Rational Irrationality: Modeling Climate Change Belief Polarization Using Bayesian Networks

John Cook

Global Change Institute, The University of Queensland, Australia

School of Psychology, University of Western Australia, Australia

Stephan Lewandowsky

School of Experimental Psychology and Cabot Institute, University of Bristol, UK

School of Psychology, University of Western Australia, Australia

Keywords: belief polarization, Bayes theorem, Bayesian updating, climate change

Mailing address: Global Change Institute, The University of Queensland, St Lucia, QLD 4072, Australia

Email: j.cook3@uq.edu.au

Abstract

Belief polarization is said to occur when two people respond to the same evidence by updating their beliefs in opposite directions. This response is considered to be ‘irrational’ because it involves contrary updating; a form of belief updating that appears to violate normatively optimal responding, as for example dictated by Bayes’ theorem. In light of much evidence that people are capable of normatively optimal behavior, belief polarization presents a puzzling exception. We show that Bayesian Networks, or Bayes Nets, can simulate rational belief updating. When fit to experimental data, Bayes Nets can help identify the factors that contribute to polarization. We present a study into belief updating concerning the reality of climate change in response to information about the scientific consensus on anthropogenic global warming (AGW). The study used representative samples of Australian and U.S. participants. Among Australians, consensus information partially neutralized the influence of worldview, with free market supporters showing a greater increase in acceptance of human-caused global warming relative to free market opponents. In contrast, while consensus information overall had a positive effect on perceived consensus among U.S. participants, there was a reduction in acceptance of human-caused global warming for strong supporters of unregulated free markets. Fitting a Bayes Net model to the data indicated that under a Bayesian framework, free market support is a significant driver of beliefs about climate change and trust in climate scientists. Further, active distrust of climate scientists among a small number of U.S. conservatives drives contrary updating in response to consensus information amongst this particular group.

Rational irrationality: Modeling Climate Change Belief Polarization Using Bayesian Networks

Imagine two people with differing beliefs about a publicly contentious issue, such as climate change. One person accepts human-caused global warming while the other is dismissive of the human role in climate change. How might the two react if told that there is a strong scientific consensus—involving over 95% of all domain experts (Doran & Zimmermann, 2009; Anderegg, Prall, Harold, & Schneider, 2010) and peer-reviewed climate research (Oreskes, 2004; Cook et al., 2013)—regarding human-caused global warming? The person who accepts the presence of a consensus might be expected to strengthen their beliefs. However, how will the same information be processed by the "dismissive"? One possibility is that the "dismissive", already distrustful of climate scientists, views the consensus as confirmation of a conspiracy or "group think" among scientists, rather than as a reflection of the strength of the scientific evidence. They may thus emerge more unconvinced when informed about the scientific consensus. While both parties received the same evidence, their beliefs changed in opposite directions.

This phenomenon is known as belief polarization, and it occurs when people receiving the same information update their beliefs in diverging directions. While belief polarization may occur relatively infrequently (Kuhn & Lao, 1996), it has been observed across a range of contentious issues. In a classic study, supporters and opponents of the death penalty became more set in their views in response to mixed information that both supported and rejected the death penalty (Lord, Ross, & Lepper, 1979). Likewise, in response to a report describing a nuclear breakdown, supporters of nuclear power focused on the fact that the safeguards worked, whereas opponents focused on the breakdown (Plous, 1991). When religious believers and non-believers were exposed to a fictitious report disproving the Biblical account of the Resurrection, the religious believers increased their faith whereas non-believers accepted the report and became more skeptical (Batson, 1975). Similarly, news stories about health impacts from climate change have been shown to have polarizing impact across party lines. Information about health impacts “backfire” among Republicans, who showed lower identification with potential victims, whereas Democrats showed greater identification with victims and increased concern about climate impacts in response to the same information (Hart & Nisbet, 2011).

Belief polarization can also be observed in response to evidence supporting a single point of view. When people receive evidence that contradicts their prior basic beliefs, it can result in strengthening of beliefs contrary to the evidence. This is known as contrary updating, or the “worldview backfire effect” (Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012). To illustrate, Nyhan and Reifler (2010) showed participants mock newspaper articles that suggested that weapons of mass destruction (WMDs) had been found in Iraq after the 2003 invasion, before issuing a correction that WMDs had not been found. This correction induced belief polarization: conservatives became more likely to believe that Iraq had WMDs, whereas the reverse was observed with liberals.

This type of belief polarization in response to unambiguous evidence is commonly considered an “irrational” response; that is, a deviation from Bayesian belief updating, which is considered to be the normative, optimal way in which a person should change their beliefs in light of new evidence (Gerber & Green, 1999). A Bayesian rational agent is thought to update prior beliefs on the basis of new evidence, to form a revised “posterior” set of beliefs. Beliefs can only be updated in the direction suggested by the evidence—hence, at first glance, a rational agent could not show an increased belief in a hypothesis (e.g., that there were WMDs in Iraq) when being presented with contrary evidence (i.e., that no WMDs were found).

We argue in this article that although a simple Bayesian view cannot accommodate belief polarization, a more sophisticated variant involving Bayesian belief networks can give rise to polarization even though agents behave entirely "rationally" (Jern, Chang, & Kemp, 2014). We begin by formalizing Bayesian belief updating before introducing Bayesian networks.

