supports their prior opinions as stronger than nonsupporting information; and attitude polarization, where this bias leads to polarization with exposure to the same set of information.

The present study aims to clarify the differences between processing compatible and incompatible evidence, separating the effects of the two types of evidence. To do this, we will examine evidence rating for both mixed evidence (as used in prior studies) and one-sided evidence (previously missing from much of the literature). Studies using mixed evidence imply that participants process congruent and incongruent information using different processes; for example, readily accepting compatible arguments while spending more time and mental resources to undermine incompatible arguments (Edwards & Smith, 1996; Taber, Cann, & Kucsova, 2009). It is not clear whether compatible and incompatible arguments must be presented together to activate these processes or whether they apply to congruent and incongruent arguments due to the nature of the evidence alone. The inclusion of mixed and non-mixed (one-sided) evidence allows for examination of potential differences.

This study further aims to examine whether belief updating behavior supports a motivated reasoning account or a Bayesian account of belief updating. This will be assessed by testing whether participants' beliefs change as a function of biased assimilation of the evidence, dependent on their prior beliefs, or whether belief change depends on the direction and/or merits of the evidence.

Methods

Participants

Participants were 124 students (75 female) enrolled in Psychology, Linguistics, or Cognitive Science courses at the University of California, San Diego (UCSD) participating as part of a course requirement. All participants provided informed consent, and procedures were approved by the Institutional Review Board (IRB) at UCSD. Participants were between 18 and 35 years old, with a mean age of 21. An additional two participants completed the survey, but their results were not included, either because their responses suggested they did not understand the rating scale (n=1), or because their age was greater than 35 years (n=1).

Materials

The study concerned six socio-political issues: abortion, animal testing, assisted suicide, climate change, the death penalty, and school uniforms. These issues were among the most popular topics covered on two debate websites, www.procon.org and idebate.org.

Attitude measurements: For each issue, a single policy statement was chosen for participants to rate in terms of how much they agree or disagree (e.g., "Animal testing should be banned."). This was followed by four position statements for each issue selected from two headings under "Points for" on the idebate.org archive, and two from "Points against." Participants responded to all five of these position statements, and these responses formed the initial attitude measurement. After the experimental treatment, participants again responded to five position statements per issue to form the subsequent attitude measurement.

Strength measurements: For each issue, participants were given four questions with a 9-point Likert scale to indicate how much they cared about, and had thought about, that issue. These four questions were combined to form a measure of strength of conviction.

Arguments: Using text from the websites, 6 supporting (Pro) and 6 opposing (Con) arguments were selected for each issue. Arguments were generally matched for content (i.e., if a Pro and a Con argument addressed the same point, both arguments were usually selected), and for length (mean argument length = 120 words, sd = 11). To create arguments of similar length, portions of longer arguments were omitted.

Procedure

The study included three phases: initial collection of attitude and conviction strength measurements, the presentation and rating of arguments, and the subsequent collection of attitude and strength measurements.

Initial collection of attitude and strength measurements proceeded one issue at a time, as participants first rated their attitude on the issue, and then responded to the questions regarding the strength of their convictions on that issue. The presentation order of the six issues was randomly determined.

Following the collection of attitude and strength measurements, each participant was asked to read and rate arguments for three randomly chosen issues from the original set of six. For these three issues, one was randomly designated as the Pro condition, such that the participant read and rated six arguments in support of the original position; one was randomly designated as the Con condition, such that the participant read and rated six arguments against the original position; and one was randomly designated as the Mix condition, such that the participant read and rated three arguments in support of the original position, and three arguments against. The order of the issues was randomized, as was the order of the arguments presented within each issue. Treatment thus included four treatment conditions: Pro, Con, Mix, and None, with the None condition comprising four issues for which participants were not presented any argument text.

After reading all arguments, participants were again asked to rate their positions on all six issues. Next, participants completed a brief political knowledge quiz to assess their political sophistication, and two questions to assess open-mindedness. Finally, they read a debriefing page that explained the goal of the study and provided links to the websites used for the argument texts.

Analysis

Opinions were scaled from -5 to 5, with -5 representing the opinion most against the issue and 5 representing the opinion most in favor of the issue (each issue is framed as a statement, e.g. "The death penalty should be banned."). Items where participants spent too long reading the argument text (more than 153 seconds, 3 standard deviations from the mean) were removed from analysis (28 items out of 2232).

