UNIVERSITY OF CALIFORNIA SAN DIEGO

Modeling Attitude Change and Cognitive Bias for Social Issues

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Cognitive Science

by

Megan Dalene Bardolph

Committee in charge:

Professor Seana Coulson, Chair
Professor Andrea Chiba
Professor Marta Kutas
Professor Craig McKenzie
Professor Ed Vul

2019

The Dissertation of Megan Dalene Bardolph is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2019

DEDICATION

This dissertation is dedicated to my husband, Jeffrey Simpson; to our daughter, who accompanied me to my defense talk; to my friend Melissa Troyer, who encouraged me in academics and so many other areas of friendship and life; and to the many friends who shared this journey with me.

Thank you to my parents and other family members who believed in my success beyond what was reasonable; and to my advisor and additional mentors who modeled curiosity, compassion, and academic rigor.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Seana Coulson for her support as the chair of my committee. Her patience as I discovered my research interests and developed as a graduate student allowed me to succeed in spite of obstacles: she taught me, literally, that life is a journey, metaphorically.

I also acknowledge Marta Kutas and the members of her lab for teaching me a lasting sense of scholarly integrity. My experiences collaborating on research, discussing experiments and statistics, and socializing with this group have brought joy and fulfillment to my time as a grad student.

To Craig McKenzie, who helped me dive deep into the field of judgment and decision making through several classes and many long discussions; to Ed Vul, who helped me with statistical modeling; and to Andrea Chiba, who reminded me that affect is present in nearly every area of life, sincerest thanks.

I would also like to acknowledge the many undergraduate student researchers of the Brain and Cognition Lab who made this research a reality.

Chapter 3, in full, is a reprint of the material as it appears in Proceedings of the 39th Annual Conference of the Cognitive Science Society. Bardolph, Megan; Coulson, Seana. The dissertation author was the primary investigator and author of this paper.

VITA

2005 Bachelor of Science, Rose-Hulman Institute of Technology, Terre Haute, IN

2012-2017 Teaching Assistant, University of California San Diego

2011-2018 Graduate Student Researcher, University of California San Diego

2019 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Bardolph, M. D. & Coulson, S. (2018). The role of affective involvement and knowledge in processing mixed evidence for social issues. In T.T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 1317-1322). Madison, WI: Cognitive Science Society.

Bardolph, M. D. & Coulson, S. (2017). Belief updating and argument evaluation. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 1586-1591). London, UK: Cognitive Science Society.

Forgács, B., Bardolph, M. D., Amsel, B. D., DeLong, K. A., & Kutas, M. (2015). Metaphors are physical and abstract: ERPs to metaphorically modified nouns resemble ERPs to abstract language. *Frontiers in human neuroscience, 9, 28*. doi: 10.3389/fnhum.2015.00028

Bardolph, M., & Coulson, S. (2014). How vertical hand movements impact brain activity elicited by literally and metaphorically related words: an ERP study of embodied metaphor. *Frontiers in human neuroscience*, *8, 23*. doi: 10.3389/fnhum.2014.01031

ABSTRACT OF THE DISSERTATION

Modeling Attitude Change and Cognitive Bias for Social Issues

by

Megan Dalene Bardolph

Doctor of Philosophy in Cognitive Science

University of California San Diego, 2019

Professor Seana Coulson, Chair

Polarization is increasing in American politics and frequently involves disagreement over basic facts. Although this phenomenon is rare, the domain of social and political issues provides an environment where emotions may influence opinions and decisions in seemingly irrational ways. This research explores how individuals judge and respond to evidence about controversial socio-political issues, considering whether behavior is more appropriately modeled by accounts of motivated reasoning or Bayesian updating rules.

We find evidence of an attitude congruency bias, where people judge information to be of higher quality when it aligns with their existing attitudes on an issue. However, we find that this bias does not necessarily lead to polarizing. Instead, people's change in attitudes is better described by Bayesian updating, where people are sensitive to the amount and quality of information they are presented with. This behavior does not seem to be driven by affect or knowledge in a domain.

Judgment of information, in the form of argument rating, was analyzed by creating separate measures of objective argument quality and individual rating bias. Both factors were found to model attitude change, indicating that participants are sensitive to both the objective quality of evidence and to the effects of their own biases. Accounts of Bayesian information processing and motivated reasoning both predict behavior when information is modeled in terms of these two factors.

Finally, this research explores the role of information choice and shows that a bias toward choosing attitude-congruent information may lead to motivated attitude change, where exposure to a biased set of evidence models attitude change in line with one's existing views. People exhibit more sensitivity to information quality when they do not choose which information to view, indicating that choice may play a special role in allowing individual bias to outweigh information quality. These findings inspire questions about the role of curated information and people's capacity for rational behavior in a tense political climate.

Chapter 1
Rational models of attitude change for social issues

Polarization seems to be increasing in American politics (Alwin & Tufis, 2016; Layman, Carsey, & Horowitz, 2006), and frequently involves disagreement over basic facts (Kahan, 2016). For example, in a survey of American voters, 75% of self-identified liberals believed that climate change was due primarily to human activity, whereas only 45% of conservatives shared this belief (McCright & Dunlap, 2011). Although repeated discussion within a group with a strong sense of shared identity can lead individuals to hold more extreme attitudes (Sunstein, 2002), polarization is relatively rare (Kahan, Peters, Dawson, & Slovic, 2017). Because social and political issues are often where cultural identity intersects with factual knowledge, they provide an arena for polarization to arise (Kahan, 2016). Here we examine how individuals respond to information regarding controversial social issues, exploring whether or not the affectively charged nature of these issues compromises a rational response. We consider whether polarization on such issues is inevitable, or whether cognitive factors allow consensus to emerge.

Motivated reasoning

One explanation for polarization on social issues is motivated reasoning (Ditto & Lopez, 1992; Kunda, 1990). On such accounts, attitude polarization occurs because people with opposing views draw opposite conclusions from the very same evidence. In a classic study, Lord, Ross, and Lepper (1979) queried participants about their views on capital punishment, then presented them with the results of two studies: one that suggested the death penalty deters crime, and one that suggested the opposite

conclusion. Participants were asked to rate the quality of each study, and then to re-characterize their views on the death penalty. Interestingly, participants tended to rate the study that supported their own beliefs as objectively better than the one that undermined them, and each group adjusted their beliefs to more strongly favor their original position (Lord et al., 1979). According to motivated reasoning, polarization occurs because different processes are invoked to evaluate arguments compatible with prior beliefs versus those that are incompatible. In particular, participants readily accept attitude congruent arguments, while spending more time and mental resources arguing against incongruent ones (Edwards & Smith, 1996, Taber, Cann, & Kucsova, 2009).

There is reason to believe that even if in many instances people learn easily and naturally from their environment (see Nisbett & Ross, 1980, Pyszczynski & Greenberg, 1987), there are situations in which it is difficult to methodically test and revise hypotheses because of motivational factors, rather than cognitive limitations. For example, if an adult is assessing religious beliefs that have been an integral part of her life, it seems unlikely that she will start with a broad set of hypotheses, or even a single hypothesis, and then explore the possibilities of abandoning religious faith or seeking out new religious beliefs in response to careful consideration of evidence and data. Religious beliefs, along with other types of personal beliefs, may function differently from hypotheses such as “What happens when I let go of this cup?” because they are emotional and may be associated with a person’s sense of self (Kunda, 1990, Pyszczynski & Greenberg, 1987). Individuals may be motivated not to update these beliefs in the same way that other beliefs are updated.

Dissonant rationalizing

Cognitive dissonance theory (Festinger, 1957, Wicklund & Brehm, 1976), an early theory of motivated reasoning, is one possible framework for understanding attitude change. According to this theory, exposure to information contradicting a strongly held belief creates cognitive dissonance and a need to reduce or resolve the dissonance. The theory proposes that people are motivated to reduce dissonance just as we are motivated to eat: the more hungry people are, the more they seek out food. Similarly, the more cognitive dissonance people experience, the more they will seek to resolve that dissonance. People can resolve dissonance by seeking out belief-confirming information, changing their mind, or adopting additional thoughts to resolve the dissonance being experienced.

In an effort to elicit motivated self-justification of beliefs in response to contradictory information, Batson (1975) presented participants with one-sided evidence in a study about religious belief. The study was designed to test the prediction of cognitive dissonance that individuals possessing strong beliefs and commitment to those beliefs will strengthen instead of weaken their position when exposed to convincing evidence that their beliefs are wrong (Festinger, Riecken, & Schachter, 1956). Participants responded to surveys on their beliefs in religion and the infallibility of the Bible, then read an article claiming that religious scholars had proved writings in the Bible to be fraudulent. When participants who reported themselves as strongly holding religious beliefs were exposed to contradictory evidence, those who believed the evidence to be true paradoxically increased their commitment to their beliefs, while those skeptical of the evidence were persuaded to slightly decrease the intensity of their

beliefs (though not by a statistically significant amount). This indicates that people can "double down" on strongly held beliefs in response to belief-incongruent evidence.

In Batson's (1975) study, the stakes were relatively high for accepting the new information, involving a belief that could not be easily changed and was strongly tied to individual identity. It may be that when people claim to believe more strongly after viewing convincing counter-arguments, they are actually trying to counteract the doubt that they are experiencing. In this case, they are attempting to maintain their self-concept (see Ecker, Lewandowsky, Fenton, & Martin, 2014 and Gal & Rucker, 2010), in line with Festinger's theory of cognitive dissonance (Festinger, 1957). We will refer to this pattern of behavior, where people express increased intensity in their beliefs after exposure to high quality incongruent evidence, as *dissonant rationalizing*.

Bayesian updating

Bayesian updating serves as a normative alternative to motivated reasoning and dissonant rationalizing. According to a Bayesian model, new evidence is combined with prior belief to form a posterior belief. In the simplest version of updating a belief about the world, data are randomly sampled and there is no uncertainty about the validity or interpretation of the evidence. For example, a person may be interested to know whether it is raining outside, and one form of evidence may be the observation of other people carrying umbrellas. Observing this evidence makes the belief that it is raining outside stronger. Although the potentially emotional topics of interest explored here do not fit simply into a Bayesian model, this concept is often used as a normative standard, so an attempt is made at constructing such a model to contrast with other accounts of belief updating.

Bayes' rule (Bayes, 1763/1958) is a formulation of basic elements of probability theory, describing the joint probability of two variables in terms of the conditional probability of one variable and the marginal probability of the other. The joint probability of a and b can be written either as:

$$P(a,b) = P(a|b)P(b)$$

or

$$P(a,b) = P(b|a)P(a)$$

Combining these equations yields:

$$P(a|b)P(b) = P(b|a)P(a)$$

which can be rewritten as

$$P(b|a) = P(a|b)P(b)/P(a)$$

As described in (Griffiths, Kemp, & Tenenbaum, 2008), this formula can be used to represent hypotheses and data. If h represents a hypothesis about a phenomenon that gives rise to data, d, then an observer can learn about the probability of h vs. other processes, h', by observing the evidence, d. The posterior probability, P(h|d), or the degree of belief in the hypothesis given the observed data, can be expressed as

$$P(h|d) = P(d|h)P(h)/P(d)$$

P(h|d) is referred to as the posterior probability, and P(d|h) is called the likelihood. Likelihoods are used to update posterior beliefs from prior beliefs in light of how well the hypothesis predicts observed data. This principle can be used to model behavior, often serving as an updating rule for a Bayesian network of hierarchical priors with prior distributions over the priors (Griffiths, Kemp, & Tenenbaum, 2008; see also Good, 1980, Gelman et al., 1995, Lee, 2006, Tenenbaum, Griffiths, & Kemp, 2006).

When people are very knowledgeable in a particular domain, their knowledge serves as a strong prior, since evidence has already been accumulated. For example, scientists have conducted many experiments verifying the value of the maximum speed of light in the universe. Each study has contributed evidence that informs the exact value. However, not every experimental result is perfect, so findings are accepted or rejected based on the quality of the research and taking into account the existence of noise in estimates of the exact parameter value.

Within a Bayesian model, not all evidence should be treated equally. In addition to a hypothesis about the value of the speed of light, there are other hypotheses about the ways in which a study may be flawed. If a study were to produce a value of the speed of light very different from what is known, scientists' hypothesis about the value of the speed of light should not be updated at all; instead, they may adjust their hypothesis about the quality of the research and reject the result entirely. A 2011 study of neutrinos indicating that they traveled faster than the speed of light provides an interesting case study of this scenario. Many scientists were hesitant to accept the results based on their deviation from accepted theory; this hesitance was later justified when it was discovered that the results were due to faulty equipment (Amelino-Camelia, 2011; Stephens, 2015).

Accounts of Bayesian consideration of evidence even allow for evidence to be treated differently depending on whether it is in agreement with one's prior beliefs. If two people hold differing prior beliefs, they may treat the same piece of evidence differently (see Gerber & Green, 1999 for an account of this). However, unless there is reason to suspect the source or otherwise question the validity of the evidence, Bayesian

normative accounts still require that people's prior beliefs be updated in the direction of the evidence. This updating can be small, and can even be zero, but it cannot be in the opposite direction. Polarization, or adopting a more extreme version of one's current beliefs, can only occur when belief congruent evidence is encountered. Polarization in response to belief incongruent evidence is a violation of normative updating in this account.

A complication of using Bayesian updating as a model of beliefs, however, is the potentially complex nature of prior beliefs. The potentially vast amount of experiences and information that underlies people's belief structures may necessitate more complex models of a hypothesis space that is updated in response to evidence. Jern and colleagues (Jern, Chang, & Kemp, 2014) illustrate examples of simple Bayes nets that can represent beliefs comprised of more than just one hypothesis. Depending on the structure of the network, it is possible for the same piece of evidence to normatively give rise to divergent updating. This type of updating can occur without including additional motivations, affective or otherwise. For example, two people could read the same findings of a study indicating the effectiveness of the death penalty at deterring crime. Both people have differing prior beliefs about the effectiveness of the death penalty that will be updated according to the study findings. If those individuals also have differing beliefs about the consensus of the scientific community with regards to its effectiveness, then one person may assume that a study's results must be extremely convincing for it to be published, while another may believe that such studies get published even if the evidence isn't very good, simply because the findings agree with the general consensus (Jern et al., 2014). Because of their differing assumptions (representing different beliefs

in a hypothesis), these individuals can update their beliefs in different directions while still following Bayesian normative principles.

The present study

The present study investigated the relationship between attitude congruency bias and polarization by presenting participants with evidence-based arguments about affectively charged social issues, and examining how individuals changed their beliefs in response to those arguments. Because Bayesian accounts suggest people are sensitive to the amount of evidence they encounter, the amount of evidence was experimentally varied to see if participants responded in a way consistent with the normative account. To discriminate between motivated and Bayesian accounts, participants were exposed to mixed evidence on some issues and to one-sided evidence on others.

All accounts of attitude change predict polarization in the presence of attitude-congruent arguments, but differ in their predictions for how people respond to attitude-incongruent evidence. According to Bayesian models of information processing, people are sensitive to the quality and quantity of evidence in a way that depends on their priors. Differential judgment of evidence depending on its alignment with prior attitudes may reflect the degree to which people are able to integrate new information with extant knowledge, and does not necessarily lead to polarization.

Under the Bayesian model, the exclusive presentation of incongruent evidence would not lead participants to have more confidence in their initial beliefs. Participants should either change their beliefs in the direction of the evidence (and perhaps more so for high quality evidence) or fail to adjust their initial attitudes (if the evidence is not

persuasive). It may be that participants with strong prior attitudes already possess background knowledge and opinions that are considered along with new information; in this case, participants with stronger attitudes may be less persuaded by new evidence either congruent or incongruent with prior attitudes, and those with less strong attitudes will be more persuaded by new evidence.

Motivated reasoning accounts suggest participants may polarize when exposed to incongruent evidence judged to be of low quality, especially when presented alongside attitude-congruent evidence (Lord et al., 1979). Because motivated reasoning accounts suggest polarization depends on people's ability to refute attitude incongruent evidence, it may only occur among a subset of participants who feel strongly enough to engage in this process. Motivated reasoning suggests polarizing is most likely to occur when participants are presented with a mixture of evidence, to occur in participants who report attitudes toward the extremes of the attitude scale and/or those who indicate a strong emotional investment in the issue, and to be most pronounced in politically sophisticated participants who are knowledgeable enough to argue against incongruent information (see Taber et al., 2009).

The dissonant rationalizing account proposes that when the existence of high quality incongruent evidence threatens important aspects of their identity, participants will engage in self-justification of their beliefs and polarize. Similar to motivated reasoning, the dissonant rationalizing account suggests polarization is most likely for those participants who are highly committed to their stance on an issue; that is, they report attitudes towards the extremes of the attitude scale and indicate a strong emotional investment in the issue. The dissonant rationalizing account differs from

motivated reasoning in predicting polarization in response to high-quality attitude incongruent arguments (rather than low quality incongruent ones). This is because only high quality incongruent arguments are likely to activate the dissonant processes (Batson, 1975).

In the present study, participants first responded to a survey regarding their attitudes on six controversial topics. They then read arguments regarding three of those issues and rated the quality of each argument. Following the presentation of the arguments, participants once again responded to a survey regarding their attitudes on the six topics. Regression models were used to examine which variables were related to participants' argument ratings, with an emphasis on factors associated with an attitude congruency bias. Next, we examined whether this attitude congruency bias produced polarization and explored other factors associated with attitude change.

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 found 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 agreed or disagreed (e.g., “Animal testing should be banned.”). Below we refer to the “Pro” side of an issue as the one that was in agreement with this policy statement, and the “Con” side as contrary to the policy statement. 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 were combined to form the initial attitude measurement. After the experimental treatment, participants again responded to five position statements per issue to form the subsequent attitude measurement.

Affective involvement 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. Responses to these four questions were combined to form a measure of affective involvement.

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 edited.