Bayes Theorem describes how a rational agent updates its prior belief in a hypothesis *H*, P(*H*), in response to new evidence *E*. The updated or posterior degree of belief in a hypothesis *H* is expressed as probability P(*H*|*E*). Bayes Theorem stipulates that the updated belief is a function of people's prior belief P(*H*) and the conditional probability P(*E*|*H*) of observing the evidence *E* given *H* is true.

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \quad (1)$$

According to Bayesian expectations, two people with differing prior beliefs should update their beliefs in the same direction when presented with the same information (Bartels, 2002).

Belief polarization presents a conundrum in light of the large body of evidence that people update their beliefs in accordance with the rules of Bayesian inference. Examples of Bayesian inference include sensorimotor skills (e.g., estimating the velocity of an approaching tennis ball; Körding & Wolpert, 2004), category learning (Sanborn, Griffiths, & Navarro, 2010) and predicting final quantities such as box office grosses, lifespan and duration of a Pharaoh's reign from a current value (Griffiths & Tenenbaum, 2006). For example, in an iterative experiment where participants repeatedly estimated lifespans from a person's age, the distribution of estimated values was consistent with the prior distribution of lifespans, indicating Bayesian reasoning among individuals (Lewandowsky, Griffiths, & Kalish. 2009). Conversely, there is also evidence that in some contexts, people make predictions in a non-Bayesian manner, placing undue weight on prior beliefs (Kahnemann & Tversky, 2013).

It is therefore not surprising that a number of studies have attempted to explain belief polarization under a Bayesian framework (Bullock, 2009). Past studies have employed constrained forms of Bayesian updating, whereby the principal tenets of Bayes theorems were augmented by non-Bayesian processes (Zimper & Ludwig, 2009; Gerber & Green, 1999; Wilkins, 2011; Dixit & Weibull, 2007; Andreoni & Mylovanov, 2012).

Our approach, by contrast, simulates belief polarization within a fully Bayesian approach, through the use of Bayesian Networks, also known as Bayes Nets (Pearl, 2000). The key to this approach lies in the introduction of other belief components into a Bayes Net (Jern et al., 2014). In our case, we include variables such as “worldview” and trust in scientists. Worldview has been variously operationalized as people’s score on a liberal-conservatism scale (Ding, Maibach, Zhao, Roser-Renouf, & Leiserowitz, 2011; McCright, Dunlap, & Xiao, 2013), or as the degree to which they endorse free markets (Heath & Gifford, 2006; Lewandowsky, Gignac, & Oberauer, 2013), or as party affiliation (Hardisty, Johnson, & Weber, 2010; Hart & Nisbet, 2011; Malka, Krosnick, & Langer, 2009) or as their position on a dichotomy between people who are “hierarchical-individualists” and those who are “egalitarian-communitarian” (Kahan, Jenkins-Smith & Braman, 2011a). Although those different operationalizations tap diverse aspects of people’s worldview, as a first approximation all those belief variables seem to explain an overlapping share of the variance of people’s attitudes towards climate change.

Trust in climate scientists has been observed to be a driving factor behind polarization over climate change (Malka et al., 2009). Similarly, trust in experts and perception of expertise is moderated by how consonant the expert's views are with a person's own worldview (Kahan et al., 2011a). Accordingly, political "ideology" correlates highly with beliefs about climate change (Heath & Gifford, 2006; Kahan et al., 2011b; Lewandowsky, Oberauer, & Gignac, 2013; Lewandowsky et al., 2013), with people who endorse unregulated free markets being more likely to reject evidence from climate science. Even among meteorologists, a survey has found that political ideology, defined on a scale from conservative to liberal in this instance, was one of the variables most strongly related to climate views (Stenhouse, Maibach, & Cobb, 2013). By incorporating extra variables, belief polarization is potentially enabled as these additional belief variables moderate people's interpretation of the evidence. From here on, we use people's endorsement of free markets (REFs) as a concise proxy variable for their personal and political worldviews.

Bayes Nets

A Bayes Net is a graphical network of causally linked variables, also referred to as Belief Nets because the probability assigned to each variable represents the degree of belief in each state of the variable. Each variable is represented by a node in the network, while the directed lines represent dependence relationships between them. To illustrate, Bayes Theorem is represented by the Bayes Net in Fig. 1(a), with the evidence E having a probabilistic dependence on the hypothesis H. We assume that H can take one of two possible values, with H=0 or H=1, and for the sake of this example, prior probabilities P(H=0)=.6 and P(H=1)=.4. We also assume, in this example, that the conditional probability P(E=1 | H=1) is 0.8. In other words, it is highly likely that if the hypothesis is true, then evidence for the hypothesis will be observed.

Suppose that such evidence has been observed, hence P(E=1) is set to 1. Bayes Theorem dictates that the updated, posterior belief P(H=1) should now be 0.73. In our graphical representation, this updated belief "flows backwards" through the arrow in Fig. 1(a) and changes the probabilities of different values of *H*. This principle holds true for all Bayes Nets regardless of their complexity: Each arrow captures a probabilistic (and causal) dependence, and when evidence is observed, this information "flows backwards" to update the probability distribution of antecedent nodes.