Participants' prior opinions and strength of conviction were analyzed to ensure uniform representation across conditions, since within each issue, experimental conditions (Pro, Con, Mix, or None) were varied between subjects. A linear model of prior opinion as a function of treatment condition and issue showed that although opinions varied by issue, there were no significant differences among conditions (Pro, Con, Mix, None), nor was there any interaction of issue and condition. Similarly, strength of conviction did not vary as a function of treatment condition.

Models of argument rating were analyzed with a linear mixed effects regression (LMER) model using the lme4 package in R (Bates, Maechler, Bolker, Walker, et al., 2014; R Core Team, 2015). All experimental factors were allowed to interact initially; more complex models were compared with more parsimonious models using model ANOVA in R. Models were fit with random intercepts for subjects and items (viz. arguments). The reported models are those that included statistically significant predictors of argument rating and are not statistically different from more complex models (using cutoff $p < .01$).

Models of belief updating were analzed with a linear model in R. Again, all experimental factors were allowed to interact initially; more complex models were compared with more parsimonious models using model ANOVA in R. This is equivalent to selecting all predictors with a significant p value ($p < .01$) in the model ANOVA.

Results

The present study was designed (i) to replicate patterns of argument evaluation shown in other studies (Lord et al., 1979; Edwards & Smith, 1996; Taber et al., 2009) and (ii) to critically examine whether biased argument rating leads to belief updating, as suggested by a motivated account of reasoning, or whether belief change can be better explained by a Bayesian account in which participants are sensitive to the merits of the evidence.

The motivated cognition account explains attitude polarization as resulting from a biased assimilation of the evidence, such that evidence compatible with participants' initial positions is weighted more heavily than incompatible evidence, and consequently has a disproportionate impact on the way participants update their beliefs. We first assessed whether participants evaluated the arguments in a biased manner by analyzing whether their ratings of these arguments differed systematically as a function of their prior beliefs. Next, we assessed the factors that influenced belief change in response to these arguments.

Argument Rating

As noted above, our first question was whether participants rated evidence differentially as a function of its compatibility with their initial attitudes about the relevant issue. To examine this question, we began by modeling participants' argument ratings with a linear mixed effects model with predictors of treatment condition (Pro, Con, or Mix), argument polarity (Pro or Con), prior opinion, strength of conviction, issue, and political sophistication. Argument polarity is coded separately from Condition and represents Pro and Con arguments irrespective of which experimental condition they were presented in. The goal of this variable coding procedure was to separate potential effects of experimental condition from effects of argument polarity.

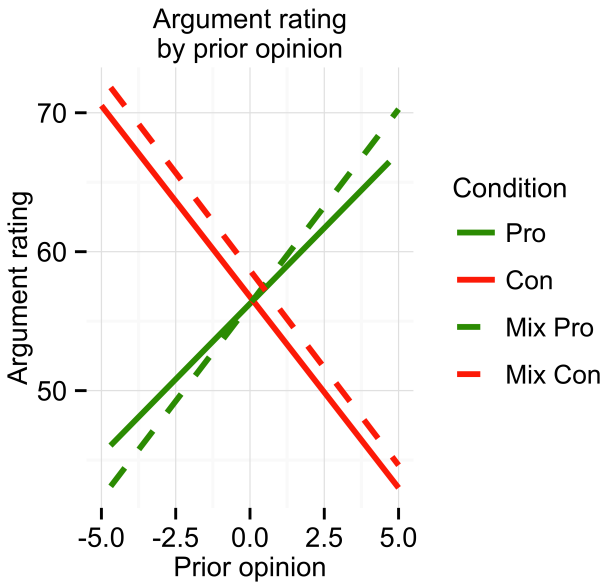

Figure 1: Average argument rating as a function of prior opinion (-5 most opposed, 5 most in favor of the issue). Green lines represent Pro arguments presented in the Pro and Mix conditions; Red lines represent Con arguments presented in the Con and Mix conditions.