Procedure

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

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

Following the collection of attitude and affective involvement 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 statement; one was randomly designated as the Con condition, such that the participant read and rated six arguments against the original position statement; and one was randomly designated as the Mix condition, such that the participant read and rated three Pro arguments and three Con argument. The order of the issues was randomized, as was the order of the arguments presented within each issue. Treatment thus included four conditions: Pro, Con, Mix, and None, with the None condition comprising issues for which participants were not presented any argument text. Participants rated each argument from “Weak” to “Strong” using a slider bar. Numeric values ranged from 0 to 100 and were not visible.

After reading all arguments, participants were again asked to rate their attitudes on all six issues. Next, participants completed a brief political knowledge quiz to assess their political sophistication, and two questions to assess open-mindedness. Participants' political sophistication was represented by their scores on the political knowledge quiz, ranging from 0 to 5 (number correct out of 5 items). Finally, participants read a debriefing page that explained the goal of the study and provided links to the websites used for the argument texts.

Analysis

Attitudes were scaled from -5 to 5, with -5 representing the opinion most opposed to the issue statement (e.g. "The death penalty should be banned.") and 5 representing maximal agreement with the issue statement. Affective involvement ranged from 0 (least involvement) to 5 (most involvement). Political sophistication ranged from 0 to 5. Items where participants spent too long reading the argument text (more than 3 standard deviations from the mean, 153 seconds,) were removed from analysis (28 items out of 2232).

Participants' prior attitudes and affective involvement were analyzed to ensure uniform representation across conditions, since within each issue treatment (Pro, Con, Mix, or None) was varied between subjects. A linear model of prior attitude as a function of treatment condition and issue showed that although attitudes varied by issue, there were no significant differences among conditions (Pro, Con, Mix, None), nor was there any interaction of issue and condition. A linear model of affective involvement similarly showed an effect of issue and not experimental condition. Mean values and standard deviation for both prior attitude and affective involvement are shown in Table 1.1.

Table 1.1: Prior attitude and affective involvement average values by condition and issue.

Issue	Condition	Mean prior attitude (sd)	Mean affective involvement (sd)
Abortion	None	1.71 (2.53)	6.36 (1.70)
	Pro	2.09 (2.79)	7.02 (1.38)
	Con	2.81 (1.73)	6.53 (1.81)
	Mix	1.36 (2.46)	6.14 (1.27)
Animal testing	None	-0.05 (2.03)	4.92 (1.61)
	Pro	0.17 (2.67)	5.19 (1.78)
	Con	-0.11 (2.33)	4.62 (1.33)
	Mix	-0.16 (2.06)	5.29 (1.63)
Assisted suicide	None	0.46 (2.50)	4.88 (1.63)
	Pro	-0.41 (2.38)	4.83 (1.41)
	Con	-1.11 (1.91)	4.98 (1.00)
	Mix	0.59 (1.90)	4.06 (1.69)
Climate change	None	2.46 (1.45)	5.76 (1.74)
	Pro	2.71 (1.10)	6.13 (1.67)
	Con	1.97 (2.47)	5.72 (1.97)
	Mix	2.25 (1.27)	5.42 (2.13)
Death penalty	None	0.32 (2.09)	4.72 (1.44)
	Pro	0.06 (2.15)	4.18 (1.32)
	Con	0.15 (2.26)	4.52 (1.68)
	Mix	-0.25 (2.68)	5.63 (1.49)
School uniforms	None	-0.10 (1.95)	3.70 (1.77)
	Pro	-0.76 (2.14)	3.96 (1.86)
	Con	0.06 (2.23)	3.58 (1.52)
	Mix	-0.47 (2.37)	3.10 (1.34)

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 and backward model selection was used to determine the best-fitting model. Accordingly, more complex models were compared with more parsimonious models using ANOVA in R. Models were fit with random intercepts and slopes for subjects and items (viz. arguments). The reported models are those that included statistically significant

experimental predictors of argument rating and were not statistically different from more complex models (generally using cut-off $p < .01$, but trending predictors are also reported).

Models of attitude change were analyzed 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 approximately equivalent to selecting all predictors with a significant p value in the model ANOVA.

For all models, argument polarity (Pro/Con) and argument congruency (Congruent/Incongruent) were sum coded; experimental condition (Pro/Con/Mix) was treatment coded, with the Mix condition set as the baseline. Issue (Abortion/Assisted suicide/Animal testing/Climate change/Death penalty/School uniforms) was also treatment coded. Climate change was used as a baseline for Issue because participants' average attitude change for this issue was close to zero.

Results

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 attitudes. Next, we assessed the factors that influenced attitude change in response to these arguments.

Argument Rating

Were participants' argument ratings biased by their initial attitudes about the issues? 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 attitude, affective involvement, issue, and political sophistication. Political sophistication was included because previous studies have suggested that sophisticated individuals are more likely to engage in motivated reasoning (Taber et al., 2009). Argument polarity was coded separately from treatment condition, and represents Pro and Con arguments irrespective of which condition (Pro, Con, or Mix) they were presented in. The goal here was to separate potential effects of experimental condition from those of argument polarity; that is, whether (for example) a given Pro argument would elicit different ratings when presented in the Pro condition, and accompanied by five other Pro arguments, versus the Mix condition, where it was accompanied by three Con arguments and two other Pro arguments.

Our model selection procedure revealed that experimental condition did not influence the ratings of individual arguments. The best model predicts argument rating as a function of prior attitude and argument polarity only (see Equation 1.1 and Table 1.2 for model results). This linear mixed effects model includes Prior attitude x Argument polarity random slopes and Prior attitude and Argument polarity intercepts by subject, as well as random intercepts for subjects and items. Arguments in favor of the Pro side of an issue were rated more highly by participants whose prior attitudes were closer to the Pro side, and rated lower by those with attitudes closer to the Con side. Likewise, arguments for the Con side of an issue were rated more highly by participants with attitudes closer to the Con side and lower by those with attitudes closer to the Pro side of an issue.

There was a trend for a 3-way interaction of prior attitude, argument polarity, and affective involvement ($p = .063$), with the slope of the Prior attitude x Argument polarity

interaction being steeper for participants with high affective involvement. None of the other factors reached significance, either as main effects or in interaction with other variables.

Argument rating ~ Prior attitude x Argument polarity (1.1)

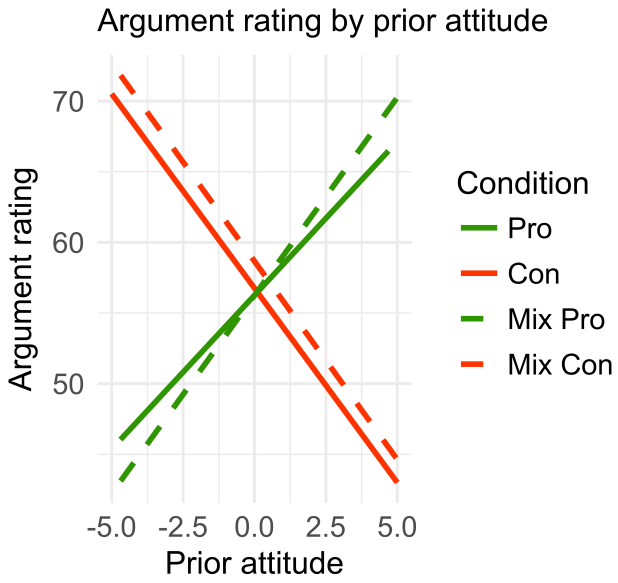

Figure 1.1: Average argument rating as a function of prior attitude (-5 most opposed, 5 most in favor of the position statement). Green lines represent Pro arguments presented in the Pro and Mix conditions. Red lines represent Con arguments presented in the Con and Mix conditions.

Table 1.2: Model results for Equation 1.1

Factor	Estimate	Std. Error	df	t-value	p-value
Intercept	56.75	1.15		49.0	< .001
Prior attitude	-0.16	0.26	1	-0.62	.536
Argument polarity	-0.78	1.02	1	-0.76	.449
Prior attitude x Argument polarity	2.43	0.30	1	8.23	< .001

Figure 1.1 shows how argument ratings differed as a function of participants' prior attitudes, 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 statement, 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 Con arguments, with opponents (-5 on the x-axis) rating attitude-compatible Con arguments higher than supporters (+5 on the x-axis), irrespective of whether opposing arguments were presented in a Con or a Mix block.

Attitude Change

Results above reveal an attitude congruency bias in participants' ratings of the arguments. Ultimately, however, our question is whether and how this bias leads to attitude change. To assess this question, we created linear models to predict attitude change from experimental treatment controlling for other relevant factors. We also used participants' argument ratings as a predictor of attitude change.

Initial examination of the data suggested that participants' attitude change from their Prior to Post-survey attitude was toward the center of the attitude scale. Even in the control condition in which participants did not read any arguments about the issue, participants on the Pro side of an issue had post-survey attitude responses that were less supportive, and participants on the Con side had post-survey attitude responses that were less opposed to the issue statement. This de-polarization behavior presumably reflects regression toward the mean. To correct for this, we used the control

condition to compute the change from pre- to post- survey attitude measurements in each issue. This average slope was then subtracted from the post-survey attitude measurements on that issue in the experimental conditions (Pro, Con, and Mix) to derive corrected measurements of attitude change. All analyses below utilized these corrected measurements.

Attitude change can be coded either as the degree to which participants polarized (moved further toward the extreme of the attitude scale in the direction of their prior attitude) or as change toward the Pro or Con position of an issue. We selected the appropriate coding for each account of attitude change in order to determine which one best describes the behavioral data. First we describe the analysis in terms of the Polarization variable.

Polarization analysis. Polarization was calculated as change in the direction of one's prior attitude (positive) or toward the opposing position (negative). For example, if a participant reported a prior attitude of 1.5 (slightly Pro) and a post-survey attitude of 4 (more Pro), their Polarization score would be 2.5, representing the presence of polarizing. If this same participant's post-survey attitude were -1.5 (slightly Con), their Polarization score would be -3, representing a moderation or anti-polarization of their initial attitude. The Polarization variable thus ranged from -10 to 5.

As noted above, there are two slightly different accounts that suggest a relationship between biased argument rating and polarization: motivated reasoning and dissonant rationalization. According to motivated reasoning, polarization results when participants over-weigh congruent arguments and under-weigh incongruent ones. This model predicts polarization will be modeled by an interaction of argument rating and

congruency that results because attitude change occurs when congruent arguments are rated high and incongruent arguments are rated low. If both of these behaviors are necessary for polarizing, we might expect a further interaction with Condition, as Polarization would only occur in the Mix condition in which participants encounter both congruent and incongruent information.

The dissonant rationalizing account of polarization predicts that highly-rated incongruent arguments will lead to polarizing in conditions where one-sided evidence is presented. Here we would expect that the Pro or Con conditions might produce evidence of polarization for participants who rated incongruent arguments highly. This account suggests that participants in the Pro and Con conditions might polarize when they read highly rated incongruent arguments, although this behavior may only be apparent in those participants with the highest degree of affective involvement.

Also of interest is whether the tendency to rate congruent arguments higher than incongruent arguments in the Mix condition results in polarizing; that is, is the presentation of mixed evidence different from presentation of one-sided evidence? We began by constructing a linear model predicting polarization from Argument rating, Congruency, and Condition. Political sophistication, Affective involvement, and Issue were included in the model and allowed to interact with the other predictors. Backward model comparison yielded the best model of Polarization, shown in Equation 2. There is an Argument rating by Congruency interaction accompanied by an interaction of Condition and Congruency, with a trending additive effect of Issue ($p = .026$). To understand the Condition x Congruency interaction, the data were split with the Mix condition analyzed separately and the Pro and Con conditions modeled together.

Polarization ~ Argument rating x Congruency + Condition x Congruency (1.2)

Table 1.3: Model results for Equation 1.2

Factor	Estimate	Std. Error	df	t-value	p-value
Intercept	-0.296	0.196		-1.51	.132
Argument rating	0.000	0.003	1	-0.10	.923
Congruency	-0.908	0.196	1	-4.63	< .001
Argument rating:Congruency	0.016	0.003	1	4.92	< .001
Condition			2		
(Pro)	0.146	0.116		1.27	.021
(Con)	0.295	0.115		2.56	.001
Condition x Congruency			2		
(Pro)	0.520	0.116		4.50	< .001
(Con)	0.380	0.115		3.32	< .001

For mixed evidence, motivated accounts predict that biased rating of arguments leads to polarizing. These accounts claim that congruent and incongruent evidence are judged by different standards, with congruent evidence being accepted while incongruent evidence may be critically examined or mentally argued against. Under these accounts, we would expect that both highly rated congruent arguments and low rated incongruent arguments can lead to polarizing. This would be shown by an interaction of argument rating and congruency, with highly rated congruent arguments and low rated incongruent arguments leading to positive polarization values. Some participants' attitudes may remain unchanged, but motivated accounts do not generally predict de-polarizing, or negative values of polarization in response to mixed evidence.

The best model of polarization for the Mix condition is shown in Equation 1.3. There was a significant Argument rating x Congruency interaction. Figure 1.2 shows how participants' average argument ratings model polarization for congruent and incongruent arguments. While the trend lines are in the direction predicted by motivated accounts of polarization, positive polarization does not seem to be present within the

confidence interval bounded by the standard error regions of the plot. Instead, only de-polarization is significantly modeled by both highly rated incongruent arguments and low rated congruent arguments.

Polarization ~ Argument rating x Congruency (1.3)

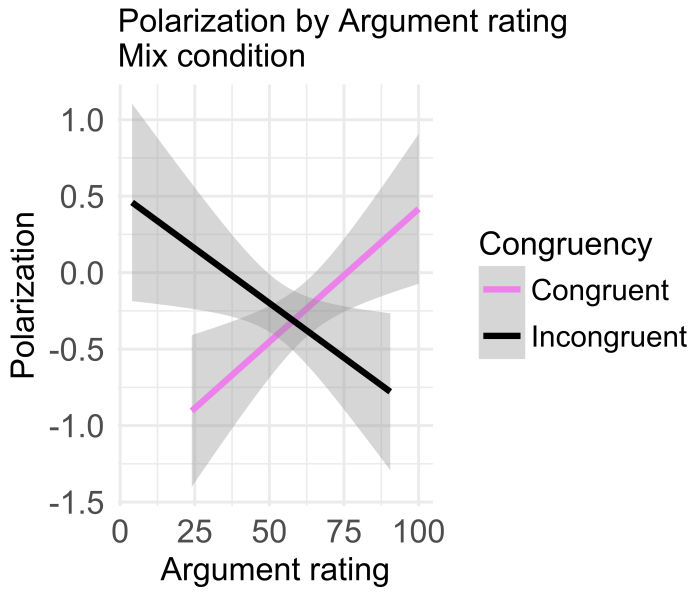

Figure 1.2: Polarization as a function of argument rating (0 weak, 100 strong) for the Mix condition. Pink line represents arguments congruent with participants' initial attitudes. Black line represents arguments incongruent with participants' initial attitudes.

For one-sided evidence, Bayesian accounts of attitude change predict that participants' attitudes will change in the direction of the evidence presented, not in the opposite direction. In terms of the Polarization variable, positive polarization should only be present when participants viewed attitude-congruent arguments. "Polarization" may be greater when the quality of the evidence is judged to be higher. De-polarization (a negative polarization value in this model) is predicted in the presence of attitude-incongruent arguments, which may be enhanced for higher-quality evidence (those arguments receiving higher ratings).

There was no significant difference by Condition of the Pro/Con data together, so the data were combined in a single model. The linear model that best predicted polarization is shown in Equation 1.3 above: the same model was found for the Pro/Con data and the Mix data.

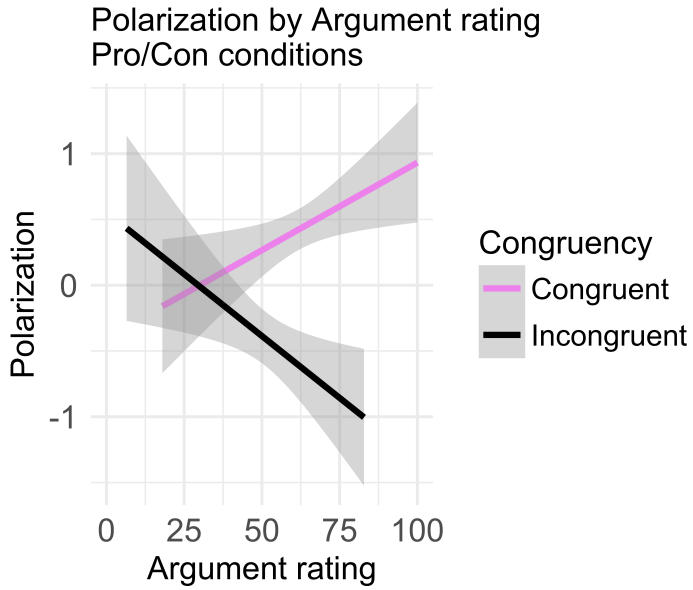

Figure 1.3: Polarization as a function of argument rating (0 weak, 100 strong) for Pro and Con conditions. Pink line represents arguments congruent with participants' initial attitudes. Black line represents arguments incongruent with participants' initial attitudes.

Figure 1.3 shows the nature of the Argument rating x Congruency interaction for the one-sided data. The highly rated congruent arguments are associated with increased polarization, while highly rated incongruent arguments are associated with enhanced de-polarization. Positive polarization was not present for either low or high ratings of incongruent arguments. These patterns better match Bayesian or information-processing accounts of attitude change. Instead of polarization/depolarization, it may make more sense to interpret this data as attitude change, in line with Bayesian or information-processing accounts of attitude change.