Fig. 1(b) shows the change in belief in *H* in response to evidence, for different prior beliefs in *H*. Regardless of prior belief, belief updating is always in the same direction, consistent with the evidence. The Bayes Net in Fig. 1(a) cannot model belief polarization: given constant conditional probabilities, there exists no distribution of prior beliefs that could cause P(H=1) to be updated in the *opposite* direction given the observation E=1.

When additional relevant variables are entered into the Bayes Net, some (but not all) configurations of Bayes Nets are capable of producing polarization (Jern et al., 2014). To illustrate, consider Fig. 1(c). Jern et al. (2014) applied this Bayes Net to Batson's (1975) study, in which participants were asked to read a story undermining Christian beliefs. Participants with strong Christian beliefs became more certain of their belief while participants with weak Christian beliefs further weakened their beliefs. The Bayes Net in Fig. 1(c) is able to capture this observed response with the extra variable *V* representing religious worldview, and *H* corresponding to the hypothesis that Jesus is the son of God. Jern et al. (2014) argued that a possible explanation of the Batson (1975) result is that strong believers *expect* their faith to be frequently challenged with contrary (but false) evidence whereas someone with little religious belief expects to see evidence against religion. Hence one's worldview influences beliefs about a hypothesis as well as one's interpretation of evidence.

Applying a Bayes Net to climate change beliefs

The focal hypothesis *H* in our Bayes Net was people's acceptance that humans are causing the Earth's climate to change, a view on which 97% of publishing climate scientists have converged on, based on the evidence (Doran and Zimmermann, 2009; Anderegg et al., 2010; Cook et al., 2013). The evidence variable *E* therefore was the scientific consensus on human-caused global warming. We chose consensus to represent evidence for several reasons: First, consensus is known to be an effective form of quasi-scientific evidence in the eyes of the public at large (Petty & Wegener, 1999). Second, presentation of information about the scientific consensus has been shown to increase acceptance of climate science, demonstrating a causal link between perceived consensus and climate attitudes (Lewandowsky, Gignac, & Vaughan, 2013). Perception of consensus has been observed to be a "gateway belief", predicting numerous climate-related beliefs (Ding et al., 2011; McCright et al., 2013; van der Linden et al, 2015; Stenhouse et al, 2013). Third, unlike the nuanced landscape of actual scientific evidence, people's perception of the consensus among scientists can be summarized in a single number and hence is readily represented by a single node in a Bayes Net.

Turning to the additional belief variables, following Jern et al.'s (2014) Bayes Net 1(b) we introduce trust (in the evidence or its source) as a third variable, represented by *T* in our Bayes Net—in this case, trust in the 97% of climate scientists whose consensus constitutes the evidence node. A final significant factor influencing climate attitudes is worldview, represented by *W*. Worldview is known to influence climate attitudes (Heath & Gifford, 2006; Kahan et al., 2011b; Lewandowsky, Oberauer, & Gignac, 2013; Lewandowsky, Gignac, & Vaughan, 2013), which is represented by a directed link between *W* and *H*. A further directed link between *W* and *T* captures the influence of worldview on trust in climate scientists (Malka et al., 2009). We use free market support as a proxy for worldview. Finally, E is linked to H in the standard manner, and there is an additional link between T and E representing the moderating influence of trust in belief updating. This extra link is implied in the religious belief Bayes Net from Jern et al. (2014), where expectation of faith-challenging evidence (presumed false and hence untrustworthy for those with religious belief) is crucial for modeling of belief polarization. The 'Worldview Bayes Net' shown in Fig. 2(a) captures the known links between our set of variables.

The relationship between variables is captured by the Bayes Net's conditional probabilities in Fig. 2(b) to (d). Conditional probabilities will be estimated by fitting the Bayes Net to the data from the experiment that is presented in this article. Fig. 2(b) shows approximate example values based on previous studies (Lewandowsky, Gignac, & Vaughan, 2013; Kahan et al, 2011a; Malka et al, 2009), and in particular the polarization model of Jern et al. (2014).

The known influence of those variables on climate belief is represented in the conditional probabilities shown in Fig. 2(b & c). High free market support [P(W=1) approaching 1] is expected to correspond with low belief in anthropogenic global warming (AGW) [P(H=1) approaching 0]. Similarly, high free market support [P(W=1) approaching 1] corresponds to low trust in climate science T [P(T=1) approaching 0]. These conditional probabilities are labeled P(H=1 | W=1) and P(T=1 | W=1).

Based on these conditional probabilities, the Bayes Net predicts that strong free market supporters will decrease their belief in AGW in response to evidence for AGW. This example of contrary updating is driven largely by the conditional probability P(E=1 | T=0 & H=0), highlighted in Fig. 2(d). This represents the expectation that evidence *for* AGW will be observed even though AGW is believed to be false. This echoes the Jern et al. (2014) interpretation of the

Batson (1975) results, suggesting that the backfire effect among religious believers was driven by the expectation that their faith would be challenged with (presumably false) evidence. Suspicion about the motives of information sources has been associated with being less easily influenced by misinformation (Lewandowsky, Stritzke, Oberauer, & Morales, 2005). Similarly, extreme suspicions about scientists may predispose people to presume the existence of an (unwarranted) consensus among climate scientists, perhaps because they are conspiring to create a "hoax" (Inhofe, 2012).