Our model selection procedure revealed that experimental condition per se was irrelevant. The best model predicts argument rating as a function of prior opinion and argument polarity only. There was a trending further interaction with strength of conviction, with the slope of the rating x prior opinion line being steeper for participants with high strength of conviction ($p = .015$ for the 3-way interaction). Other experimental variables did not show main effects or interact with experimental variables. The mixed effects linear model includes random subject intercepts and individual argument intercepts. See Equation 1 and Table 1 for model results.

$$\text{Argument rating} \sim \text{prior opinion} * \text{argument polarity} \quad (1)$$

Table 1: Model results for Equation 1.

Factor	df	F value
Prior opinion	1	1.5
Argument polarity	1	0.3
Prior x Argument polarity	1	125.1

Figure 1 shows how argument ratings differ as a function of participants' prior opinions, with separate green regression lines shown for supporting arguments presented in the Pro condition and in the Mix condition, and separate red regression lines for opposing arguments presented in the Con condition and in the Mix condition. The positive slope of both green lines reflects the fact that the more participants support the issue, the higher they rate the Pro arguments compatible with their position. The similarity in the slope of the Mix and the Non-Mix line indicates that participants' ratings of these arguments were similar, regardless of whether they were presented in the context of other Pro arguments, or with a mixture of Pro and Con arguments. Likewise, the negative slope of both red lines reflects systematic bias in the ratings of opposing arguments, with opponents (-5 on the x-axis) rating those arguments higher than supporters (+5 on the x-axis), irrespective of whether opposing arguments were presented in a Con or a Mix block.

Belief Updating

We are interested in what factors lead to belief updating, or opinion change, after participants read and rate the arguments. Specifically, experimental condition might interact with participants' prior opinions, showing that belief updating due to different types of evidence (i.e., that presented in the Pro, Con, and Mixed conditions) differs as a function of their original position regarding that issue. Strength of conviction may also influence opinion change if participants whose beliefs are stronger are either more motivated to defend their position or rely on a greater body of knowledge to form their prior opinion. Because participants may change their opinions differently by issue, issue is also included as a predictor. Finally, we included a measure of political sophistication because previous studies have suggested that sophisticated individuals are more likely to engage in motivated reasoning (Taber et al., 2009).

Opinion change was modeled as a function of treatment condition (Pro/Con/Mix), prior opinion, strength of conviction, issue, and political sophistication. Linear models as described in the Analysis section were created to investigate the effects of these factors on opinion change. The best model to predict opinion change is shown in Equation 2.

$$\text{Opinion change} \sim \text{condition} + \text{prior opinion} * \text{strength} \quad (2)$$

Table 2: Model results for Equation 2.

Factor	df	Estimate	F value	p value
Condition	2		19.4	$< .001$
Prior opinion	1	-.41	96.2	$< .001$
Strength	1	-0.02	0.78	.38
Prior x Strength	1	0.03	7.55	$< .01$

The effect of experimental condition on opinion change is shown in Figure 2. On average, independent of prior beliefs, participants' opinions were more in favor of an issue after viewing and rating arguments in the Pro condition; more opposed to the issue after viewing and rating arguments in the Con condition; and unchanged after viewing arguments in the Mix condition.

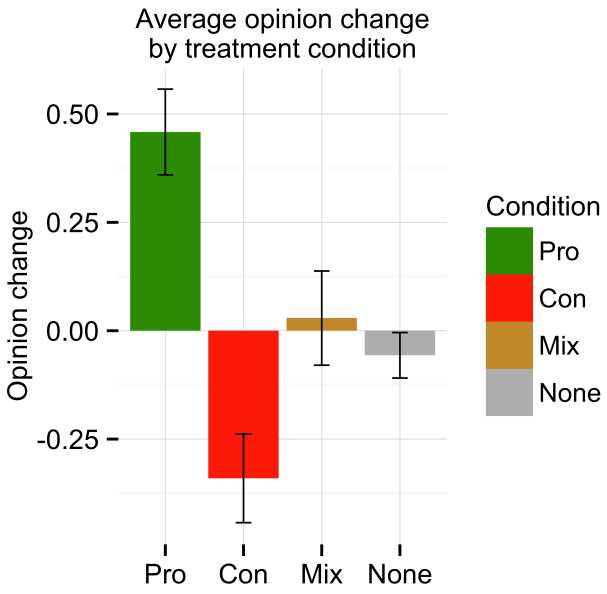

Figure 2: Average opinion change for each treatment condition. Lines represent standard error.