Attitude change analysis. To better describe the data in terms of Bayesian accounts of information processing and attitude change, we next constructed models posed in terms of attitude change instead of polarization. Attitude change was calculated simply as participants' (corrected) post-survey attitude minus their pre-survey (prior) attitude. Positive values represented change more in favor of an issue statement, and negative values represented change more opposed to it. As in the Polarization analysis above, the control condition was used to correct for regression to the mean in Attitude change in the experimental conditions. Because Prior attitude is incorporated in the Attitude change variable (Post-survey attitude – Prior attitude), it was not included in the model of Attitude change. However, in order to potentially distinguish the behavior of those holding extreme positions from behavior of those holding moderate positions, we created a categorical variable, Position, ranging from -2 to 2, representing participants' relative position with respect to each issue. Data were split into approximate quintiles while preserving the same spacing on the positive and negative portions of the attitude scale (-5 to -2.5, -2.5 to -1, -1 to 1, 1 to 2.5, 2.5 to 5). Instead of coding argument congruency as a single term, we represented rating behavior by including argument rating, argument polarity, and participant position in the linear model predicting attitude change.

Under a Bayesian or information-processing account, participants will change their attitudes in line with the evidence, especially when the evidence is of high quality. We therefore expect that participants in the Pro condition will on average show attitude change toward the Pro position on an issue, those in the Con condition will show change toward the Con position, and those in the Mix condition may not show attitude

change overall. This attitude change may be modulated by the quality of arguments: in every condition, highly rated arguments would be predicted to lead to greater attitude change in the direction of the arguments. This would be modeled as a 2-way interaction of Rating and Argument polarity, where highly-rated arguments lead to attitude change in the direction of Argument polarity independent of participants' position on an issue.

By contrast, motivated accounts predict a 3-way interaction among argument rating, argument polarity, and position, where highly rated arguments congruent with participants' positions lead to attitude change in the direction of the arguments, and low rated incongruent arguments lead to attitude change in the opposite direction of the arguments (i.e. in the direction of the participant's position). This pattern may further modulated by condition, arising more prominently in the Mix condition, and by participants' levels of affective involvement with the issues, being more evident in participants who feel the most strongly.

A linear model was created predicting corrected attitude change with factors of Argument rating, Argument polarity, Position, treatment Condition (Pro/Con/Mix), Affective involvement, and Political sophistication. Backward model comparison was used to compare more complex models to more parsimonious models retaining significant predictors. The best model of attitude change included an Argument rating x Argument polarity interaction, a Condition x Position interaction, and an interaction of Position and Issue. This model is shown in Equation1.4.

Attitude change ~ Argument rating x Argument polarity + Condition x Position + Position x Issue (1.4)

Table 1.4. Partial model results for Equation 1.4

Factor	Estimate	Std. Error	df	t-value	p-value
Intercept	-0.866	0.499		-1.73	.084
Argument Polarity	2.054	0.381	1	5.40	< .001
Argument rating	0.022	0.004	1	5.11	< .001
Argument rating:Polarity	-0.036	0.006	1	-5.69	< .001
Condition			2		
(Pro)	0.095	0.241		0.39	0.695
(Con)	-0.656	0.240		-2.74	0.006
Position			4		
(-2)	1.888	0.857		2.20	0.028
(-1)	-0.057	0.336		-0.17	0.865
(1)	-0.530	0.462		-1.15	0.251
(2)	-0.988	0.475		-2.08	0.038
Issue			5		
(Abortion)	-0.761	0.500		-1.52	0.129
(Assisted suicide)	-0.455	0.488		-0.93	0.352
(Animal testing)	0.193	0.461		0.42	0.676
(Death penalty)	-0.632	0.490		-1.29	0.198
(School uniforms)	-0.310	0.462		-0.67	0.503

As shown in Figure 1.4, attitude change following highly rated arguments is on average more in line with the arguments. There appears to be change in the opposite direction of arguments for low rated Pro arguments, but this pattern is primarily due to behavior in the Mix condition, where Pro arguments are also viewed along with Con arguments, meaning some of the change may be due to highly rated Con arguments.

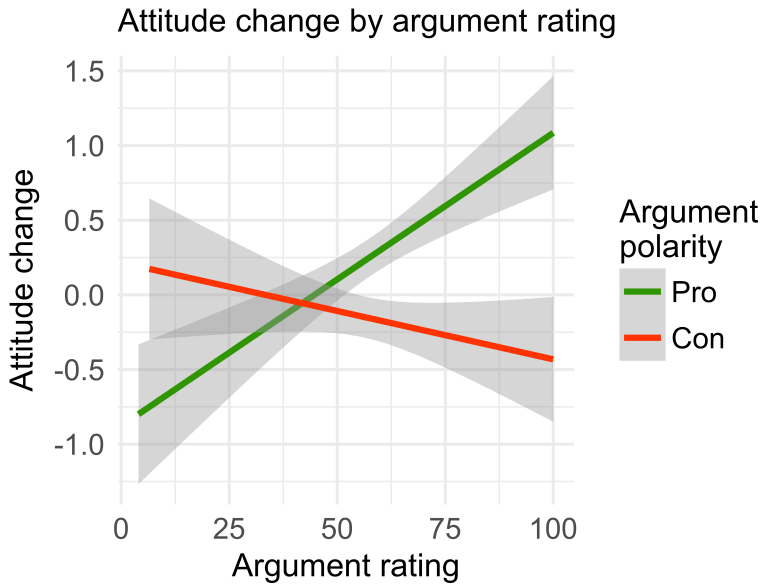

Figure 1.4: Interaction of argument rating and argument polarity. The red line represents average attitude change following Con arguments in the Con or Mix condition. The green line represents average attitude change following Pro arguments.

There is a Condition x Position interaction, with participants in the Mix condition showing de-polarizing behavior: those initially most in favor of the issue reported attitudes less in favor, and those most opposed reported attitudes less opposed to the issue after viewing arguments. This interaction is shown in Figure 1.5. There is no significant Condition x Position interaction when the Pro and Con conditions are modeled separately, although there is a main effect of Condition, with participants in the Pro condition changing their attitude in the Pro direction and those in the Con condition changing their attitude in the Con direction.

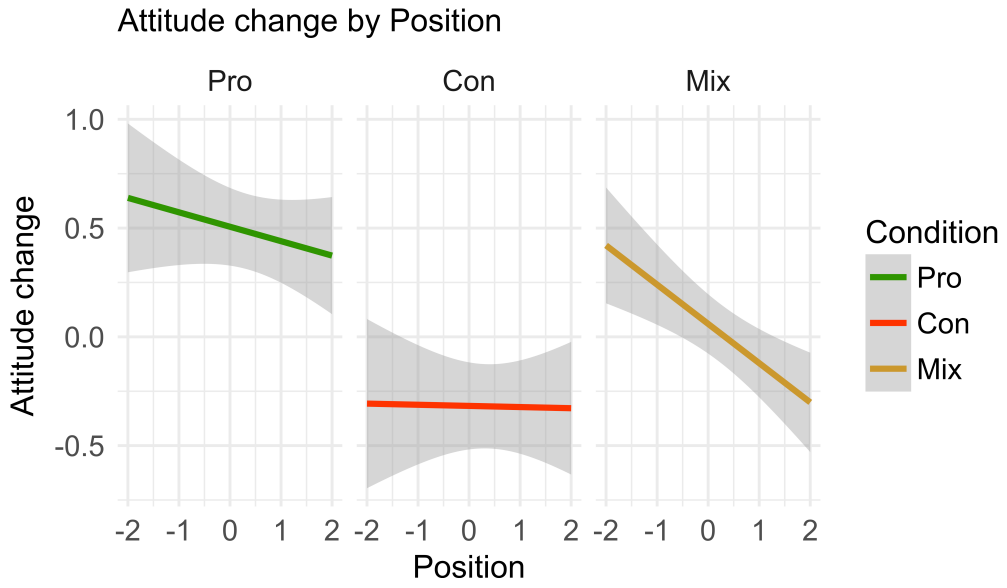

Figure 1.5: Attitude change as a function of Position. Position represents participants' initial attitudes, from -2 (most opposed) to 2 (most in favor of an issue statement). Position is modeled as a categorical variable, but is plotted as a numerical range to show trends. Lines represent the Pro (green), Con (red), and Mix (yellow) conditions.

The Position x Issue interaction results because attitude change varied depending on the issue presented and participants' initial attitudes. Individual estimates for each issue relative to Climate change (where participants showed the least amount of attitude change) and position relative to 0 are shown in Table 1.5.

Table 1.5. Continued model results for Equation 1.4: Condition x Position and Issue x Position interaction coefficients

Factor	Estimate	Std. Error	df	t-value	p-value
Condition:Position			8		
(Pro : -2)	-0.218	0.378		-0.58	0.565
(Pro : -1)	0.490	0.364		1.35	0.179
(Pro : 1)	1.066	0.313		3.41	< 0.001
(Pro : 2)	0.209	0.324		0.64	0.519
(Con: -2)	-0.242	0.395		-0.61	0.540
(Con: -1)	-0.120	0.353		-0.34	0.734
(Con: 1)	0.963	0.318		3.03	0.003
(Con: 2)	0.294	0.324		0.91	0.364
Issue:Position			19		
(AB: -2)	0.185	0.949		0.19	0.846
(AB: -1)	1.234	0.789		1.57	0.118
(AB: 1)	0.622	0.584		1.06	0.288
(AB: 2)	1.291	0.548		2.36	0.019
(AS: -2)	-0.506	0.906		-0.56	0.576
(AS: -1)	1.192	0.444		2.68	0.008
(AS: 1)	-0.079	0.551		-0.14	0.886
(AS: 2)	0.852	0.649		1.31	0.189
(AT: -2)	-2.096	0.875		-2.39	0.017
(AT: -1)	-0.243	0.414		-0.59	0.557
(AT: 1)	-0.289	0.538		-0.54	0.592
(AT: 2)	0.770	0.552		1.40	0.164
(DP: -2)	-1.528	0.900		-1.70	0.090
(DP: -1)	0.379	0.434		0.87	0.383
(DP: 1)	0.124	0.557		0.22	0.824
(DP: 2)	0.670	0.590		1.14	0.257
(SU: -2)	-1.507	0.881		-1.71	0.088
(SU: 1)	-0.588	0.535		-1.10	0.272
(SU: 2)	0.766	0.614		1.25	0.213

Discussion

The present study was designed to replicate patterns of argument evaluation shown in previous studies (Edwards & Smith, 1996; Taber et al., 2009), and to critically examine whether biased argument rating leads to the polarization of attitudes. The latter is suggested by a motivated account of reasoning, and contrasts with a Bayesian

account in which participants are sensitive to the merits of the evidence. Although our participants produced biased argument ratings, they did not polarize. In keeping with Bayesian accounts, attitude change was largely a function of the number of arguments participants read and their perceived quality. When attitude incongruent arguments were considered credible, participants changed their attitudes in response to them.

Argument Rating

Participants rated arguments that were congruent with their prior policy opinions as objectively better than arguments that were incongruent with those opinions. Moreover, this attitude congruency bias scaled linearly with participants' prior opinions, as those at either end of the scale showed the greatest bias in argument ratings. Edwards and Smith (1996) similarly showed that argument strength ratings were correlated with participants' prior beliefs for most of the issues in their study. Similarly, Taber and colleagues (2009) found that prior attitude significantly predicted an attitude congruency bias when participants read long arguments (with the same pattern trending for short and two-sided arguments). The findings of the present study are in line with prior work, potentially supporting a motivated account of argument strength rating. However, in conflict with accounts that suggest motivated reasoning processes are triggered by the presentation of a mixture of attitude congruent and incongruent information (Tabor, et al., 2009), our participants' argument ratings showed similar patterns in response to mixed and one-sided evidence. The present study thus suggests that attitude congruency bias does not require the presentation of mixed evidence to arise.

In fact, attitude congruency bias might be interpreted as consistent with a Bayesian reasoning account in which participants at the ends of the scale are assumed to assign a high likelihood to data consistent with their own position. In terms of cognitive models, prior probabilities may represent background knowledge on a topic, which serves as a standard by which new information can be judged. (Griffiths, Kemp, & Tenenbaum, 2008). Evidence that is incompatible with a well-informed attitude may be correctly judged as weak (see Gerber & Green, 1999), although there is some debate regarding the appropriateness of using one's prior beliefs to judge the quality of evidence (cf. Koehler, 1993, Kahan, 2016). To dissociate motivated from Bayesian reasoning, it is necessary to examine the attitude change data.

Attitude Change

Our initial analysis indicated that while attitude congruent evidence did indeed lead to polarization, incongruent evidence did not. Contrary to the predictions of dissonant rationalizing, strongly rated incongruent evidence led to depolarization (see Figures 1.2 and 1.3), that is, attitude change *away* from participants' original positions. Moreover, while attitude change in response to mixed evidence was somewhat in keeping with the predictions of motivated reasoning, showing a trend for polarization in response to weak incongruent evidence, attitude change never rose significantly above zero (Figure 1.2). So while our participants did not polarize in response to attitude incongruent evidence that was judged weak, they did de-polarize in response to incongruent evidence that was judged to be strong (Figures 1.2 and 1.3).

Our failure to find polarization in response to attitude incongruent arguments thus contrasts with previous accounts of polarization following exposure to mixed evidence

(Lord et al., 1979, Taber et al., 2009). This may be due to methodological differences between the studies, particularly in the administration of pre- and post- treatment attitude measurements. However, when the original study by Lord and colleagues was replicated with a similar before and after scale, findings of polarizing were not observed (Miller, McHoskey, Bane, & Dowd, 1993). Taber et al. (2009) point out that polarizing only arises under specific circumstances, and that highly affective content is needed to activate the motivated processes that lead to polarizing. Our failure to find polarization may be because our materials did not sufficiently evoke affective responses from our participants.

In fact, attitude change observed in the present study was largely a function of the quantity and the quality of the arguments presented to the participants. As outlined in Equation 1.4, three independent factors contributed to attitude change. The first involved argument rating and argument polarity; the second involved experimental condition and participant position; and, the third involved participant position and issue. The interaction of argument rating and argument polarity results because participants responded to highly rated arguments by moving in the direction of those arguments. Regardless of their positions, participants responded to highly rated Pro arguments by moving in the Pro direction, and to highly rated Con arguments by moving in the Con direction (see Figure 1.4).

The interaction of condition and position is illustrated in Figure 1.5, and generally reflects attitude change in the direction of the evidence. Participants in the Pro condition were slightly more supportive of the issue statement after reading six Pro arguments, with the strongest opponents changing more than those who already favored the Pro

position. Participants in the Con condition were slightly more opposed to the issue statement after reading six Con arguments. Participants in the Mix condition, that is, those who viewed 3 arguments on either side tended to de-polarize, with opponents moving in the Pro direction and supporters moving in the Con direction. Finally, the interaction of position and issue merely reflects the fact that participants de-polarized more for some issues than others.

As noted above, while prior attitude (position) was relevant for attitude change, we found no evidence for the polarization phenomenon predicted by motivated reasoning. Instead we found that for a subset of issues, participants showed a moderating response to the evidence; that is, their attitudes shifted toward the center of the attitude scale. This is in line with Bayesian models of beliefs in which prior distributions over belief in hypotheses are revised in response to new evidence (see Griffiths, Kemp, & Tenenbaum, 2008 for an overview).

Griffiths et al. (2008) describes cognitive models where inferences are made in light of background knowledge, allowing for inductive reasoning. People can learn distributions of data in an environment by comparing competing hypotheses with new information. New information allows hypotheses to be revised, and in the case of hierarchical priors (prior distributions over other distributions), revision of one hypothesis can lead to the revision of additional hypotheses. Although these models do not specifically cover revisions of belief about social issues, the principles could reasonably apply to this domain, since individual beliefs may be composed of several related hypotheses, and new information can help distinguish among competing hypotheses about the structure of the socio-political environment.

Conclusion

The present study suggests the relationship between argument ratings and attitude 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 were associated with movement in the direction of the arguments themselves, regardless of their congruency with participants' prior beliefs. This is evidence in favor of a Bayesian account and shows that in spite of the biases evident in participants' ratings of arguments, their attitude change seems to be based on the evidence presented.

Chapter 2
The role of active arguing in predicting attitude change

Study 1 revealed an attitude congruency bias, but this bias did not lead to polarization of participants' attitudes. The present study replicates and extends Study 1 by adding participant text response. This addition provides an opportunity for participants to list what they were thinking about when rating arguments, which may activate an active arguing strategy suggested by previous accounts of motivated reasoning (Edwards & Smith, 1996, Taber & Lodge, 2006, Taber et al., 2009).

The attitude change results from Study 1 seem to support a rational information-processing model better than the dissonant rationalizing account. On this motivated account, participants who are strongly committed to an issue would be predicted to polarize in response to incongruent evidence. We did not find this pattern of polarizing by participants with either strong prior opinions or strong affective commitment to an issue. Instead, we found that while prior attitude may have mattered differentially for some issues, participants showed a slight de-polarizing effect overall.

One limitation of the prior study was the use of a highly educated sample from an elite university. The rational behavior evidenced by these participants may not generalize to a larger sample. It may be that, as Taber and colleagues suggest, some studies are not able to elicit polarizing because participants are not engaged enough or not exposed to highly affective stimuli (Taber et al., 2009). It is possible that we did not find evidence of polarizing because participants were not actively judging the arguments by writing down their thoughts, as they did in Taber et al. (2009). This is addressed in the present study.

In Study 1, more highly rated arguments resulted in greater attitude change in the direction of the arguments. However, it is unknown whether this reflects the effects of the participants' bias or persuasion by what participants believe to be quality arguments. Argument rating was predicted by prior attitudes in line with the arguments, meaning this study does somewhat confound the variables of argument rating and prior attitude. To assess the effects of bias vs. argument quality, we would need some objective rating of argument quality. This issue is preliminarily explored in the present study and further addressed with a detailed model in Study 3.