By contrast, participants with low free market support are expected to increase their belief in AGW in response to evidence for AGW, as there is no conflict between personal ideology and the evidence. The other conditional probabilities $P(E{=}1 \mid T{=}1 \;\&\; H{=}0)$ and $P(E{=}1 \mid T{=}0 \;\&\; H{=}1)$ reference low probability outcomes, given the correlation between belief in AGW (H=1) and trust in climate scientists (T=1), and we do not expect them to be a significant factor. Consequently, the Bayes Net can explain belief polarization based on plausible values of prior probabilities derived from the existing literature. Our experiment explores whether people polarize in response to consensus information, and by permitting estimation of the conditional probabilities underlying such belief updating, it may highlight the cognitive processes underlying polarization.

Method

We report an experiment that presented scientific-consensus information and expert opinion to Australian and U.S. participants and measured subsequent acceptance of human-caused global warming, as well as worldview, trust in scientists, perceived consensus and perceived expertise. The theoretical expectations of the Worldview Bayes Net were tested by fitting the model to the observed prior and posterior values of W, T, H and E. In this article, single letter variables refer to nodes in the Bayes Net while full words (e.g., Worldview, Trust) refer to experimental design variables.

Design. The experiment featured a 2 × 2 between-subjects design with two independent variables – a consensus intervention and an expertise intervention, which was included for exploratory reasons. By fully crossing the presence or absence of each intervention, the design featured a control group (no intervention), a consensus group (no expertise intervention), an expertise group (no consensus intervention), and a group that received a combined consensus/expert intervention. The consensus intervention (Fig. 3) featured text and an infographic explaining that there is 97% agreement among climate scientists that humans are causing global warming (Doran & Zimmerman, 2009; Anderegg et al., 2010; Cook et al., 2013). The expertise intervention featured a quote about climate change from a highly credentialed climate scientist along with a photo of the scientist. Intervention text and survey items are available in the Supplemental Information.

Participants. The experiment was conducted online with U.S. (N=325, conducted February 2013) and Australian (N=400, conducted April 2013) samples. Participants were recruited via the online survey firm Qualtrics.com, which specializes in representative online surveys. Qualtrics samples their participants from a panel maintained by uSamp.com (for more details, see the uSamp.com website), using propensity sampling based on gender, age and region, which has been shown to reasonably approximate representativeness (Berrens et al., 2003). Participants were compensated with cash-equivalent points by Qualtrics. The two countries were chosen in order to replicate and compare results of earlier research (Lewandowsky, Gignac, & Vaughan, 2013; Ding et al., 2011; McCright & Dunlap, 2011). All survey items were compulsory. Only participants who passed attention filter questions associated with the experimental manipulations (ensuring attentive reading of intervention text) and completed all items were included in the final sample. The overall group of participants were selected to approximate a representative sample, with participants randomly allocated to experimental conditions.

Test items. The survey comprised 33 items plus 2 attention filter questions. Six constructs were measured: Worldview, Trust in climate scientists, Perceived Expertise of scientists, Perceived Consensus, acceptance of AGW (Climate) and the percentage Attribution of human activity to long-term climate trends. Five additional items measuring support for mitigation policies were included at the end of the Australian survey and are not analyzed in this article. Five items measuring support for free markets, developed by Heath and Gifford (2006), were used as a proxy for Worldview. Trust in climate scientists and perceived expertise of scientists used 5 items each, adapted from Ohanian (1990). Climate attitudes were measured using 5 items previously used by Lewandowsky, Gignac, and Vaughan (2013). Attribution of human activity used 3 items representing 3 long-term climate metrics (percentage from 0 to 100% that human activity contributed to warming temperatures, sea level rise and extreme weather events) that were also taken from Lewandowsky, Gignac, and Vaughan (2013). Five constructs (Worldview, Trust, Perceived Expertise, AGW, Attribution) were measured by averaging survey items while Perceived Consensus was derived from a single survey item.

Results

Our analysis examined the interplay between Worldview and the design variables, namely country and the consensus and expertise manipulations. Data was analysed with R (R Development Core Team, 2011), using the Car package in R to perform an ANOVA with country and the consensus intervention as fully-crossed factors, and the continuous Worldview variable as a further continuous predictor. All reported F-values are based on Type II sums of squares to accommodate differences in group size. Worldview was standardized to mean zero and standard deviation one.

The expertise intervention caused a small but significant increase in Perceived Consensus, $F(1, 717) = 6.29$, $p = 0.01$, and Climate, $F(1, 717) = 5.06$, $p = 0.02$. However, the effect was additive with respect to the other experimental variables on all measures (i.e., interactions were non-significant, shown in Table S2). As this analysis is concerned with the interplay between worldview and the experimental manipulation, the expertise independent variable is thus not considered further and analysis focused on comparison of the control and consensus intervention groups.

Table 1 summarizes the influence of the independent variables (consensus intervention, country, Worldview) as well as their interaction terms on five dependent variables: Perceived Consensus, acceptance of AGW, attribution, trust, and perceived expertise. All *p*-values and statistical information are available in the table and are not explicitly repeated in the text.