Overall, participants shifted their opinion toward a more moderate point of view (and also in the direction of the evidence), with participants more in favor of an issue changing their opinion to be less in favor, and those opposed changing their opinion to be more in favor. This center-trending behavior is represented in the negative coefficient of prior opinion in the model. Prior opinion further interacts with strength of conviction such that participants with lower strength show more center-trending than do those with higher strength of conviction.

The prior opinion x strength interaction is shown in Figure 3. Values for opinion change were baseline corrected by subtracting prior opinion * opinion change slope for the None condition to show how much opinion changed when participants viewed and rated arguments. This visually removes the overall center-trending pattern observed for all conditions. Participants with high strength of conviction did not show a difference in opinion change compared to baseline. Those with low strength of conviction show an additional center-trending pattern, with participants more in favor of an issue changing to be more opposed, and participants more opposed to an issue becoming more in favor.

Finally, we were interested in whether participants' argument ratings would influence their beliefs in addition to the other factors. Motivated cognition accounts would predict that participants who exhibit biased rating behavior will be more likely to polarize, updating their beliefs in the direction

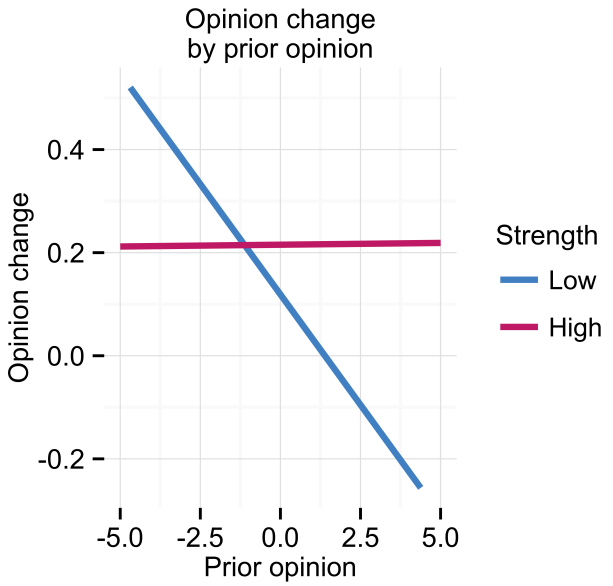

Figure 3: Interaction of prior opinion and strength. The blue line represents participants with low strength of conviction for a given issue, and the pink line represents those with high strength of conviction. Values have been corrected to remove the center-trending slope of the None condition to show their difference from baseline.

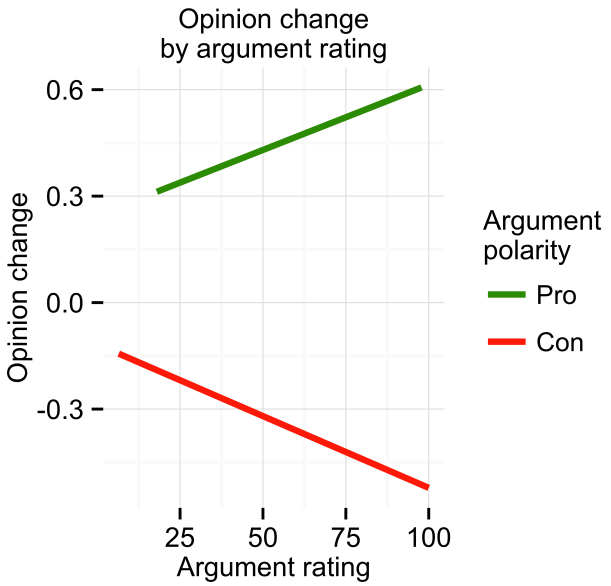

Figure 4: Interaction of argument rating and argument polarity. The red line represents average opinion change for Con arguments in the Con or Mix condition; the green line represents average opinion change for Pro arguments.

of their initial opinion. By contrast, Bayesian updating predicts that participants will rely only on the evidence. Consequently, they will either move in the direction of the evidence (irrespective of their prior beliefs), or maintain their original point of view.