Study 1 demonstrates that the attitude congruency bias, where people judge information compatible with their beliefs as stronger than incompatible information, is a robust phenomenon. The presence of the attitude congruency bias did not lead to polarization of participants' attitudes. Instead, the attitude change measurements revealed sensitivity to both the quantity and the quality of the argumentative texts that participants viewed. These data are in line with Bayesian accounts of information processing.

Previous work has found evidence of a disconfirmation bias, where participants spend more time reading arguments incompatible with their prior attitudes (Edwards & Smith, 1996). Participants also produce more denigrating arguments in response to incompatible arguments. This bias has been shown to model attitude polarization in a study involving highly affective political issues (Taber et al., 2009). The present study tests for this disconfirmation bias using four socio-political issues used in a previous study.

Accounts of motivated reasoning suggest that both biased processing of belief-compatible information and a disconfirmation bias lead to attitude polarization following mixed evidence. The present study provided a mixture of mixed and one-sided evidence, similar to Study 1, with the addition of active arguing in the form of prompts following the presentation of information to elicit participants' thoughts as they rated arguments congruent and incongruent with their prior attitudes.

If active arguing does activate biased processing that was absent in Study 1, we should find evidence that participants polarized in their attitudes, adjusting in line with their initial opinions. We tested for polarizing by comparing two models of attitude change, one that models attitude change based on the quality of the evidence and one that allows for polarizing. To test for the presence of disconfirmation bias, we modeled time spent reading arguments as a function of their congruency with participants' attitudes. We also examined the nature of participants' text to see whether more negative affect is present in response to incongruent arguments. Finally, we explored whether attitude change in response to highly-rated arguments is due to the objective quality of the arguments, or whether participants' bias may be driving attitude change.

Methods

Participants

Participants were 128 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 32 years old, with a mean age of 21. An additional

three 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 written responses were insufficient in length or content (n=2). Data from a single issue was removed for seven participants due to survey error (e.g., one or more responses failed to be recorded correctly).

Materials

Four socio-political issues were examined: abortion, animal testing, assisted suicide, and the death penalty. These issues were among the most popular topics found on two debate websites, www.procon.org and idebate.org. These issues were selected from those used in Study 1. We elected not to use the climate change issue because attitudes on this topic were homogeneous as virtually all of the participants in our pool agreed strongly with the statement that human activity contributes to climate change. The school uniforms issue was dropped because Study 1 suggested it engendered low levels of affective involvement. Otherwise, materials for attitude measurements, affective involvement measurements, and argument texts were identical to those used in Study 1.

Attitude measurements. For each issue, a single policy statement was chosen for participants to rate in terms of how much they agreed or disagreed (e.g., “Animal testing should be banned.”). Below we refer to the “Pro” side of an issue as the one that was in agreement with this policy statement, and the “Con” side as contrary to the policy statement. 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 were combined to form the initial attitude measurement. After the experimental treatment, participants again responded to five position statements per issue to form the subsequent attitude measurement.

Affective involvement 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. Responses to these four questions were combined to form a measure of affective involvement.

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

As in Study 1, there were three phases: initial collection of attitude and affective involvement measurements; the presentation and rating of arguments, along collection of written responses to arguments; and the subsequent collection of attitude and affective involvement measurements.

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

Following the collection of attitude and affective involvement measurements, each participant was asked to read and rate arguments for three randomly chosen

issues from the original set of four. 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 issue statement; one was randomly designated as the Con condition, such that the participant read and rated six arguments against the issue statement; and one was randomly designated as the Mix condition, such that the participant read and rated three Pro arguments and three Con arguments. The order of the issues was randomized, as was the order of the arguments presented within each issue.

After half of the arguments, participants were instructed to type into the text box what they were thinking about as they rated the previous argument. Participants could spend as much time on this free response task as they desired, but they were not allowed to advance to the next section until at least 60 seconds had elapsed. This was intended to encourage participants to respond during this interval.

After reading all arguments, participants were again asked to rate their positions on all four issues. Following the post-treatment attitude measurements, participants completed a brief political knowledge quiz to assess their political sophistication and answered two questions to assess their 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

Attitudes were scaled from -5 to 5, with -5 representing maximal disagreement with the issue statement, (e.g. "The death penalty should be banned,") and 5 representing maximal agreement with it. Items where participants spent too long

reading the argument text (more than 3 standard deviations from the mean, 332 seconds) were removed from analysis (21 items out of 2262).

Participants' prior opinions and affective involvement were analyzed to ensure uniform representation across experimental conditions, since within each issue, the 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 ($F = 22.1$ $p < .001$), there were no significant differences among conditions (Pro, Con, Mix, None), and no interaction of issue and 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 ANOVA in R. Models were fit with random intercepts for subjects and items (viz. arguments). The reported models are those that included statistically significant experimental predictors of argument rating and were not statistically different from more complex models (generally using cut-off $p < .01$, but trending predictors are also reported).

Models of attitude change were analyzed 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 approximately equivalent to selecting all predictors with a significant p value in the model ANOVA.

For all models, argument polarity (Pro/Con) was sum coded; experimental condition (Pro/Con/Mix) was treatment coded, with the Mix condition set as the

baseline. Issue (Abortion/Assisted suicide/Animal testing/ Death penalty) was also treatment coded. Abortion was used as a baseline for Issue because participants' average attitude change for this issue was close to zero.

Results

As in Study 1, 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 attitudes. Next, we assessed the factors that influenced attitude change in response to these arguments.

Argument Rating

To examine whether participants' argument ratings were biased by their initial attitudes, as in the previous study 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 attitude, affective involvement, issue, and political sophistication. Argument polarity was coded separately from treatment condition, and represents Pro and Con arguments irrespective of which condition (Pro, Con, or Mix) they were presented in. This is to separate the potential effects of experimental condition from those of argument polarity; that is, whether (for example) a given Pro argument would elicit different ratings when presented in the Pro condition, and accompanied by five other Pro arguments, versus the Mix condition, where it was accompanied by thee Con arguments and two other Pro arguments.

The best model predicts argument rating as a function of prior opinion, argument polarity, and issue (see Equation 2.1 and Table 2.1 for model results). Experimental condition did not influence the ratings of individual arguments (though there was a

marginal difference between a 4-way interaction including Condition and the 3-way interaction, p = .047). None of the other factors were significant, either as main effects or in interaction with other variables.

To further explore this interaction, the rating data were modeled by issue. Argument ratings for every condition showed a significant Prior attitude x Argument polarity interaction, with arguments in the attitude-congruent condition rated higher than arguments in the incongruent condition. Although the slopes of the lines differ by issue, the direction of the interaction is consistent across all conditions. Figure 2.1 illustrates the three-way interaction by showing the two-way Prior attitude x Argument polarity interaction for each issue.

Argument rating ~ Prior attitude x Argument polarity x Issue (2.1)

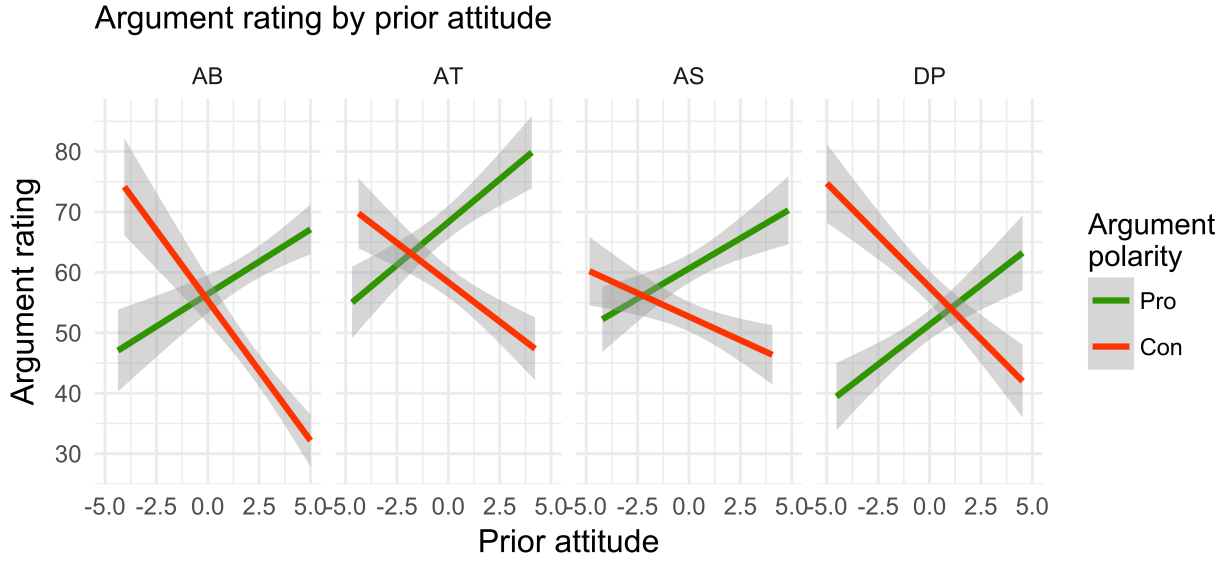

Figure 2.1. Average argument rating as a function of prior attitude (-5 most opposed, 5 most in favor of the position statement). Green lines represent Pro arguments presented in the Pro and Mix conditions. Red lines represent Con arguments presented in the Con and Mix conditions. Separate graphs are shown for the four different issues: Abortion, Animal testing, Assisted suicide, and Death penalty.

Table 2.1: Model results for Equation 2.1

Factor	Estimate	Std. Error	df	t-value	p-value
Intercept	56.963	3.356		16.97	< 0.001
Prior attitude	1.932	0.654	1	2.95	0.004
Argument polarity	-1.329	4.699	1	-0.28	0.778
Issue			3		
(Assisted suicide)	3.318	4.481		0.74	0.462
(Animal testing)	9.694	4.497		2.16	0.036
(Death penalty)	-4.831	4.515		-1.07	0.289
Prior attitude x Argument polarity	-6.600	0.934	1	-7.07	< 0.001
Prior attitude x Issue			3		
(Assisted suicide)	0.484	0.922		0.53	0.600
(Animal testing)	1.530	0.985		1.55	0.121
(Death penalty)	0.797	0.990		0.81	0.421
Argument polarity x Issue			3		
(Assisted suicide)	-6.517	6.341		-1.03	0.309
(Animal testing)	-7.013	6.415		-1.09	0.279
(Death penalty)	7.864	6.404		1.23	0.225
Prior attitude x Argument polarity x Issue			3		
(Assisted suicide)	2.424	1.268		1.91	0.056
(Animal testing)	1.048	1.389		0.75	0.451
(Death penalty)	0.368	1.362		0.27	0.787

Argument rating data in Study 2 thus replicate the attitude congruency bias seen in Study 1, in keeping with prior reports by previous investigators (Edwards & Smith, 1996; Lord, Ross, & Lepper, 1975; Taber et al., 2009). Below we examine whether these biased ratings were associated with polarized attitude change.

Participant reading time and sentiment

To test for the presence of a disconfirmation bias, we modeled participants' reading time for each argument as a function of the argument's congruency with prior attitudes. A linear mixed effects regression model with reading time as a function of prior attitude and argument polarity failed to reach significance, as did a model of log reading times. Neither model was improved by the addition of affective involvement and/or political sophistication. Participants in the present study did not spend more time

reading incongruent arguments, as found in prior research (Edwards & Smith, 1996, Taber et al., 2009).

Participants' text responses were coded for sentiment on a positive-negative scale using the Stanford CoreNLP server (Manning et al., 2014) with a Python package. Each text response was processed as a single item and assigned a numeric score. Positive values represent positive sentiment, and negative values represent negative sentiment. Numeric sentiment scores ranged from -1.43 to 1.54 and were skewed negative.

We first assessed whether participants generated more negative response to attitude-incongruent arguments than to congruent arguments. To test this, we modeled sentiment as a function of prior attitude and argument polarity. A linear mixed effects regression model with sentiment as a function of Prior attitude x Argument polarity failed to reach significance. The addition of affective involvement and/or political sophistication did not cause this interaction to reach significance.

We next modeled sentiment as a function of argument rating because, although rating is highly correlated with arguments' congruency with prior attitudes (Eq. 2.1), participants' responses to arguments may have more to do with their objective quality than their alignment with participants' attitudes. A linear mixed effects regression model of sentiment as a function of argument rating reached significance compared to an intercept-only model ($X^2[1,1120] = 50.7$, $p < .001$). This model was not improved by the addition of affective involvement and/or political sophistication. Participants' sentiments were thus related to the ratings they provided assessing each argument's strength, and

not to their congruency with prior attitudes (although these variables are highly correlated).

Attitude change

As in Study 1, we next assessed whether and how the attitude congruency bias shown above leads to attitude change. We created linear models to predict attitude change from experimental treatment and participants' argument ratings. Attitude change was calculated as participants' post-survey attitude minus their pre-survey (prior) attitude. Positive values represented change in the Pro direction (more supportive of the issue statement), and negative values represented change in the Con direction (more opposed to the issue statement).

An overall regression toward the mean was present in the attitude measurements in all conditions, even the control condition where no arguments were viewed. On average, participants' attitude change from their Prior to Post-treatment attitude was toward the center of the attitude scale. Consequently, we used participants' responses in the control condition to calculate a correction factor. Accordingly, initial attitude measurements in each issue were used to predict attitude scores in that issue at the end of the study. This average slope, calculated for the control condition in which participants did not read any arguments, was subtracted from each participants' attitude change score. The calculation was performed separately for each issue because the correction factor differed as a function of issue. This effectively corrects for regression toward the mean in the experimental conditions. All analyses of attitude change utilize these corrected values.

To test whether participants' attitude change is best modeled by the amount and quality of the evidence, we began with the model of attitude change from Study 1.

Attitude change ~ Argument rating x Argument polarity + Condition x Position + Position x Issue (2.2)

We compared the model in (2.2) with a model including a 3-way interaction of argument rating, argument polarity, and position (2.3).

Attitude change ~ Argument rating x Argument polarity x Position + Condition x Position + Position x Issue (2.3)

Model ANOVA suggested the addition of the Position factor did not improve the model ($F[12,2241] = 1.13$, $p = .33$). Additionally, the 3-way interaction in Equation 2.3 did not reach significance ($F[4,2241] = 0.87$, $p = .48$). An ANOVA performed on the 2-way interaction model (Equation 2.2) revealed that the Position x Issue interaction was trending, but did not reach significance ($p = .022$). To simplify the reporting of results, estimates in Table 2.2 are shown for Equation 2.4, the best model predicting attitude change for the present study.

Attitude change ~ Argument rating x Argument polarity + Condition x Position + Issue (2.4)

Table 2.2: Model results for Equation 2.4

Factor	Estimate	Std. Error	df	t-value	p-value
Intercept	-0.060	0.253		-0.24	0.812
Argument Polarity	-0.936	0.214	1	-4.38	< 0.000
Argument rating	0.000	0.003	1	-0.04	0.966
Argument rating:Polarity	0.017	0.004	1	4.66	< 0.001
Condition			2		
(Pro)	0.591	0.236		2.51	0.013
(Con)	-0.369	0.234		-1.58	0.115
Position			4		
(-2)	1.410	0.258		5.46	< 0.001
(-1)	0.985	0.235		4.18	< 0.001
(1)	-0.279	0.201		-1.39	0.165
(2)	-0.210	0.208		-1.01	0.313
Issue			3		
(Assisted suicide)	0.120	0.143		0.84	0.403
(Animal testing)	-0.070	0.156		-0.45	0.654
(Death penalty)	-0.307	0.151		-2.03	0.043
Condition:Position			8		
(Pro : -2)	-1.170	0.393		-2.98	0.003
(Pro : -1)	-0.620	0.403		-1.54	0.124
(Pro : 1)	0.087	0.360		0.24	0.809
(Pro : 2)	-0.541	0.348		-1.55	0.121
(Con: -2)	-1.231	0.418		-2.94	0.003
(Con: -1)	0.352	0.434		0.81	0.418
(Con: 1)	-0.254	0.342		-0.74	0.458
(Con: 2)	0.000	0.327		0.00	0.999

The Argument rating x Argument polarity interaction is shown in Figure 2.2. As in Study 1, attitude change following highly rated arguments is on average in the direction of the arguments.

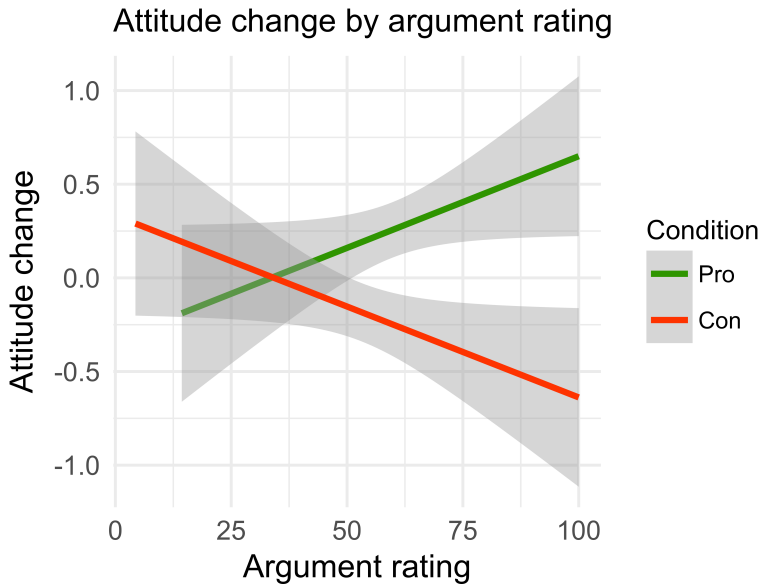

Figure 2.2: Interaction of argument rating and argument polarity. The red line represents average attitude change following Con arguments in the Con or Mix condition. The green line represents average attitude change following Pro arguments.