Perceived Consensus. For both Australian and U.S. participants, Perceived Consensus in the control group averaged below 60%, consistent with other research reporting that people under-estimate the scientific consensus (Nisbet & Myers, 2007). Fig. 4(a) and 4(b) show that the perception of consensus varied significantly with worldview. Table 1 demonstrates a main effect of the consensus intervention on Perceived Consensus (control 57%, consensus intervention 91%). There was also a significant three-way interaction between Worldview, the consensus intervention and country on Perceived Consensus, indicating a difference between the two countries in how consensus information changes perceived consensus across the ideological spectrum. Fig. 4(a) shows how the increase in Perceived Consensus amongst Australian participants was highest among conservatives while Fig. 4(b) shows that for Americans, the increase in Perceived Consensus was uniform for different levels of worldview.

Climate and Attribution. The main effect of the consensus intervention was significant on both Climate and Attribution. The three-way interaction between Worldview, Country and the consensus intervention was significant for Attribution and close to significance for Climate. Fig. 4(c) and 4(e) shows that for Australian participants, consensus information partially neutralized the influence of worldview on rejection of climate science. Fig. 4(d) and 4(f) show that for U.S. participants, the interaction between Worldview and Consensus was in the opposite direction, such that greater endorsement of free markets was associated with a reduced effectiveness of the consensus intervention. This indicates that while consensus information partially neutralized Worldview in Australia, in replication of Lewandowsky, Gignac, and Vaughan (2013), it had a polarizing effect in the U.S. (The online supplement reports separate ANOVAs for each country that provide statistical confirmation of the statements about the data made here in the text.)

Trust and Perceived Expertise. Across both countries, trust in climate scientists was significantly and negatively correlated with Worldview. Fig. 4(g) and 4(h) shows that the stronger the support for free markets, the lower the Trust. The consensus intervention had a significant main effect in increasing Trust. In addition, there was an interaction between the consensus intervention and country, indicating different reactions between U.S. and Australian participants. Consensus information activated further distrust of scientists among Americans with high free market support, while the consensus intervention had no effect on trust for the Australian sample.

Perceived expertise varied significantly with Worldview, consistent with the finding of Kahan et al. (2011a) that the perceived expertise of climate scientists is influenced by political ideology. The consensus intervention had no overall significant effect on Perceived Expertise. Fig. 4(i) and 4(j) show that consensus information slightly lowered Perceived Expertise among Americans, except for those who were least likely to endorse unregulated free markets, whereas it had a slight positive effect among Australians.

Fitting Bayes Net to observations

We fitted the Worldview Bayes Net to the data, which were rescaled to the range 0-1 to represent probabilities of each Bayes Net variable. The Bayes Net was fitted to each country's data separately, obtaining a unique set of Bayes Net parameters for each country. Each participant's (rescaled) support for free market was input for W, trust in scientists for T, belief in AGW for H and perception of consensus for E. Participants who were shown no consensus information (control condition) were used for "prior" values in the Bayes Net, whereas participants shown the consensus information were used for "posterior" values. While indicated perceived consensus was used for E for control participants, E was set to 1 for posterior participants. Given that the attention filter for the consensus intervention ensured the participant remembered the actual level of consensus, the difference between setting E to 1 and using posterior data for E was negligible, and for simplicity we therefore set E to 1.

The Bayes Net was fitted to the data using the Bayes Toolbox in Matlab. SIMPLEX was used to minimize the RMSD discrepancy between the experimental data and the Bayes Net predictions for prior and posterior W, T, H and E. This allowed the estimation of 8 parameters representing the conditional probabilities or relationship between the variables of the Bayes Net, from 1177 data points with the U.S. data and 1400 data points with the Australian data. Note that the Bayes Net minimizes the discrepancy across the group of prior and posterior data and hence doesn't require prior and posterior values from the same individuals.

The conditional probability obtained from the model fit that is of greatest interest is P(E=1 | H=0 & T=0). This represents the expectation that there is a scientific consensus about AGW while also believing that AGW is false and while distrusting climate scientists. We interpret this probability to represent the expectation that climate scientists will "collude" to agree on human-caused global warming—thereby creating an impression of consensus—even though AGW is false. This parallels the reasoning of Jern et al. (2014), who interpreted belief polarization over challenges to religious belief to reflect believers *expecting* to encounter false evidence attacking their faith.

Fig. 5 shows the modeled prior and posterior beliefs in H (acceptance of AGW) and T (trust in climate scientists) given the estimated conditional probabilities. Within the Bayes Net, the independent variable W (worldview) varies from 0 (no support for free markets) to 1 (strong support for free markets). The greyed area represents the range of worldview values capturing 95% of the experimental data when they are rescaled to be commensurate with the 0-1 range required by the Bayes Net. The fact that Fig. 5(a) and 5(c) show smaller grey areas compared to 5(b) and 5(d) indicates that the Australian sample has a narrower distribution of worldview values, with fewer strong free-market supporters than in the U.S. sample. Fig. 5(a) captures the worldview-neutralizing effect of consensus information on the Australian sample, with a greater increase in climate belief occurring for strong supporters of free markets. In contrast, Fig. 5(b) captures the polarization in the U.S. sample, with strong free market supporters showing contrary updating in response to consensus information. Fig. 5(c) captures the lack of change in trust for Australians over the range of observed worldview values; note that the model's extrapolation as

W approaches 1 is well beyond the observed values of *W*. However, Fig. 5(d) shows the drop in trust among Americans in the observed range of *W* values.