Model comparison revealed that when participants' average argument rating was included as a predictor, the most parsimonious account of opinion change is given by the factors in Equation 3. As in Equation 2, the main effect of treatment condition and the interaction between prior opinion and strength of conviction were present. In addition to the previous predictors, opinion change is further predicted by an interaction of argument polarity and argument rating. As shown in Figure 4, this interaction term results because participants' opinions on average change to be more congruent with the position of those arguments that participants rated highly. The higher a given participant rated Pro arguments, the more their opinion changed in the positive direction. The higher they rated Con arguments, the more their opinion changed in the negative direction.

$$\text{Opinion change} \sim \text{condition} + \text{prior opinion} * \text{strength} + \text{argument polarity} * \text{argument rating} \quad (3)$$

Figure 4 shows this interaction of argument rating x argument polarity (Pro/Con). The occurrence of prior opinion and argument ratings in separate, additive terms in Equation 3 suggests that the relationship between argument rating and opinion change was independent of participants prior beliefs. That is, whether or not participants initially agreed with the policy embraced in a given argument, they changed their positions to be more congruent with the arguments, especially for highly rated arguments.

Discussion

Argument rating

Participants rated arguments that were compatible with their prior policy opinions as objectively better than arguments that were incompatible with those opinions. Moreover, this bias scaled linearly with participants' prior opinions, as those at either end of the scale showed the greatest bias in argument ratings. This argument rating bias is consistent with previous findings, potentially supporting the motivated reasoning account. However, these findings are also consistent with a Bayesian reasoning account in which participants at the ends of the scale are assumed to assign a high prior probability to their own position, and naturally assess the likelihood of congruent evidence to be higher than that of incongruent evidence. To dissociate motivated from Bayesian reasoning, it is necessary to examine the opinion change data.

Belief updating

The belief updating data provide support for a Bayesian account and show that even in the presence of biased argument ratings, participants changed their beliefs in response to the evidence. The final model of opinion change suggested that

for any given issue, participants' beliefs at the end of the experiment depended on three independent factors: treatment condition, an interaction between prior opinion and strength of conviction, and an interaction between argument polarity and argument rating. Whereas a motivated reasoning account predicts that treatment condition will interact with prior opinion, we instead found that condition had an independent effect. Participants who read Pro arguments adjusted their beliefs in a positive direction, those who read Con arguments adjusted their beliefs in a negative direction, and those in the Mix condition made almost no adjustment to their beliefs.

Further, while prior opinion was highly relevant for belief change, we found no evidence for the polarization phenomenon predicted by motivated reasoning. In fact, participants with weaker convictions moved a small amount away from their original positions, while those with strong convictions tended to maintain their existing beliefs.

Finally, the relationship between argument ratings and belief change was more consistent with a Bayesian account than the biased assimilation process predicted by motivated reasoning. That is, with motivated reasoning we would expect both highly-rated congruent arguments and low-rated incongruent ones to lead to opinion change in the direction of participants' prior opinions. Instead, we saw that highly-rated arguments, regardless of their congruency with participants' prior beliefs, were associated with movement in the direction of the arguments themselves. This is strong evidence in favor of a Bayesian account and shows that even in the presence of biased argument rating, belief change seems to be based on the quality of the evidence itself.

Acknowledgments

This research was supported by a grant from the Frontiers of Innovation Scholars Program (FISP) at UC San Diego.

References

Bates, D., Maechler, M., Bolker, B., Walker, S., et al. (2014). lme4: Linear mixed-effects models using eigen and s4. *R package version*, *1*(7).

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.

Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology*, *71*(1), 5.

Gerber, A., & Green, D. (1999). Misperceptions about perceptual bias. *Annual review of political science*, *2*(1), 189–210.

Jern, A., Chang, K.-M. K., & Kemp, C. (2014). Belief polarization is not always irrational. *Psychological review*, *121*(2), 206.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological review*, *94*(2), 211.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, *108*(3), 480.

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.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.

Pyszczynski, T., & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. *Advances in experimental social psychology*, *20*, 297–340.

R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.

Taber, C. S., Cann, D., & Kucsova, S. (2009). The motivated processing of political arguments. *Political Behavior*, *31*(2), 137–155.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.