As in Study 1, Study 2 revealed a Condition x Position interaction that reflects de-polarizing. Figure 2.3 shows that opponents in the Pro condition (to the left of zero on the x-axis) displayed more positive attitude change than supporters, and supporters in the Con condition (to the right of zero on the x-axis) show more negative attitude change than opponents. In fact, participants in the Mix condition showed slightly more de-polarizing behavior than participants in the Pro and Con conditions. This result is contrary to the hot cognition account that suggests the presentation of a mixture of evidence triggers motivated reasoning processes that result in polarized attitude change. These data suggest that although attitude change effects were small, attitude incongruent arguments were actually more influential than congruent ones.

Study 2 also replicated the interaction between Argument rating and Argument polarity observed in Study 1. That is, irrespective of participants' positions on the issues, attitude change was associated with viewing highly rated arguments. These findings are

consistent with the Bayesian framework outlined in the previous chapter in showing that the direction of attitude change depended on the evidence, with greater attitude change after encountering evidence that was deemed to be of high quality (see Figure 2.2). However, interpretation of these data is complicated by the presence of the attitude congruency bias in the argument ratings. That is, participants systematically inflated the ratings for attitude congruent arguments so that observed data might reflect the influence of these biases rather than increased sensitivity to arguments that were objectively better.

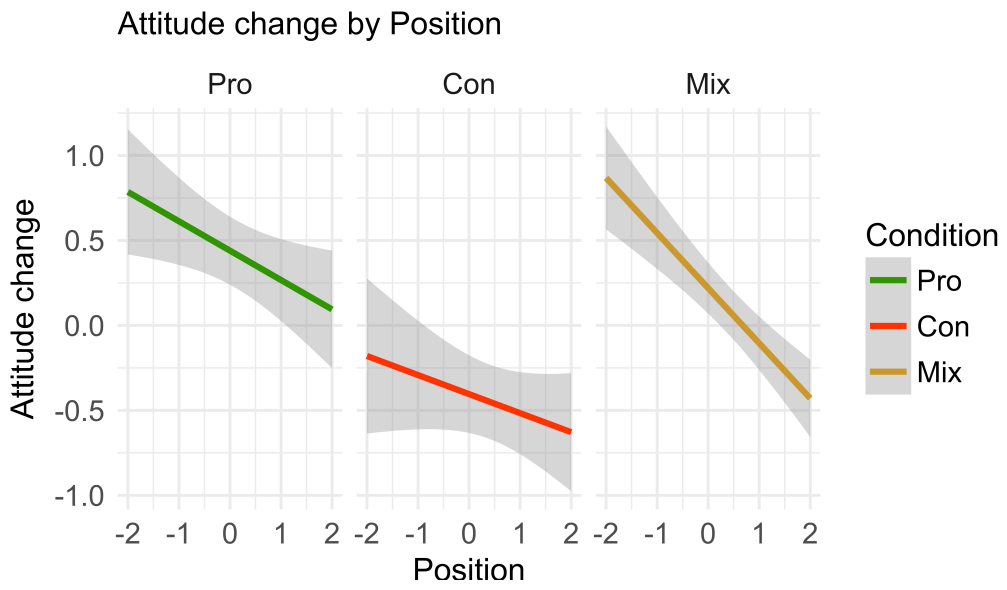

Figure 2.3: Attitude change as a function of Position. Position represents participants' initial attitudes, from -2 (most opposed) to 2 (most in favor of an issue statement). Position is modeled as a categorical variable, but is plotted as a numerical range to show trends. Lines represent the Pro (green), Con (red), and Mix (yellow) conditions.

To better test whether high quality attitude incongruent arguments led to attitude change in the direction of the arguments, we used argument ratings from a previous study (Study 1) to model attitude change. This separates the effect of the quality of arguments themselves from the effect of participants' biased ratings. For this

exploratory analysis, argument ratings from all participants in Study 1 were averaged for each argument. These averaged values were used in place of subjects' ratings in the present study to represent average argument quality as judged by a separate set of participants with similar characteristics (participants are drawn from the same subject pool of UCSD students).

The above model of attitude change (Equation 2.4) was modified to include Objective rating as a predictor instead of Argument rating. The modified model is shown in Equation 2.5 below.

Attitude change ~ Objective rating x Argument polarity + Condition x Position + Issue (2.5)

AIC values were used to compare this model to the model using participants' own ratings. The models were shown to be significantly different (AIC = 1549 vs. 1566), with the model using participants' own ratings outperforming the model with Objective rating. Model ANOVA on the objective rating model showed that this variable achieved a conventional level of significance in predicting attitude change ($p = .019$ for the Objective rating x Argument polarity interaction).

These results show that at least some of the attitude change modeled by argument ratings is due to the quality of arguments themselves, although some of this change may be due to the participants' own bias that was present in the argument ratings. The relative contribution of bias versus objective quality is addressed more directly in the study described in Chapter 3.

Overall, Study 2 replicates the findings of Study 1, even though the present study required participants to justify their argument ratings with a textual response. This suggests our earlier failure to find the phenomenon of attitude polarization reported by Taber and colleagues (2009) did not result because of this methodological detail. While participants still exhibited a robust attitude congruency bias, favoring arguments in line with their prior attitudes, they did not appear to engage in argumentative strategies while reading and responding to argument text. Rather than spending more time reading incongruent texts and producing more negative responses, participants' reading time did not vary with argument congruency, and the affective content of their responses was instead related to ratings of the arguments.

Keeping in mind that participants' ratings of arguments are confounded with their prior attitudes, these results still suggest that individuals are capable of processing arguments with sensitivity to their quality despite maintaining a preference for those arguments congruent with individual attitudes. As in Study 1, the quality and the quantity of information modeled belief change, which suggests that an information-processing account of attitude change is appropriate even in the presence of controversial issues.

Chapter 3
The role of affective involvement and knowledge in processing mixed evidence for social issues

The Role of Affective Involvement and Knowledge in Processing Mixed Evidence for Social Issues

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

Exposure to mixed evidence can lead to polarization, or adopting a more extreme version of one's initial attitude. One potential reason for this is attitude congruency bias, rating evidence that supports one's attitude as stronger than evidence that undermines it. Here we explore factors associated with this bias and their relationship to attitude change following exposure to mixed evidence. We conducted several tests, including an attitude survey on two controversial social issues, a poll regarding participants' affective involvement in each issue, an argument rating task, and assessments of knowledge about social issues and political sophistication. We replicated the attitude congruency bias. Ratings bias was associated with affective involvement, but not with measures of topic knowledge or political sophistication. Attitude change was predicted by a linear combination of objective argument strength and rating bias. Participants' sensitivity to objective argument strength suggests the attitude congruency bias does not inevitably lead to polarization.

Keywords: decision making; reasoning; motivated reasoning; rationality; language and thought; attitude congruency bias

Introduction

Over time, people form attitudes about objects, people, and issues. These attitudes may change through any number of affective, cognitive, and/or behavioral mechanisms (see Petty & Wegener, 1998 for an account of the Elaboration Likelihood Model of persuasion). New information can lead to attitude change, but often that information is judged in light of extant attitudes (prior opinions).

Judging information differentially in consideration of its agreement with prior opinions may be natural and beneficial in environments where one's beliefs are true most of the time (Alloy & Tabachnik, 1984). One study showed that scientists judged research as being of higher quality when its conclusions were in agreement with their own prior opinions (Koehler, 1993). Researchers rejected the quality of studies based only on their outcomes, not the merits of the design. This can be viewed as rational when findings are inconsistent with a body of scientific knowledge, but outside of established fields could clearly present difficulties in reaching consensus.

Fortunately, polarization, where exposure to the same evidence leads people to opposite attitude adjustments in the direction of their prior attitudes, is relatively rare (Kahan, Peters, Dawson, & Slovic, 2017), although repeated discussion within a group with a strong sense of shared identity can lead individuals to hold more extreme attitudes (Sunstein, 2002). The domain of social and political issues may be one such special case, since in many areas there can arise "two sides" of a seemingly factual issue, with supporters on each side failing to change their beliefs toward an objective consensus (Kahan, 2016). For example, in a survey of American voters, 75% of self-identified liberals believed that climate change was due primarily to human activity, whereas only 45% of conservatives shared this belief (McCright & Dunlap, 2011).

One explanation for polarization on social issues is motivated reasoning (Ditto & Lopez, 1992; Kunda, 1990). On such accounts, attitude polarization occurs because people with opposing views draw opposite conclusions from the very same evidence. In a classic study, Lord, Ross, and Lepper (1979) queried participants about their views on capital punishment, and then presented them with the results of two studies, one that suggested the death penalty deters crime, and one that suggested the opposite conclusion. Participants were asked to rate the quality of each study, and then to re-characterize their views on the death penalty. Interestingly, participants tended to rate the study that supported their own beliefs as being objectively better than the one that undermined them, and each group adjusted their beliefs to more strongly favor their original position (Lord, Ross, & Lepper, 1979).

Subsequent studies have shown that exposure to mixed evidence can lead people to polarize, changing their beliefs to be more in line with their initial attitudes (Edwards & Smith, 1996; Taber & Lodge, 2006). One explanation for this change in belief is that people accept information congruent with their extant opinions without critical examination, while incongruent information is critically examined and judged more negatively in the presence of negative affect (Taber, Cann, & Cucsova, 2009).

If the process by which the attitude congruence bias leads to polarization does involve emotional processing of evidence, then people with emotional commitment to their attitudes may be more likely to display the bias and also to polarize. If, however, knowledge about a topic influences the way evidence is processed, we may see topic knowledge influencing belief change.

We first examine which variables are related to attitude congruency bias, specifically whether affective involvement leads to increased bias or whether cognitive factors of knowledge about the topic or political sophistication can predict argument ratings. Next, we examine the role of attitude congruency bias in attitude change and explore the contribution other factors. Findings are interpreted in light of their consistency with accounts of motivated reasoning.

Methods

Participants

Participants were undergraduate students enrolled in Psychology, Linguistics, or Cognitive science courses at the University of California, San Diego (UCSD) (n=141, 99 female) participating as part of a course requirement. Participants ranged from 18 to 29 years of age (mean = 20). All participants provided informed consent, and procedures were approved by the Institutional Review Board (IRB) at UCSD.

Procedure

After consenting to participate in the study, participants first completed Initial Attitude and Affective involvement measurements for each issue (described in the Materials section). One of the issues was randomly assigned to the Mix condition, meaning that the participant would read arguments for and against the position articulated in the issue statement (henceforth: the issue). The other issue was assigned to the Control condition. Participants were not exposed to any arguments regarding the issue in the Control condition.

During the treatment phase of the study, each participant read 3 Pro and 3 Con arguments regarding the issue in the Mix condition. Arguments were presented in a random order. Following the presentation of each argument, participants used a 100-point slider to indicate the argument's strength on a scale from Weak to Strong (numbers not visible). After half of the arguments, participants were asked to describe their thoughts about the argument via a typed response in a text box.

After the treatment phase, participants again completed the attitude measurement survey for both issues to determine their Post-treatment Attitude scores. Next, participants completed a Topic Knowledge test for each issue and a brief political knowledge quiz to assess their political sophistication. Finally, they viewed a debriefing page that explained the goal of the study and provided links to the websites used for the argument texts.

Materials

The survey used for the present study contains a subset of materials used in a previous study (Bardolph & Coulson, 2017). Two socio-political issues were included: animal testing and the death penalty, selected from the most popular topics on two debate websites, www.procon.org and idebate.org. Text from both sides of debate arguments was used to create one-paragraph arguments that either supported or opposed the related issue.

Attitude measurement For both issues, participant attitude was measured using 5 survey questions: A single policy statement ("Animal testing should be banned"; "The death penalty should be illegal") with a rating slider from Disagree to Agree (0 to 100, numbers not visible), followed by four position statements for each issue. These position statements were selected from "Points for" and "Points against" on the idebate.org archive (e.g., "Animals involved in animal research are mostly well treated."). Each position statement was rated using a 9-point scale of agreement/ disagreement. Ratings from the policy statement and the four position statements were scaled and combined to form an average initial Attitude, ranging from -5 (most opposed to the issue) to 5 (most in favor of the issue).

After the experimental treatment, participants responded again to the same five statements for each issue. Responses were combined as before to form an average post-treatment attitude score.

Affective involvement For each issue, affective involvement was measured using 4 survey questions with a 9-point rating scale indicating: how much participants care about the issue, how strong their feelings are, how certain they are of their feelings, and how much they have thought about the issue. These four measurements were combined to form a measure of affective involvement.

Arguments Six supporting (Pro) and six opposing (Con) arguments were selected using text from the debate sites 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 edited. A study of ratings for these arguments that drew from the same participant pool indicated slightly higher ratings of arguments regarding animal testing than the death penalty, but revealed similar ratings for participants supporting and opposed to the position statement for each issue (Supporting: mean = 61.5 and Opposed: mean = 62.6 for animal testing arguments, and Supporting: mean = 54.4 and Opposed: mean = 55.8 for death penalty arguments).

Topic knowledge For each issue, topic knowledge was measured using eight multiple choice factual questions (e.g., "Which animal is used most frequently for research?"; "The death penalty was ruled to be constitutional in the US under which amendment?"). These questions were piloted in an earlier norming study. Items that were too easy or too difficult were not included in the present study. Topic knowledge for each issue is represented by the percentage of questions the participant answered correctly.

Analysis

Initial attitude and Post treatment attitude were scaled from -5 to 5, representing, respectively, the opinion most opposed to each issue, and most in favor of the issue. In this coding, a positive score represents an attitude in favor of the legality of animal testing and the legality of the death penalty. A measure of attitude change was created by subtracting each participant's Prior attitude from their Post treatment attitude. Consequently, Attitude change could range (in principle) from -10 to 10. Affective involvement ratings for each participant for each issue ranged from 1 to 9 (least to most strong). Items for which participants spent less than 3 standard deviations below the median log reading time or more than 3 standard deviations above the median log reading time were removed (9 items out of 846).

Argument ratings A linear mixed effects regression (LMER) model was used to analyze argument rating data. Models were constructed with the lme4 package in R (Bates, Maechler, Bolker, Walker, et al., 2014 R Core Team, 2015). Analysis involved construction of an LMER model to predict argument ratings and the use of backward model comparison using ANOVA. Models were fit with random intercepts for participants and for arguments (items). The use of random intercepts helps control for individual variability in participants' use of the rating scale, as well as for differences in the quality of particular arguments (some arguments are intrinsically better and consequently tend to be rated as stronger by all participants). Backward model comparison yielded the most parsimonious model that included all significant predictors ($p < .01$ used as cutoff).

Linear models Predictive relationships among experimental variables, including Attitude change, were analyzed with a linear model in R. Analysis involved backward model comparison using ANOVA to establish the optimal model. This is roughly equivalent to selecting all predictors below a threshold p value in the model ANOVA.

Objective argument rating To obtain an approximately objective rating of argument strength, we used ratings from two prior experiments where participants viewed the same arguments in a mixed condition (Pro and Con arguments presented together). N=39 for animal testing, N=48 for death penalty, with participants' opinion approximately evenly distributed for each issue (prior opinion ranged from -4.1 to 3.9, mean = -0.2 for animal testing; -4.7 to 5.0, mean = 0.1 for death penalty).

Ratings from the two prior experiments and the current study did not differ significantly by study. Argument label was used to predict an average rating of argument strength for each argument. These predicted values from the two prior studies were used as a measure of argument quality. In the model, this factor is referred to as Argument.

Rating Bias To create a single measure of bias for each participant, each argument's objective rating was subtracted from the participant's rating. For example, consider a participant who gave a Pro argument a rating of 90 (very strong). If this argument's objective rating were 65, the residual rating would be 25, indicating that this participant is biased toward the Pro position.

An average rating bias was created by subtracting each participant's bias in favor of Con arguments from bias in favor of Pro arguments. Using this scale, bias reflects how much stronger participants rated Pro arguments than Con arguments for a given issue. For example, if a participant were on average biased by 25 points in favor of Pro arguments and 5 points against Con arguments (average bias = -5 for Con arguments), their Bias score would be (25 - -5) = 30, indicating that they rate Pro arguments more highly than Con arguments. A participant who rated Con arguments more highly than Pro arguments would receive a negative Bias score. Bias scores were transformed to a z-score variable for modeling. A measure of Folded bias, the absolute value of the bias term, was also used when correlating the magnitude of bias with other variables.

Correcting Post-survey attitude regression to the mean An overall regression toward the mean was present in both the Experimental and Control conditions. On average, participants' change from their Prior to Post-survey attitude was toward the center of the attitude scale. This means that participants in favor of an issue changed their opinion to be more opposed to that issue and vice versa, even when they do not read any arguments. For example, one participant in favor of animal testing reported a prior attitude of 3.59 and a post-survey attitude of 2.97 for animal testing. Although this participant was not exposed to any arguments about animal testing, their attitude moved toward the center of the attitude scale. We refer to such "changes" in attitude as regression to the mean.

To correct for regression to the mean, we used attitude scores in the Control condition to determine a correction factor that could be applied to the Treatment condition. Beginning with data from the Control condition in which participants read no arguments, we used linear regression to predict post-survey attitude change from their initial attitude reports. This average slope (-0.17) was subtracted from attitude change measures in both the Control and the Treatment conditions, effectively correcting for regression to the mean. For example, the attitude change score for the participant described above was corrected from -0.62 to 0.06.

Results

Histograms of participant variables used as predictors are shown in Figure 1. The frequency of values of the predictor is shown for Affective involvement, Topic knowledge, Political sophistication, and Rating Bias (z-scored). Table 1 shows a correlation matrix for these four variables. The only correlation trending toward significance is the correlation of Affective involvement and Topic knowledge ($p = .07$).