In contrast to previous applications that presupposed the conditional probabilities of the Bayes Net (Jern et al., 2014), here the relationships between the Bayes Net variables were estimated empirically (see Fig. 6) and were found to be consistent with the relationships between these variables observed in previous studies. Our emphasis is not on the accuracy with which the Bayes Net reproduced the observed patterns, but rather on what the Bayes Net can tell us about the qualitative patterns of belief updating. In particular, we focus on the fact that $P(E=1 \mid H=0 \;\&\; T=0)$ is high for the U.S. sample (.84), indicating that participants with low belief in H and low trust T (mainly people with high W, viz. political conservatives) nonetheless have a high expectation that a consensus among climate scientists exists, perhaps because they will collude to fabricate a consensus or because they engage in “group think”. In contrast, $P(E=1 \mid H=0 \;\&\; T=0)$ was comparatively low for Australians (.48), indicating that conservatives in the Australian sample have a lower expectation of a “fabricated” consensus.

Discussion

Summary of results

The present experiment replicated previous results investigating the role of worldview and perceived scientific consensus on climate beliefs. We observed that worldview is a dominant influence on climate beliefs, and that providing consensus information raises perception of consensus. The detailed pattern of belief updating on the Climate and Attribution items differed between countries and was a function of Worldview, with consensus information having a slightly worldview-neutralizing effect on Australians but a backfire effect on a small proportion of Americans with strong conservative (free-market) values.

The observed polarization among U.S. conservatives meshes with some previous results, but stands in contrast to others. On the one hand, consensus messaging was found to have a worldview neutralizing effect on U.S. participants in van der Linden et al. (2015), with conservatives exhibiting a greater increase in climate belief compared to liberals. One possible contributor to the contrasting result is that van der Linden uses party affiliation as a proxy for political ideology rather than free-market support as used in this study. Another contributor may be differences in the intervention content, which is significantly shorter and less informative (only mentions scientific consensus with less climate science information) in van der Linden (2015) and uses different imagery (pie-chart) to communicate the consensus.

Similar to the present study, Kahan et al. (2011a) found that consensus information was potentially polarizing, with hierarchical individualists (i.e., mainly people who endorse free markets) attributing less expertise to climate scientists relative to egalitarian communitarians (who believe in regulated markets). The worldview-neutralizing effect on Australians that was observed here replicates existing work involving an Australian sample by Lewandowsky, Gignac, and Vaughan (2013).

Our results also support other precedents, namely that trust in climate science is lower among conservatives (Malka et al., 2009). One novel element to our research is the observed change in trust in response to consensus information. Among Australians, trust was unchanged. Among U.S. participants, the consensus-intervention polarized trust with free-market supporters becoming more distrustful of scientists when informed about the scientific consensus. We cannot offer an explanation why the two countries differ in this article. One potential limitation of our results involves the generally small effect sizes of our experimental manipulation (see Table 1), especially compared to the large effect of worldview. In response, we note that the size of the effects may well be commensurate with the brevity of our manipulation: We presented participants with a brief passage and a simple graphical stimulus. We consider it remarkable that this subtle manipulation had a statistically detectable effect, however small.

Implications from Bayes Net modeling

The relative success of the Worldview Bayes Net in capturing the response to consensus information suggests that it is possible to simulate seemingly "irrational" responses such as belief polarization as normatively rational, Bayesian responses (cf. Jern et al., 2014). Specifically, Bayesian networks show that when other prior beliefs such as trust in evidence and worldview are incorporated into belief updating models, contrary updating can be simulated under a normatively-optimal framework.

Using the principal variables known to affect people's attitudes towards climate change, we found that the Bayes Net could be fit to experimental data and qualitatively reproduce the pattern in the prior (control) and posterior (consensus intervention) data. By estimating the underlying conditional probabilities, the Worldview Bayes Net offers a possible explanation of the psychological processes driving belief polarization. The estimated conditional probabilities from the Bayes Net showed that political conservatives who are dismissive of AGW exhibited an active distrust of climate scientists, with the distrust greater in the U.S. sample relative to the Australian sample. We suggest that the high distrust among U.S. conservatives is indicative of a degree of skepticism that some authors have identified as being present in conspiratorial thought (Keeley, 1999). The estimate implies that a person who does not accept AGW and distrusts scientists would, with high certainty, expect scientists to manufacture the appearance of a scientific consensus. The findings of the Worldview Bayes Net are therefore arguably consistent with previous findings of a small but significant link between the rejection of human-caused global warming and conspiratorial thinking (Lewandowsky, Oberauer, & Gignac, 2013; Lewandowsky, Gignac, & Vaughan, 2013; Smith & Leiserowitz, 2012; Lewandowsky et al., 2015).

This study presents opportunities of further research using Bayes Nets to investigate belief updating with respect to polarizing issues. One insight from the Worldview Bayes Net is recognition of the powerful influence of worldview on both scientific beliefs and trust in scientists. It follows that any intervention that can reduce the influence of worldview may indirectly also reduce or reverse polarization. Examples may be interventions that emphasize how scientific information is not in conflict with personal ideology by framing it in world-consonant terms (e.g., Hardisty et al., 2010) or through self-affirmation (Cohen et al., 2007).