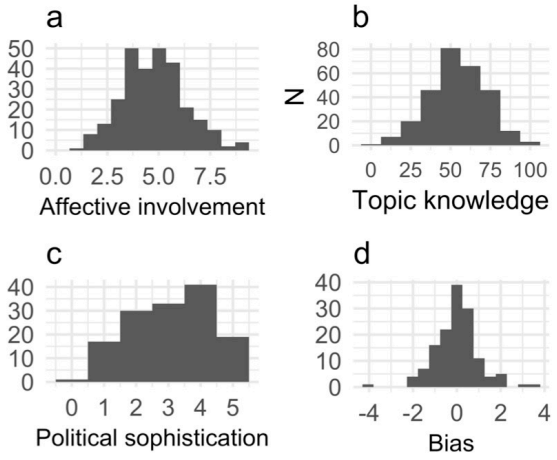

Figure 1: Histograms showing the frequency of (a) Affective involvement: two measures per participant, one for each issue; (b) Topic knowledge: two measures per participant, one for each issue; (c) Political sophistication: one measure per participant; and (d) Bias: one measure per participant, calculated for the Treatment condition and transformed to a z-score variable.

Argument rating

This analysis tests for the attitude congruency effect in participants' ratings of arguments regarding animal testing and the death penalty by exploring how each participant's rating for a given argument varies as a function of their prior opinion on the issue (opposed to supportive), and argument polarity, that is, whether the argument itself was Pro or Con. On such an analysis, attitude congruency bias is revealed by a cross-over interaction of these factors, as participants rank arguments congruent with their prior opinions as stronger than arguments that are incongruent with those opinions. Further, to see if attitude congruency was related to participants' affective involvement with the issue, their political sophistication, or their degree of extant knowledge regarding the topic, we included these factors as additional predictors in the model.

Participants' argument ratings were analyzed with an LMER model (as described in the Analysis section). The initial model included predictors of Prior opinion, Affective involvement, Argument polarity (Pro/Con), Issue (Animal testing, Death penalty), Political sophistication, and Topic knowledge, with random intercepts for participants and for arguments. More complex models were compared to models with fewer predictors using model ANOVA, yielding the most parsimonious model that still contained all significant predictors of Argument rating (using $p < .01$ cutoff).

Argument rating ~ Prior opinion * Argument polarity (1)

Table 1: Correlations of predictor variables.

	Affective Involvement	Topic Knowledge	Political Sophistication
Topic knowledge	0.11		
Political sophistication	-0.06	0.03	
Bias	-0.09	-0.12	0.06

The model that best predicts argument rating is shown in Equation 1.[1] Results of the model are listed in Table 2. The relationship between Prior opinion and Argument polarity is shown in Figure 2. The predicted cross-over interaction reflects the fact that participants who were extreme supporters (5 on the Prior opinion axis) rated attitude-congruent Pro arguments as stronger than incongruent Con arguments. Similarly, extreme opponents (-5 on the Prior opinion axis) rated the attitude congruent Con arguments as stronger than incongruent Pro arguments. However, while the data suggest argument ratings were indeed subject to attitude congruency bias, we failed to detect a relationship between argument ratings and any of our other measures, including affective involvement, topic knowledge, or political sophistication. Affective involvement was correlated with Prior opinion ($R = 0.53$, $p < .001$), but did not have an additional effect on Argument rating.

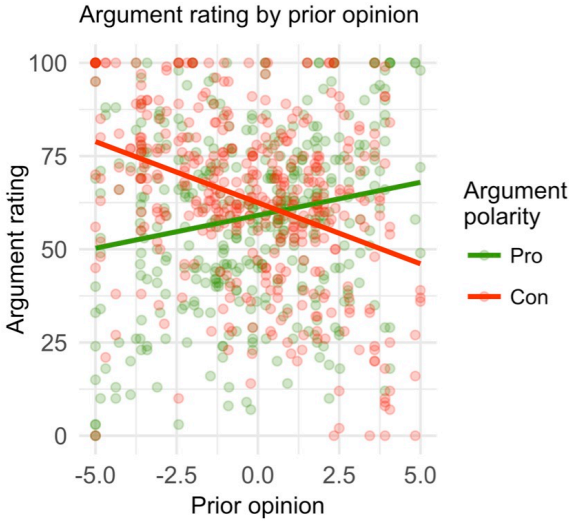

Figure 2: Interaction of argument polarity and prior opinion (-5 most opposed, 5 most in favor of the issue). Circles represent individual argument ratings. Green and red lines represent average rating of Pro and Con arguments respectively.

[1] This same analysis can be performed using Position and Folded prior opinion (magnitude). The results are the same as Equation 1, and the two models of Argument rating are not statistically different.

Table 2: Model results for Equation 1.

Factor	df	F value
Prior opinion	1	4.7
Argument polarity	1	1.3
Prior opinion x Argument polarity	1	86.9

Attitude change

Apart from the cause of attitude congruency bias, another question of interest concerns its effects on attitude change. Were participants who displayed attitude congruency bias in their argument ratings more or less likely to change their opinions? On some accounts, biased assimilation of the evidence can have an undue effect on belief change and lead to attitude polarization. To explore the impact of biased assimilation on belief updating, we used linear models to test whether there was a relationship between attitude change and rating bias.

If biased ratings lead to polarizing, we expect a positive relationship bias and attitude change: that is, positive Bias, rating Pro arguments more highly, will lead to positive Attitude change (more supportive of the issue). A negative Bias, rating Con arguments more highly, will lead to negative Attitude change (more opposed to the issue).

Participants' argument ratings, however, do not reflect bias alone. Each argument may have a degree of strength relative to other arguments, or an objective quality. An individual's rating, therefore, may reflect the objective argument strength and individual bias. For this reason, argument ratings were split into a measure of objective argument quality and individual bias.

Individual bias (labeled Bias) was calculated as described in the Methods section and used to predict corrected Attitude change in a linear model. This model tests whether biased assimilation leads to polarizing. Objective argument rating was included as an additional, separate predictor (Argument). Further, to assess the impact of affective involvement, topic knowledge, and political sophistication on Attitude change, we included these factors as additional predictors.

Analysis involved construction of a linear model to predict Attitude change and the use of backward model comparison to establish the optimal model. The initial linear model included factors of Bias, Argument, Affective involvement, Issue (Animal testing, Death penalty), Political sophistication, and Topic knowledge. Nested linear models were compared using model ANOVA in R as described in the Methods section.

The model that bests predicts Attitude change is shown in Equation 3. There was a significant main effect of Argument ($p < .005$), and a main effect of Bias ($p = .013$). The coefficient for both predictors was positive: Argument ratings in favor of Pro arguments predict opinion change in support of the issue, Bias in favor of Pro arguments predicts opinion change in support of the issue, and vice versa for Argument/Bias in favor of Con arguments. No significant effects were found for other measures, including affective involvement, topic knowledge, or political sophistication.

$$\text{Attitude change} \sim \text{Argument} + \text{Bias} \quad (3)$$

Table 3: Model results for Equation 3.

Factor	Estimate	df	F value	P value
Argument	0.27	1	10.9	< .005
Bias	0.20	1	6.3	.013

Discussion

Here we explored the importance of topic knowledge and political sophistication on the one hand and affective involvement on the other to different phenomena related to reasoning about controversial social issues.

Our initial analyses explored the role of these factors in how participants evaluate arguments that are congruent vs. incongruent with their prior attitudes on the issue. We found attitude congruency bias, but no evidence for contribution of knowledge, political sophistication, or affective involvement as moderators of this phenomenon. Replication of the attitude congruency bias is consistent with previous findings (Lord, Ross, & Lepper, 1979; Edwards & Smith, 1996; Taber, Cann, & Cucsova, 2009; Bardolph & Coulson, 2017) and consistent with accounts of motivated reasoning. The present study does not indicate that this bias is related to how knowledgeable individuals are about the topic under discussion. Because prior attitudes are highly correlated with affective involvement, the degree to which individuals care about an issue may contribute to their bias, although a precise relationship cannot be established by these data.

Further, we explored the role of topic knowledge, political sophistication, and affective involvement in ratings bias. Although we found no relationship between either topic knowledge or political sophistication in participants' degree of ratings bias, we did find a positive association between bias and affective involvement. The more affectively involved participants were with a given issue, the more biased their argument ratings were. These data are in keeping with motivated reasoning accounts.

Finally, we explored the relative importance of ratings bias, objective argument quality, affective involvement, topic knowledge, and political sophistication for attitude change. Of these factors, only objective argument quality and ratings bias were significant predictors of attitude change. While the relationship between ratings bias and attitude change is in line with motivated reasoning, our models suggest objective argument quality is a slightly better predictor of attitude change. The latter finding indicates participants were sensitive to the quality of the evidence, changing their opinions more when they were exposed to strong arguments than when they were exposed to weak ones.

These data argue against prior studies that suggest the attitude congruency bias demonstrated in the present study is likely to lead to the polarization of opinions (Taber, Cann, & Cucsova, 2009). Attitude change in the present study was in fact more influenced by the objective quality of the arguments than the participants' ratings bias. The present study suggests that while people are more skeptical of evidence that contradicts their existing attitudes, they are also sensitive to the quality of that evidence.

One limitation of the present study was the use of a highly-educated sample from a leading public university in the United States. The behavior of these student participants may not generalize to a larger sample. It is also possible that measurements of participants' attitudes do not reflect a single, stable opinion, but a combination of response instability and multiple response effects (see Zaller, 1992 for a model of survey response). Our method of correcting for regression to the mean addresses some of this variability, but there may be aspects of participants' opinions that are not fully captured by the survey methods.

Overall, these data are consistent with accounts of motivated reasoning, replicating the phenomenon of attitude congruency bias and revealing a relationship between bias and affective involvement in a controversial social issue. However, they also reveal participants' rational sensitivity to the quality of the evidence with which they are presented. This sensitivity to argument quality could potentially mitigate attitude polarization. It also highlights the possible impact of exposure to media of varying quality: although consumers of media may indeed be biased by their own attitudes, persuasive arguments from quality sources may have an impact on attitude change.

Acknowledgments

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

References

Alloy, L. B., & Tabachnik, N. (1984). Assessment of co-variation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological review, 91*.

Bardolph, M. D. & Coulson, S. (2017). Belief updating and argument evaluation. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 1586-1591). London, UK: Cognitive Science Society.

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 non-preferred conclusions. *Journal of Personality and Social Psychology, 63*(4).

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

Kahan, D. M. (2016). The politically motivated reasoning paradigm, part 1: What politically motivated reasoning is and how to measure it. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*.

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy, 1*(1).

Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational behavior and human decision processes, 56*(1).

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).

McCright, A. M., & Dunlap, R. E. (2011). Cool dudes: The denial of climate change among conservative white males in the United States. *Global environmental change, 21*(4).

Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables (In D. Gilbert, S. Fiske, & G. Lindzey (Eds.). *The handbook of social psychology (Vol. 1)*.

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

Sunstein, C. R. (2002). The law of group polarization. *Journal of political philosophy, 10(2)*.

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

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*(3).

Zaller, J., & Feldman, S. (1992). A simple theory of the survey response: Answering questions versus revealing preferences. *American journal of political science*.

Chapter 3, in full, is a reprint of the material as it appears in Proceedings of the 39th Annual Conference of the Cognitive Science Society. Bardolph, Megan; Coulson, Seana. The dissertation author was the primary investigator and author of this paper.

Chapter 4
The effects of information choice and rating bias on attitude change

The preceding chapters examined how individuals process varying amounts of evidence that either supports or contradicts their prior attitudes. In these studies, participants were presented with information in the context of an online survey where their compliance with instructions appeared to be high: participants' text in response to reading arguments nearly always indicated they had read and thought about the information presented. This context did not allow participants the option of ignoring information or selecting which arguments to read. The present study addressed attitude change in the context of participant choice, allowing some participants to select information to read from a list of available arguments; it also directly compared the choice vs. no choice contexts by employing a yoked design in which a second group of participants viewed the same information as those in the choice condition, allowing for the effects of choice to be separated from the effects of simply viewing and rating arguments.

After the introduction of the cognitive dissonance framework (Festinger, 1957), many studies investigated how individuals selectively choose to read or otherwise expose themselves to new information in light of their personal decisions, attitudes, or behaviors. A meta-analysis of studies where participants reported prior attitudes and then selected information to view found that the phenomenon of congeniality bias was, although small, robust (Hart et al., 2009). Congeniality bias is defined as selection of a greater proportion of attitude congruent information. Hart and colleagues (2009) additionally found that this congeniality bias was larger when information available for

selection was higher in quality and when individuals were more committed to their initial attitudes or beliefs.

In a prior study of information choice, Taber and Lodge (2006) allowed participants to choose text arguments to read from political interest groups regarding either affirmative action or gun control. Participants first viewed eight total arguments for one issue and were allowed to choose how many of those arguments to read from each labeled source. In the second part of the experiment, participants viewed four Pro and four Con arguments for the second issue. For both issues, participants rated each argument's strength using a slider bar. The authors found a rating bias, where participants rated arguments compatible with their prior attitudes as stronger than incompatible arguments. They also found that participants chose to read more compatible arguments when given a choice, especially those with higher political sophistication. Participants polarized in both tasks (choice and balanced information). In both tasks, polarization was driven by participants with extreme prior attitudes and those with high political sophistication, with participants low on these measures failing to polarize. Polarization was further driven by bias in favor of pro-attitudinal (congruent) arguments in the information board task and biased argument rating in the balanced information task, with the least biased participants in both tasks failing to show polarization.

Because studies in the preceding chapters showed evidence of the attitude congruency bias without attitude polarization, we wanted to test whether information choice could produce attitude polarization. We believe that the quantity of information viewed by participants is an important factor in attitude change, with more evidence in

favor of a position predicting attitude change toward that position. If participants are able to skew the proportion of arguments that they read in favor of a pre-existing attitude, they may polarize in response to biased information. However, we do not know if choice itself is important for driving this polarization; some studies indicate that polarization is a consequence of mixed evidence combined with extreme attitudes, biased processing of arguments, and high political sophistication (Taber and Lodge, 2006, Taber et al., 2009).

Based on the previous results (Study 1, Study 2), we expect that participants may show evidence of polarization if they view a greater amount of attitude-compatible information. This pattern should not necessarily depend on choice. If this is the case, then we predict that the proportion of pro-attitudinal arguments chosen will correlate with attitude polarization independent of whether participants chose that proportion in a biased manner. In a model of attitude change, proportion of Pro vs. Con arguments viewed should predict change in the direction of the arguments, with more change toward the Pro side of an issue following higher proportion of Pro arguments chosen and vice versa. This pattern would be found for participants who chose which information to view and for participants who were instead simply presented with the same information.

If the proportion of pro-attitudinal arguments chosen is modeled by the extremeness of prior attitudes, then a motivated account of polarization is indeed supported. We are interested in whether there is a special role for this congeniality bias; that is, do participants who choose to view more information compatible with their prior attitudes polarize due to the act of choosing? If so, we would expect to see attitude

change in line with prior attitudes only for those participants who exhibit the congeniality bias in the choice condition. This would provide further support to accounts of motivated reasoning.

The studies in prior chapters were concerned not just with the quantity of evidence, but also its quality. The quality of the arguments in this study was represented by ratings from a set of participants in a prior study, providing an independent and relatively objective measure. This allowed for the separation of argument quality and individual bias, the extent to which people rated arguments congruent with their beliefs as stronger than incongruent arguments beyond what would be expected due to quality alone. The design of this study makes it possible to see whether participants are still sensitive to the quality of arguments when they are able to choose which arguments to view.

In the present study, the first set of participants in the Choice phase were allowed to select which items to read from a list of Pro and Con arguments for each issue. This allowed us to calculate a bias toward viewing congruent arguments in order to model its effect on attitude change. The second set of participants in the Matched phase were yoked with participants holding similar views (prior attitude and affective commitment) and presented with arguments chosen by these participants. This allowed us to separate the effects of choice from the effects of merely reading and rating a set of arguments. Overall, the unique design of this study makes it possible to model attitude change in terms of three factors of interest: proportion of arguments selected/viewed, objective argument quality, and individual rating bias.

Methods

Participants

Participants for Phase 1 (Choice) were 161 students (99 female) ranging in age from 18 to 27 years old, M=20; participants for Phase 2 (Matched) were 96 students (60 female) between 18 and 29 years old, M=21. Students were 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.

Materials

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

The arguments were a subset of materials used in Study 1, consisting of 6 supporting (Pro) and 6 opposing (Con) arguments for each of the four issues listed above. Attitude measurements and affective involvement measurements were calculated as described in Study 1.

Procedure

The present study comprised two separate phases that varied between participants: in Phase 1 (Choice), participants chose arguments to read from a list of Pro and Con arguments. In Phase 2 (Matched), participants' initial Attitude and Affective involvement scores were used to match each Phase 2 participant with a participant from Phase 1 with similar scores. Phase 2 participants then viewed the same arguments in

the same order as the Phase 1 participant to which they were yoked. The goal of this yoked design was to compare argument ratings and attitude change of Phase 1 participants who chose their own arguments to read versus Phase 2 participants with similar attitudes who read the exact same arguments. This design allowed us to examine whether the ability to choose which arguments to examine impacts the way participants evaluate and assimilate evidence on these topics.

Phase 1 (Choice). As in Study 1, the experiment included three parts: initial collection of attitude and affective involvement measurements, the presentation and rating of arguments, and the subsequent collection of attitude and affective involvement measurements. Measurement of attitudes and affective involvement proceeded exactly as in Study 1. However, the procedure during the intermediate stage in which participants read arguments diverged somewhat from that in the previous study.

Following the initial collection of attitude and affective involvement measurements, each participant was asked to read and rate arguments for three randomly chosen issues from the original set of four. For each of these three issues, participants were presented with a list of Pro and Con arguments in a labeled list including a topic statement for each argument (i.e., “Animal research is regulated to protect animals from mistreatment.”). Initially, one Pro and one Con argument were available. Participants selected one argument, read the full text of that argument, and rated it from “Weak” to “Strong” using a slider bar. Numeric values ranged from 0 to 100 but were not visible to participants.