Conclusions

This study has demonstrated that belief polarization can be simulated within a normatively rational framework using Bayesian Networks. Fitting a Bayes Net model to experimental data involving the scientific consensus on climate change indicates that contrary updating is driven by worldview, which in turn influences trust in scientific sources. Specifically, an active distrust and expectation that scientists would “manufacture” a “fake” consensus drives contrary updating among some American conservatives (van der Linden, 2013). The Bayes Net model was also able to distinguish psychological differences between Australian and U.S. populations, finding that higher levels of distrust are evident in the polarized U.S. sample in contrast to the Australian sample. While Bayesian networks show that contrary updating is consistent with a normative framework, the question of whether the expectation of a manufactured scientific consensus could be considered rational is an open question.

Understanding why scientific messages lack efficacy or indeed may backfire among certain groups is of importance to scientists and science communicators, given the known role of perceived consensus as a gateway belief influencing a range of other climate attitudes (Ding et al., 2011; McCright & Dunlap, 2011; Stenhouse et al., 2013; van der Linden et al., 2015). The body of research into consensus messaging poses a complex, nuanced picture. Across both countries, consensus messaging significantly increases perceived consensus across the ideological spectrum. However, when it comes to changing other climate beliefs such as acceptance of AGW, the patterns of belief updating differ across countries. Consensus messaging is wholly positive in increasing acceptance of AGW with Australian participants, even partially neutralizing the biasing influence of worldview. While there is some evidence that consensus messaging also neutralizes ideology amongst U.S. participants (van der Linden et al., 2015), the present study finds evidence for belief polarization with a small number of conservatives exhibiting contrary updating.

References

Anderegg, W. R. L., Prall, J. W., Harold, J., & Schneider, S. H. (2010). Expert credibility in climate change. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 12107-12109.

Andreoni, J., & Mylovanov, T. (2012). Diverging opinions. *American Economic Journal: Microeconomics*, *4*(1), 209-232.

Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, *24*, 117-150.

Batson, C. D. (1975). Rational processing or rationalization - effect of disconfirming information on a stated religious belief. *Journal of Personality and Social Psychology*, *32*, 176-184.

Berrens, R. P., Bohara, A. K., Jenkins□Smith, H., Silva, C., & Weimer, D. L. (2003). The advent of Internet surveys for political research: A comparison of telephone and Internet samples. *Political analysis*, *11*(1), 1-22.

Bullock, J. (2009). Partisan bias and the bayesian ideal in the study of public opinion. *Journal of Politics*, *71*, 1109-1124.

Cohen, G. L., Sherman, D. K., Bastardi, A., Hsu, L., McGoey, M., & Ross, L. (2007). Bridging the partisan divide: Self-affirmation reduces ideological closed-mindedness and inflexibility in negotiation. *Journal of personality and social psychology*, *93*(3), 415.

Cook, J., Nuccitelli, D., Green, S.A., Richardson, M., Winkler, B., Painting, R., Way, R., Jacobs, P., & Skuce, A. (2013). Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental Research Letters*, *8*(2), 024024+.

Ding, D., Maibach, E. W., Zhao, X., Roser-Renouf, C., & Leiserowitz, A. (2011). Support for climate policy and societal action are linked to perceptions about scientific agreement. *Nature Climate Change*, *1*(9), 462-466.

Doran, P., & Zimmerman, M. (2009). Examining the scientific consensus on climate change. *Eos, Transactions American Geophysical Union*, *90*, 21-22.

Dixit, A., & Weibull, J. (2007). Political polarization. *Proceedings of the National Academy of Sciences*, *104*, 7351.

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

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. Psychological Science, *17*, 767–773.

Hardisty, D. J., Johnson, E. J., & Weber, E. U. (2010). A dirty word or a dirty world? Attribute framing, political affiliation, and query theory. *Psychological Science*, *21*(1), 86-92.

Hart, P. S., & Nisbet, E. C. (2011). Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Communication Research*, *39*(6), 701–723.

Heath, Y., & Gifford, R. (2006). Free-market ideology and environmental degradation - the case of belief in global climate change. *Environment and Behavior*, *38*, 48-71.

Inhofe, J. M. (2012). The greatest hoax: How the global warming conspiracy threatens your future. Washington, DC: WND Books.

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

Kahan, D., Jenkins-Smith, H., & Braman, D. (2011a). Cultural cognition of scientific consensus. *Journal of Risk Research*, *14*, 147-174.

Kahan, D. M., Wittlin, M., Peters, E., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. N. (2011b). The tragedy of the risk-perception commons: culture conflict, rationality conflict, and climate change. *Temple University Legal Studies Research Paper*, (2011-26).

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, *80*(4), 237.

Keeley, B. L. (1999). Of conspiracy theories. *The Journal of Philosophy*, *96*(3), 109-126.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244-247.

Kuhn, D., & Lao, J. (1996). Effects of evidence on attitudes: Is polarization the norm?. *Psychological Science*, 115-120.

Lewandowsky, S., Stritzke, W. G., Oberauer, K., & Morales, M. (2005). Memory for Fact, Fiction, and Misinformation The Iraq War 2003. *Psychological Science*, *16*, 190-195.

Lewandowsky, S., Griffiths, T. L. & Kalish, M. L. (2009). The Wisdom of Individuals: Exploring People's Knowledge About Everyday Events Using Iterated Learning, *Cognitive Science*, *33*, 969-998.

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106-131.

Lewandowsky, S., Gignac, G. E., & Vaughan, S. (2013). The pivotal role of perceived scientific consensus in acceptance of science. *Nature Climate Change*, *3*(4), 399-404.