After reading the first argument, the remaining unread argument and one additional Pro and Con argument were available from a list of arguments. An example

screen shot from the survey is shown in Figure 4.1. This procedure continued, adding two additional arguments after each selected argument was viewed and rated. Participants thus had a total of 6 Pro and 6 Con arguments available for each issue.

Following a randomly chosen half of the arguments, participants were instructed to type into a text box what they were thinking about as they rated the previous argument. They were allowed to advance to the next section after one minute, but could spend as much time as they desired typing these responses.

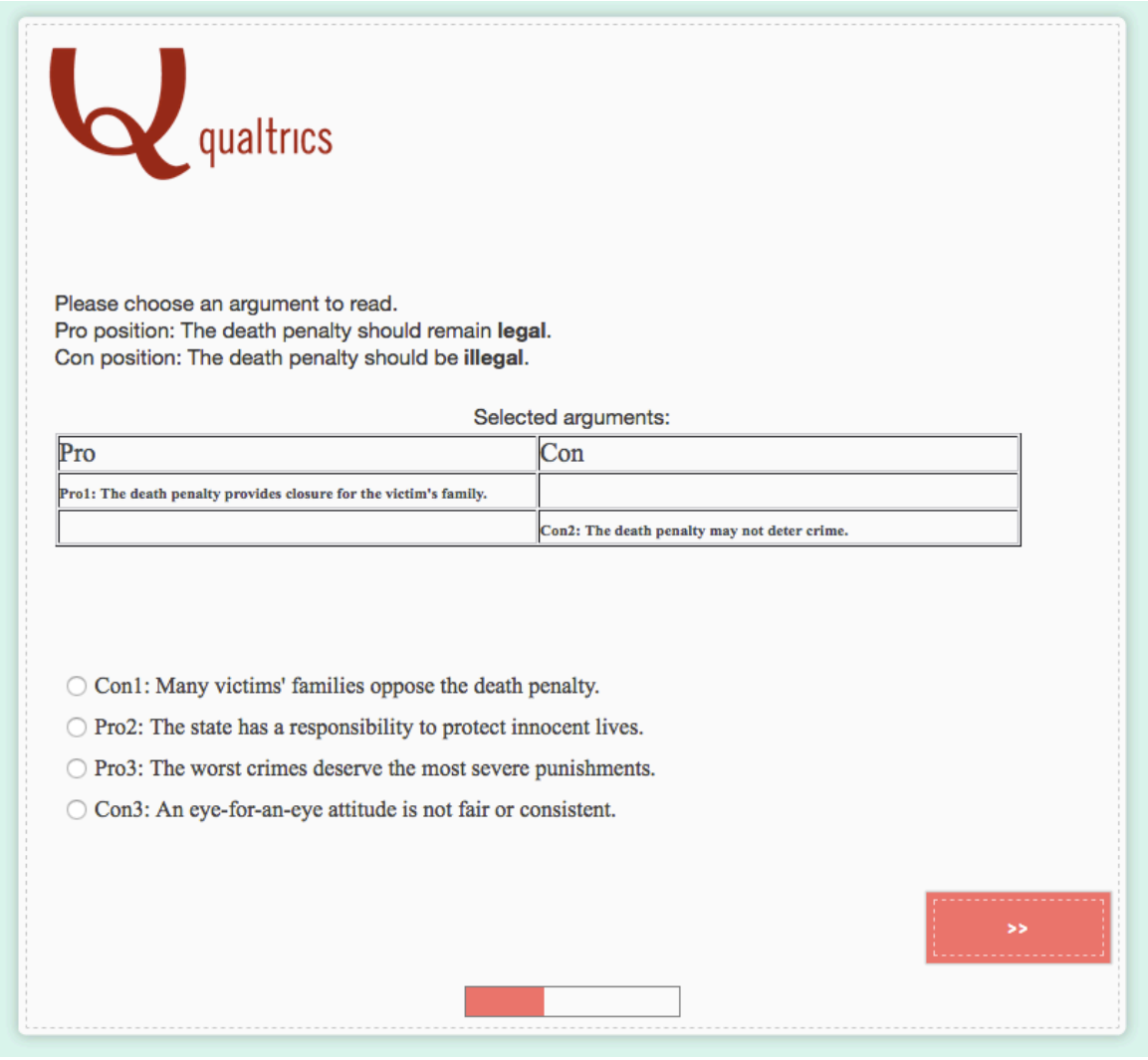

Figure 4.1. Screen shot showing sample participant selected arguments and list of available arguments for the death penalty issue.

After reading and rating arguments regarding all three issues, participants were once again asked to rate their attitudes on the four issues tested in the first stage of the study. Next, they completed a brief political knowledge quiz to assess their political sophistication and two questions to assess open-mindedness. Political sophistication ranged from 0 to 5 and was based on participants' scores on the political knowledge quiz (i.e., number correct out of 5 items). Finally, participants read a debriefing page that explained the goal of the study and provided links to the websites used for the argument texts.

Phase 2. As in to Phase 1, the experiment included three parts: initial collection of attitude and affective involvement measurements, the presentation and rating of arguments, and the subsequent collection of attitude and affective involvement measurements. The collection of attitude and affective involvement measurements proceeded exactly as in Phase 1; the procedure in the intermediate stage differed somewhat.

Following the initial collection of attitude and affective involvement measurements, each participant was asked to read and rate arguments for three randomly chosen issues from the original set of four. For each issue, each Phase 2 participant was matched with a participant from Phase 1. The matching process was done separately for each issue and involved the division of Phase 1 participants into four groups of approximately the same size. Division was based on scores for attitude (Pro or Con) and affective involvement (low or high). Participants in Phase 2 were thus matched with a Phase 1 participant with a similar attitude (Pro or Con) and a similar level of affective involvement (low or high). Phase 2 participants then viewed the exact

same arguments chosen by their yoked match from Phase 1. Arguments were presented one at a time and rated as described above from “Weak” to “Strong” with a slider bar. As in Phase 1, participants were prompted to type their thoughts following half of the arguments.

As in Phase 1, after reading all of the arguments, Phase 2 participants re-rated their attitudes on all four original issues, completed a five-item political knowledge quiz, responded to two questions on open-mindedness, and read the debriefing page.

Analysis

Attitudes were scaled from -5 to 5, with -5 representing the opinion most opposed to the issue statement (e.g. “The death penalty should be banned.”) and 5 representing maximal agreement with the issue statement. Affective involvement ranged from 0 (least involvement) to 5 (most involvement). Political sophistication ranged from 0 to 5. Items where participants spent too long reading the argument text (more than 3 standard deviations from the mean log time, 388 seconds) were removed from the analysis (28 of 2848 items from Phase 1, 9 of 1700 items from Phase 2).

Participant variables from Phase 1 and Phase 2 were analyzed to ensure uniformity across experimental phase. Prior attitude and affective involvement for each issue were compared using linear regression models. A linear model of prior attitude as a function of issue and phase showed that while prior attitude varied by issue, there was no significant interaction of issue and phase, nor was there a significant difference between Phase 1 and Phase 2 prior attitude. A linear model of affective involvement similarly showed an effect of issue and not experimental phase. Political sophistication

scores were compared using a Chi-squared test with simulated p-values. Values did not significantly differ between Phase 1 and Phase 2 (X-squared = 4.71, p = .45).

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 and backward model selection was used to determine the best-fitting model. Accordingly, more complex models were compared with more parsimonious models using ANOVA in R. Models were fit with random intercepts and slopes for subjects and items (viz. arguments). The reported models are those that included statistically significant experimental predictors of argument rating and were not statistically different from more complex models (generally using cut-off $p < .01$, but trending predictors are also reported).

Models of attitude change were analyzed 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 approximately equivalent to selecting all predictors with a significant p value in the model ANOVA.

For all models, argument polarity (Pro/Con) was sum coded. Issue (Abortion/Assisted suicide/Animal testing/Death penalty) was treatment coded. Abortion was used as a baseline for Issue because participants' average attitude change for this issue was close to zero.

Results and Discussion

Argument selection

Argument selection was modeled as a function of participants' prior attitudes to assess whether or not they tended to select more arguments in line with their prior attitudes, that is, whether they exhibited a congeniality bias.

To appropriately model the proportion variable, a GLM with a binomial function was used to predict the number of Pro and Con arguments selected (modeled as successes and failures). A model ANOVA with a Chi-squared test was used to test the fit of each model. AIC comparison showed that a model with categorical position was not significantly different from a model with continuous prior attitude. A model allowing position to interact with affective involvement and political sophistication failed to reveal effects of either variable, or significant interactions with prior attitude. The best model predicting argument selection is shown in Equation 4.1:

Frequency(Pro), Frequency(Con) ~ Prior attitude (4.1)

Overall, participants chose to read slightly more arguments that were consistent with their initial attitudes. Participants selected 52% congruent and 48% incongruent arguments on average. This proportion is displayed in Figure 4.2.

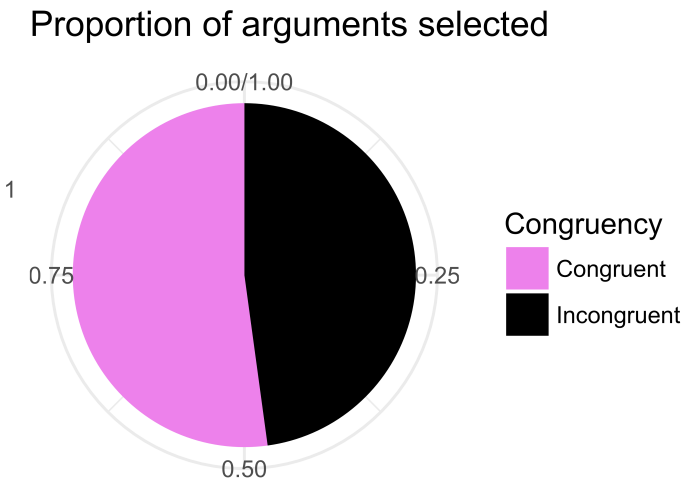

Figure 4.2: Pie chart showing proportion of congruent (pink) and incongruent (black) arguments selected by participants for each issue.

The pattern of argument selection was in line with prior studies (Taber & Lodge, 2006), albeit with a weaker congeniality bias. Besides the weaker bias toward attitude congruent arguments, the present study diverged from Taber and Lodge (2006) in that our participants' argument choice was not modulated by political sophistication or by affective involvement. One potential reason for this discrepancy is the source of the arguments. Participants in the present study were told that arguments came from online debate sites, whereas participants in Taber & Lodge (2006) could see that arguments were from potentially polarizing interest groups such as the NRA, Brady Anti-Handgun Coalition, and the platforms of American political parties. It is also possible that because students were aware they were participating in a Psychology or Cognitive science experiment, they made an effort to behave in line with researchers' imagined expectations (see Chaiken, Liberman, & Eagly, 1989 for a discussion of accuracy motivation). In an optional feedback question, some students responded that they

specifically sought out counterarguments in order to improve their understanding of an issue.

In sum, Phase 1 participants displayed a weak congeniality bias, in line with accounts of motivated reasoning. It is notable that this bias emerged even in an educated sample of students participating in an experiment pool of social science studies. These findings are mostly in line with a meta-analysis of the congeniality bias (Hart et al., 2009), which found a reduced bias for participants with greater motivation for accuracy and with higher open-mindedness.

Argument rating

Based on the results of Study 1 and Study 2, as well as prior research (Edwards & Smith, 1996; Taber et al., 2009), participants were expected to show an argument rating bias, rating attitude congruent arguments as stronger than incongruent arguments. This tendency could potentially vary depending on whether participants were able to select which arguments to view.

To test for the presence of an argument rating bias and to assess the effects of choice, we modeled participants' argument ratings with a linear mixed effects model with predictors of experimental phase (Choice/Matched), argument polarity (Pro or Con), prior attitude, affective involvement, issue, and political sophistication. Model selection resulted in Equation 4.2. Argument rating was predicted by a 3-way interaction of Prior attitude, Argument polarity, and affective involvement, with the slope of the Prior attitude x Argument polarity interaction being steeper for participants with high affective involvement, as shown in Figure 4.3. Allowing the model to further interact with Phase did not significantly improve the model (X^2 [8, N = 4511] = 11.21, p = .19).

Rating ~ Prior attitude x Argument polarity x Affective involvement (4.2)

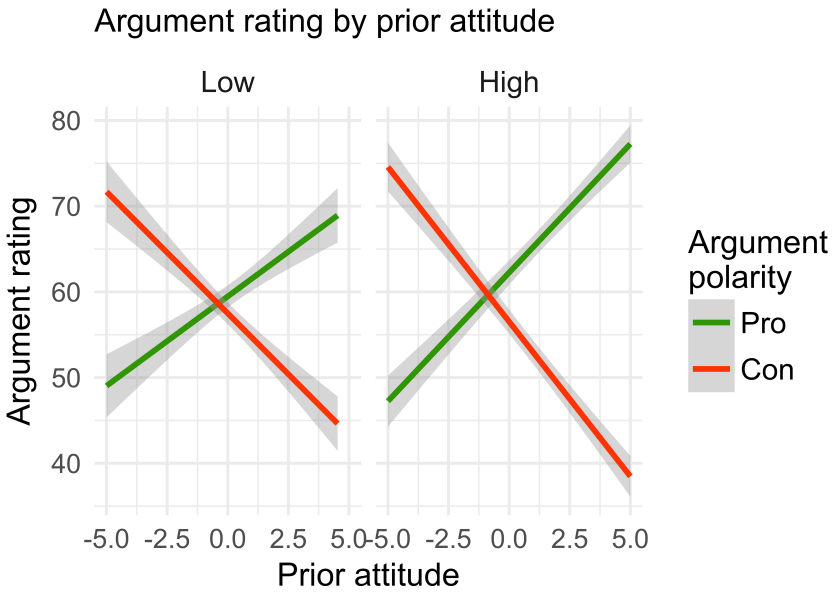

Figure 4.3. Average argument rating as a function of prior attitude (-5 most opposed, 5 most in favor of the position statement). Green lines represent Pro arguments. Red lines represent Con arguments. Ratings for participants with low and high Affective involvement (0 to 4.5, 4.5 to 9) are shown on the left and right respectively.

Table 4.1. Model results for Equation 4.2

Factor	Estimate	Std. Error	df	t-value	p-value
Intercept	58.749	1.727		34.01	< 0.001
Prior attitude	-0.444	0.491	1	-0.91	0.365
Argument polarity	-0.194	1.635	1	-0.12	0.906
Affective involvement	-0.103	0.230	1	-0.45	0.654
Prior attitude x Argument polarity	1.104	0.525	1	2.10	0.036
Prior attitude x Affective involvement	0.062	0.076	1	0.82	0.414
Argument polarity x Affective involvement	0.310	0.224	1	1.39	0.166
Prior attitude x Argument polarity x Affective involvement	0.263	0.081	1	3.25	0.001

Argument ratings in the present study were thus sensitive to the same factors as in Study 1 and Study 2 in displaying an attitude congruency bias. Participants on the Pro side of an issue rated Pro arguments higher than Con arguments, and those on the Con side rated Con arguments higher than Pro arguments. As in our previous studies,

this attitude congruency bias was largest for the participants with the most extreme positions. The present study also revealed a further interaction with affective involvement that results because these attitude congruency biases were greater among participants who reported feeling more strongly about the issues. The latter finding was present as a non-significant trend in the previous studies, and may be evident here because the pooling of Phase 1 and Phase 2 participants yielded a more powerful design.

The fact that the model was not improved by Phase suggests that argument ratings were affected by similar factors when participants chose which arguments to view as when the arguments a given participant viewed were selected by the computer. In both cases participants displayed an attitude congruency bias, preferring arguments that bolstered their own attitudes over those that undermined them. This suggests that the cognitive processes that underlie attitude congruency effects operate similarly when participants are confronted with information presented by others and when they actively seek out information themselves.

Attitude change

Attitude change was first modeled with all experimental data, then separately for the Choice phase and the Matched phase. The corrections and calculations described below were performed separately for each dataset.

As in the preceding studies, there was an overall regression toward the mean in the post-treatment attitude measurements. As in Study 1 and Study 2, we determined a correction factor by using participants' initial attitude measurements to predict their attitude change at the end of the survey. This average slope, calculated only for

measurements where the participants did not read arguments, was subtracted from the data to be analyzed, effectively correcting for regression to the mean. Because the Post-survey attitude / Prior attitude slope differed by Issue, this calculation was performed separately for each issue. Analyses below all use these corrected values for post-treatment attitude scores.

As in Study 3, an objective argument rating was calculated for each argument by averaging participants' ratings from Study 1 and Study 2. These values were then used in the computation of an Argument quality score intended to convey the relative quality of the Pro arguments a given participant viewed compared to that of the Con arguments they viewed. Argument quality was thus calculated by taking the average objective rating of Pro arguments each participant viewed minus the average objective rating of their Con arguments. Positive values reflect higher quality Pro arguments, on average, than Con arguments, while negative values reflect higher quality Con arguments. Importantly, Argument quality does not reflect the number of Pro versus Con arguments that participants viewed. The latter was reflected in a separate variable so that we could dissociate the impact of the *quantity* of Pro arguments a given participant viewed from the quality of the arguments viewed.

Rating bias was calculated as the residual rating of Pro arguments (average participant rating for Pro arguments – Argument quality for Pro arguments) minus the residual rating of Con arguments (average participant rating – Argument quality), and reflects the extent to which participants rated arguments differently than expected. That is, a positive value reflects bias in favor of Pro arguments and a negative value reflects bias in favor of Con arguments. This value does not reflect the number of arguments

participants chose. Because this value could not be calculated for issues where participants selected either all 6 Pro arguments or all 6 Con arguments, 22 out of 483 items were excluded from this analysis.