Lewandowsky, S., Oberauer, K., & Gignac, G. E. (2013). NASA faked the moon landing—therefore,(climate) science is a hoax an anatomy of the motivated rejection of science. *Psychological science*, *24*(5), 622-633.

Lewandowsky, S., Gignac, G. E., & Oberauer, K. (2013). The role of conspiracist ideation and worldviews in predicting rejection of science. *PLOS ONE*, *8*(10), e75637.

Lewandowsky, S., Cook, J., Oberauer, K., Brophy, S., Lloyd, E. A., & Marriott, M. (2015). Recurrent Fury: Conspiratorial Discourse in the Blogosphere Triggered by Research on the Role of Conspiracist Ideation in Climate Denial. *Journal of Social and Political Psychology*, *3*(1), 142-178.

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

Maibach, E., Myers, T., & Leiserowitz, A. (2014). Climate scientists need to set the record straight: There is a scientific consensus that human-caused climate change is happening. *Earth's Future, 2*(5), 295-298.

Malka, A., Krosnick, J. A., & Langer, G. (2009). The association of knowledge with concern about global warming: Trusted information sources shape public thinking. *Risk Analysis*, *29*, 633-647.

McCright, A. M., & Dunlap, R. E. (2011). The politicization of climate change and polarization in the American public's views of global warming, 2001-2010. *Sociological Quarterly*, *52*, 155-194.

McCright, A. M., Dunlap, R. E., & Xiao, C. (2013). Perceived scientific agreement and support for government action on climate change in the USA. *Climatic Change*, *119*(2), 511-518.

Nisbet, M. C., & Myers, T. (2007). The polls—trends twenty years of public opinion about global warming. *Public Opinion Quarterly*, *71*(3), 444-470.

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, *19*(3), 39-52.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303-330.

Pearl, J. (2000). *Causality: models, reasoning and inference* (Vol. 29). Cambridge: MIT press.

Petty, R. E. & Wegener, D. T. (1999). Dual-Process Theories in Social Psychology (eds Chaiken. S. & Trope, Y.), Guilford Press.

Plous, S. (1991). Biases in the assimilation of technological breakdowns - do accidents make us safer. *Journal of Applied Social Psychology*, *21*(13), 1058-1082.

R Development Core Team (2011). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria.

Sanborn, A. N.; Griffiths, T. L. & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117*, 1144-1167.

Smith, N., & Leiserowitz, A. (2012). The rise of global warming skepticism: exploring affective image associations in the United States over time. *Risk Analysis*, *32*(6), 1021-1032.

Stenhouse, N., Maibach, E., & Cobb, S. (2013). Meteorologists' views about global warming: A survey of American Meteorological Society professional members. *Bulletin of the American Meteorological Society*. doi: 10.1175/BAMS-D-13-00091.1.

van der Linden, S. (2013). What a Hoax. *Scientific American Mind*, *24*(4), 40-43.

van der Linden, S., Leiserowitz, A., Feinberg, G., & Maibach, E. (2015). The Scientific Consensus on Climate Change as a Gateway Belief: Experimental Evidence. *PLoS ONE*. doi: 10.1371/journal.pone.0118489

Wilkins, J. (2011). Are creationists rational? *Synthese*, *178*(2), 207-218.

Zimper, A., & Ludwig, A. (2009). On attitude polarization under bayesian learning with non-additive beliefs. *Journal of Risk and Uncertainty*, *39*(2), 181-212.

Fig. 1. (a) Bayes Net visually representing Bayes Theorem with example conditional probabilities and prior/posterior belief in H. (b) Example of parallel updating in response to receiving evidence in a 2-node Bayes Net. (c) Bayes Net configuration from Jern et al. (2014) capable of producing belief polarization.

Fig. 2. (a) Worldview Bayes Net. *W* represents support for free markets, *T* represents trust in climate scientists, *H* represents the hypothesis that humans are causing global warming and *E* is the evidence for *H*: the scientific consensus on human-caused global warming. (b) Example conditional probabilities represent relationships between variables, approximately estimated based on previous studies.

Fig. 3. Intervention text communicating scientific consensus on human-caused global warming. The 97% figure has been independently confirmed by Doran and Zimmerman (2009), Anderegg et al. (2010), and Cook et al. (2013).

Fig. 4. Predicted response from linear regression of observed data. Triangles with dotted line represent control group, circle with solid line represents group receiving consensus intervention. Horizontal axis represents support for free market. Left column shows Australian data, right column shows USA data. (a, b) Change in perceived consensus. (c, d) Change in belief in AGW. (e, f) Percentage attribution of AGW to long-term climate trends. (g, h) Trust in climate scientists. (i, j) Perceived expertise of climate scientists.

Fig. 5. Bayes Net model output based on conditional probabilities estimated from data fit. Horizontal axis represents support for free market, where higher support corresponds to a more conservative worldview. Grey areas represent 95% range of the observed range of W values, demonstrating that model plots outside of grey areas represent extrapolation beyond the empirical data. (a) Belief in the hypothesis H (Australia). (b) Belief in the hypothesis H (USA). (c) Trust in climate scientists (Australia). (d) Trust in climate scientists (USA).

Fig. 6. Estimated conditional probabilities from fitting Bayes Net to experimental data for the U.S. sample (a) and Australian sample (b).