To compare overall rating bias, argument quality, and attitude change values across experimental phase, t-tests were performed on these variables for participants in Phase 1 and Phase 2. None of the tests revealed significant differences between groups.

The question of interest concerns the effects on Attitude change of three main predictive variables: the proportion of arguments participants either chose or were assigned to view, the objective quality of the arguments viewed, and participants' rating bias.

According to a Bayesian account of information processing, both the quality and the quantity of information should contribute to attitude change in the direction of the evidence. When participants view more arguments in favor of a position, they will change their attitudes in the direction of the evidence, especially for high quality arguments. This means that the proportion of Pro vs. Con arguments should be related to attitude change. If higher quality arguments are more persuasive, then we should also see an effect of Argument quality. Because higher positive values reflect an overall higher quality of Pro arguments seen, positive Argument quality scores would be expected to attitude change values that are more positive (toward the Pro side), while negative Argument quality values should lead to attitude change scores which are more negative more negative (toward the Con side).

According to motivated accounts of information processing, both the congeniality bias (the tendency to choose belief-congruent arguments) and biased argument rating predict attitude change in the direction of one's prior attitude. Under this account, we should see that Bias predicts attitude change, with positive bias (toward Pro arguments) leading to attitude change toward the Pro side of an issue and negative bias (toward Con arguments) leading to attitude change toward the Con side. The congeniality bias should not necessarily interact with this factor, but may contribute an independent effect similar to the effect described above. That is, exposure to more pro-attitude information leading to "polarizing" is predicted under both motivated and Bayesian accounts.

To assess the effects of these three variables on attitude change, first all three predictors as well as issue were allowed to interact in a linear model of attitude change. A Phase variable was also allowed to interact with these predictors. Backward model comparison was used to select the best model predicting attitude change. This model is shown in Equation 4.3 and reflects all predictors with $p < .03$.

Attitude change ~ Argument proportion x Phase + Objective quality x Phase + Rating bias + Argument proportion + Issue (4.3)

Because Phase interacted with predictors of interest, attitude change was modeled separately for Phase 1 (Choice) and Phase 2 (Matched) data, as described in more detail below.

Choice phase

A linear regression model was created to predict attitude change in the Choice phase from Argument proportion, Objective quality, Rating bias, and Issue. Affective involvement and political sophistication were allowed to interact with each independent

predictor to determine whether these effects were modulated by participants' individual "motive and opportunity" to further reason in favor of or against the arguments, and to determine whether the effects of argument quality or bias varied by issue. Neither factor reached significance, either alone or as part of an interaction. The initial additive model is shown in Equation 4.4, with results shown in Table 4.2.

Attitude change ~ Argument proportion + Objective quality + Rating bias + Issue (4.4)

Table 4.2. Model results for Choice phase, Equation 4.4

Factor	Estimate	Std. Error	df	t-value	p-value
Intercept	-0.253	0.112		-2.25	.025
Argument proportion	0.152	0.054	1	2.83	< .005
Objective quality	-0.063	0.102	1	-0.62	.538
Rating bias	0.324	0.044	1	7.39	< .001
Issue			3		
(Assisted suicide)	0.574	0.124		4.62	< .001
(Animal testing)	0.289	0.140		2.07	.039
(Death penalty)	0.169	0.253		0.67	.505

Matched phase

As above, a linear regression model was created to predict attitude change in the Matched phase from Argument proportion, Objective quality, Rating bias, and Issue. Affective involvement and political sophistication were again allowed to interact with each independent predictor. Neither factor reached significance, either alone or as part of an interaction. The initial additive model is shown above in Equation 4.4. Because none of the individual issues reached significance, Issue was removed as a predictor in the model. The revised additive model is shown in Equation 4.5. Results are shown in Table 4.3.

Attitude change ~ Argument proportion + Objective quality + Rating bias (4.5)

Table 4.3. Model results for Matched phase, Equation 4.5

Factor	Estimate	Std. Error	df	t-value	p-value
Intercept	0.000	0.059		0.00	1.000
Argument proportion	-0.025	0.060	1	-0.42	.674
Objective quality	0.159	0.060	1	2.67	.008
Rating bias	0.177	0.060	1	2.94	.004

The attitude change results for the Choice phase support a motivated account, where participants' attitude change depends on the proportion of arguments selected and participants' bias toward one side of an issue. The proportion of arguments viewed did not contribute to measures of attitude change in the Matched phase, indicating that perhaps the act of choosing arguments in a biased fashion sets off biased processing that can lead to attitude change in line with one's prior attitudes. Studies 1 and 2 showed that attitude change is sensitive to the amount of evidence presented, so it may seem surprising that argument proportion was not a significant predictor of change in the Matched phase. However, participants' tendency to select a greater proportion of congruent arguments was very small in this study, meaning that individuals in the Matched phase on average viewed an even number of Pro and Con arguments or only one additional attitude congruent argument due to this bias.

In both the Choice and the Matched phase, participants' argument rating bias modeled attitude change in the direction of the bias: a bias toward Pro arguments models change toward the Pro position of an issue and vice versa for a Con argument bias. This is again in line with a motivated account of attitude change and is similar to the findings of Study 3.

Also similar the findings of Study 3, objective argument quality modeled attitude change for participants in the Matched phase. This is an important finding within the

unique yoked paradigm: participants were sensitive to the quality of arguments when they did not select which arguments to view, whereas those participants who selected their own arguments did not appear sensitive to the quality of those arguments. This does point to a special role for information choice, possibly leading to motivated reasoning in this scenario vs. more rational reasoning when participants do not choose which content to view.

Further work could explore the underlying cognitive processes at work during selective information exposure. It could be that participants choose familiar content that only serves to reinforce their existing attitudes, meaning that quality matters less because there is less new information being consumed. Alternatively or in addition to this process, participants may form an initial judgment of incompatible arguments based only on a small summary and then not process the text as thoroughly as they would if it were presented without participant choice.

The emergence of bias in response to information choice indicates a potential important difference between situations in which people are presented with a curated collection of information vs. situations where people choose what information they would like to consume. The mere act of choosing content, perhaps by selecting news articles according to their headlines or listening to only certain radio or television shows, can reinforce a belief in one's own attitudes, enhancing the potential for individual polarization for controversial issues.

Chapter 5
Summary

This series of experiments presented participants with arguments pertaining to controversial socio-political issues. The quantity of information was varied so that sometimes participants viewed one-sided evidence, either all congruent or incongruent with their initial attitudes on the issue; and sometimes they viewed a mix of supporting and disconfirming evidence. This allowed us to test the effects of the quantity of information on attitude change.

Participants were asked to rate each argument on its strength, providing a preliminary estimate of argument quality in the first two studies. Argument ratings showed an attitude congruency bias, where congruent arguments were rated more highly. Because this rating measure was confounded with bias, Chapters 3 and 4 utilized ratings from the prior studies to create a more precise estimate of objective argument quality and rating bias to separate these individual effects on attitude change.

In each study, we tested for the effects of affective involvement (how much participants cared about and had thought about an issue) and political sophistication on measures of bias and on attitude change. Political sophistication did not appear to play a role in these processes, and affective commitment emerged only as a predictor of a stronger argument rating bias for those participants who reported greater commitment to an issue. In Chapter 3, we explored the role of knowledge in attitude but were unable to find evidence of a significant relationship.

In Chapter 4, we explored a new yoked paradigm, allowing one set of participants to choose which information to view. In this Choice phase, we were able to test for the presence of a congeniality bias, where participants choose to view

information congruent with their attitudes. By yoking a second set of participants to the first based on their similarity in attitudes and affective involvement, we separated out the effects of information choice on biased evidence processing.

Chapter 1

Participants viewed a combination of mixed and one-sided evidence for six socio-political issues, rating individual arguments on how weak or strong they were. Participants' attitudes toward each issue and affective involvement were measured at the beginning of the experimental survey, and their attitudes were measured again after viewing all information.

We found evidence of an attitude congruency bias, where ratings depended on participants' initial attitudes and argument polarity: arguments congruent with initial attitudes were rated as stronger than incongruent arguments. We next tested for the presence of polarization, a measure of participants' beliefs moving toward the extremes of the rating scale for an issue. Polarization appeared to emerge only in response to congruent arguments, whereas de-polarization (a trend toward the center of the rating scale) emerged in the presence of highly-rated incongruent arguments. Because this pattern did not match accounts of motivated reasoning, we measured attitude change toward the Pro or Con position of an issue instead.

We found that participants' attitudes changed in response to highly-rated arguments regardless of their congruency with prior attitudes. Attitude change was also related to experimental condition (Pro, Con, or Mixed evidence), with a greater quantity of Pro arguments leading to change in participants' attitudes toward the Pro position, Con arguments leading to change toward the Con position, and Mix leading to little

overall change, but eliciting enhanced de-polarization. Because biased processing of evidence in the form of an argument rating bias did not appear to lead to polarizing, we concluded that the attitude change better supported a Bayesian account of belief revision: information is judged in terms of its perceived quality, but overall participants are sensitive to the amount and direction of information, updating their attitudes in the direction of the information viewed.

Chapter 2

The paradigm of this study was similar to that of the Chapter 1 experiment, with participants viewing mixed and one-sided evidence for a subset of the original issues. To examine the effects of active arguing on attitude change, to assess whether participants spent more time arguing against incongruent arguments, and to examine whether positive or negative statements accompanied congruent/incongruent arguments, we added participant text response to this study.

We replicated the attitude congruency bias found in Chapter 1 as well as the attitude change behavior, with participants again showing sensitivity to the quantity and quality of information instead of displaying polarizing in response to highly rated congruent arguments. We did not find evidence of a disconfirmation bias, where participants spend more time reading incongruent arguments. We also did not find that participants generated more bolstering statements (positive sentiment) in response to congruent arguments or denigrating statements (negative sentiment) in response to incongruent arguments. Instead, we found that sentiment varied with argument rating, with more positive sentiment following highly-rated arguments, regardless of their congruency with prior attitudes. Taken together, these results further support a

Bayesian account of information processing in which the quality of evidence (and its direction) matters more than its interaction with people's attitudes.

Some questions still remain after these two studies: Why do these results differ from those showing that the most biased, politically sophisticated participants polarize in response to mixed evidence? Do people treat information differently because they have stronger vs. weaker feelings about an issue or because they have different amounts of knowledge about the topic? And to what extent does argument rating reflect individual bias vs. the underlying quality of the information?

The participants in these studies are an educated group of undergraduate students at a top research university who are aware that they are participating in psychology-related experiments. Compared to political science students, these participants may be less motivated to argue and more motivated to pursue accuracy in their judgments and more open to attitude change. If the nature of the experiment leads more objective processing, this could point toward the importance of encouraging thoughtful consideration of new information.

To address the questions of the role of affect vs. knowledge and the effects of argument quality vs. individual bias, Chapter 3 specifically measures participants' knowledge for two of the issues in these experiments, and Chapters 3 and 4 explore the role of specific biases calculated in a new manner. Results at the end of Chapter 2 suggest that attitude change following highly-rated arguments may be partially due to the arguments' objective quality, but also due to individual bias driven by extant attitudes.

Chapter 3

In this study, participants viewed and rated only mixed evidence for one of two issues: animal testing and the death penalty. Their topic knowledge was measured for both issues, as well as their affective involvement and other measures the same as in the previous studies. To form a measure of objective argument rating, ratings from participants in Studies 1 and 2 were combined and averaged to obtain an average rating for each argument. These objective values were subtracted from participants' ratings in Study 3 to obtain a new measure, rating bias, which measured the extent to which participants rated arguments differently than what would otherwise be expected based on quality. This more sensitive measurement of bias was predictive of attitude change, showing biased processing leading to attitude change in the direction of the bias for the first time in this series of experiments. Along with bias, objective argument quality was also predictive of attitude change, again showing that participants were sensitive to this variable independent of their own attitudes.

Neither affective involvement nor topic knowledge appeared to play a significant role in either biased processing of arguments or attitude change. There was a relationship between affective involvement and the argument rating bias, but this did not further affect attitude change. Overall, results pointed toward a role of bias in attitude change, but also supported Bayesian information processing, indicating that some elements of both may be present. This study used materials from two issues where participants were overall very balanced in their views and may have been more receptive to changing their attitudes. Each participant also received only mixed evidence, which could further encourage sensitivity to the merits of each argument. As

seen in previous studies, this pool of participants seems to exhibit consideration of the information and does not tend to argue strongly against incongruent argument.

In an effort to elicit stronger potential effects of bias, Chapter 4 allows some participants to choose which arguments to read, giving them an opportunity to favor congruent arguments, which may set off further biased processing of evidence.

Chapter 4

This study comprised two phases: the Choice phase, where each participant was able to choose arguments to read from a set of labeled Pro and Con items, and the Matched phase, where each participant received a set of arguments to read from a matched Choice phase participant. This allowed us to test for the presence of a congeniality bias, a tendency to choose attitude-congruent arguments, in the Choice phase, and examine its potential effects on argument rating and attitude change. It also allowed us to test whether the act of choice enhances other biases by comparing the behavior participants in the Choice phase to those in the Matched phase.

We found evidence of a small but significant congeniality bias, with participants choosing slightly more congruent arguments to read when given a choice. This bias was modeled by participants' attitudes, but not their affective involvement. We replicated the attitude congruency bias from prior studies, where participants rate congruent arguments as stronger. This bias did not differ by experiment phase, indicating that it was not sensitive to participants' argument choice.

The rating bias variable significantly modeled attitude change for both phases, indicating that participants' biased treatment of evidence led to attitude change in line with their preference for Pro or Con arguments separate from their quality. The

proportion of Pro vs. Con arguments chosen affected attitude change only for the Choice condition, indicating that the act of choice itself may elicit an additional bias not captured by other measures. Both of these findings support a motivated account of attitude change.

As in Chapter 3, objective argument quality and argument rating bias were calculated as separate predictive factors of attitude change. Objective quality modeled attitude change in the Matched phase only, in line with the results of Study 3. The fact that objective quality did not reach significance for data in the Choice condition suggests that different cognitive processes are at work; individuals may be less sensitive to evidence in the presence of choice because they have already made a judgment in the act of choosing.

These findings are important especially in view of the current political climate in the United States and elsewhere, where polarization appears to be on the rise and individuals have more information available for consumption than ever before. The potential to choose one's own information for any topic, especially controversial social and political issues, may have significant drawbacks if the process of choice exacerbates divides between groups on opposing sides. It is difficult to imagine how individuals might agree on which information is high in quality and decide to view an appropriate mix of content that is congruent and incongruent with pre-existing attitudes. Individual choice seems to influence whether people process information in a motivated or information-processing fashion, showing that rationality may be elusive when individuals choose their own content.

References

Alwin, D. F., & Tufiş, P. A. (2016). The Changing Dynamics of Class and Culture in American Politics: A Test of the Polarization Hypothesis. *The ANNALS of the American Academy of Political and Social Science*, *663*(1), 229-269.

Amelino-Camelia, G., Freidel, L., Kowalski-Glikman, J., & Smolin, L. (2011). OPERA neutrinos and relativity. *arXiv preprint arXiv:1110.0521*.

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).

Batson, C. D. (1975). Rational processing or rationalization? The effect of disconfirming information on a stated religious belief. *Journal of Personality and Social Psychology*, *32*(1), 176.

Bayes, T. (1763/1958). Studies in the history of probability and statistics: IX. Thomas Bayes's Essay towards solving a problem in the doctrine of chances. *Biometrika, 45*, 296-315.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). New York: Guilford Press.

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.

Ecker, U. K., Lewandowsky, S., Fenton, O., & Martin, K. (2014). Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation. *Memory & cognition*, *42*(2), 292-304.

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

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

Festinger, L., Riecken, H. W., & Schachter, S. (1956). *When prophecy fails: A social and psychological study of a modern group that predicted the destruction of the world.* Minneapolis, MN: University of Minnesota Press.

Gal, D., & Rucker, D. D. (2010). When in doubt, shout! Paradoxical influences of doubt on proselytizing. *Psychological Science, 21,* 1701–1707.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. New York: Chapman & Hall.

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

Good, I. J. (1980). Some history of the hierarchical Bayesian methodology. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. M. Smith (Eds.), *Bayesian statistics* (pp. 489–519). Valencia: Valencia University Press.

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: a meta-analysis of selective exposure to information. *Psychological bulletin, 135*(4), 555.

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

Kahan, D. M. (2016). The politically motivated reasoning paradigm, part 1: What politically motivated reasoning is and how to measure it. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*.

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy, 1*(1).

Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational behavior and human decision processes, 56*(1).

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

Layman, G. C., Carsey, T. M., & Horowitz, J. M. (2006). Party polarization in American politics: Characteristics, causes, and consequences. *Annu. Rev. Polit. Sci.*, *9*, 83-110.

Lee, M. D. (2006). A hierarchical Bayesian model of human decision-making on an optimal stopping problem. *Cognitive Science, 30*, 555–580.

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.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55-60).

McCright, A. M., & Dunlap, R. E. (2011). Cool dudes: The denial of climate change among conservative white males in the United States. *Global environmental change*, *21*(4), 1163-1172.

Miller, A. G., McHoskey, J. W., Bane, C. M., & Dowd, T. G. (1993). The attitude polarization phenomenon: Role of response measure, attitude extremity, and behavioral consequences of reported attitude change. *Journal of Personality and Social Psychology*, *64*(4), 561.

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

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.

Stephens, R. (2015, May 7). *The data that threatened to break physics: What does a rational scientist do with an impossible result?* Retrieved from http://nautil.us/issue/24/error/the-data-that-threatened-to-break-physics

Sunstein, C. R. (2002). The law of group polarization. *Journal of political philosophy, 10(2)*.

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

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755-769.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. Trends in Cognitive Science, 10, 309-318.

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.

Wicklund, R. A., & Brehm, J. W. (1976) *Perspectives on cognitive dissonance*. Psychology Press.