Updating Prior Beliefs Based on Ambiguous Evidence

Stephen H. Dewitt[1], David A. Lagnado[1] and Norman Fenton[2]

[1]Department of Experimental Psychology, University College London, 26 Bedford Way, WC1H 0AP

[2]School of Electronic Engineering and Computer Science, Queen Mary University of London, Mile End Rd, London E1 4NS

Abstract

This paper investigates a problem where the solver must firstly determine which of two possible causes are the source of an effect where one cause has a historically higher propensity to cause that effect. Secondly, they must update the propensity of the two causes to produce the effect in light of the observation. Firstly, we find an error commensurate with the 'double updating' error observed within the polarisation literature: individuals appear to first use their prior beliefs to interpret the evidence, then use the interpreted form of the evidence, rather than the raw form, when updating. Secondly, we find an error where individuals convert from a probabilistic representation of the evidence to a categorical one and use this representation when updating. Both errors have the effect of exaggerating the evidence in favour of the solver's prior belief and could lead to confirmation bias and polarisation.

Keywords: Bayesian; Confirmation Bias; Polarisation; Belief Updating; Base Rates

Two nations, X and Y are independently testing new missile technology. Each has made six detonation attempts: X has been successful twice and Y four times. You observe another detonation on the border between X and Y but cannot determine the source. Based only on the provided information, what is the probability that X (or Y) is the source of this missile? Further, what is your best estimate of the propensity for success of X and Y after this latest observation (i.e. the probability, for each nation, that a future missile they launch will detonate)?

The general form of this problem is ubiquitous in many areas of life. There are several possible causes of an effect (the problem readily extends to more than two causes), with one having a historically higher propensity[1] to cause that effect. The effect is then observed, but the source is unknown. The causal responsibility must be attributed, and the propensities updated. This latter requirement to provide an estimate not only for the individual case, but also a higher-level update of propensity, distinguishes this problem from Bayesian problems studied previously, which tend to focus only on assessing the lower level, individual instance, and assume a stable propensity estimate.

One common manifestation of the general form of this problem arises where two or more social groups are perceived to have a varying historical record of causing some negative social incident (e.g. a crime, an anti-social effect or a terror attack) and news media circulate that an incident has occurred, but the perpetrator (and therefore the social group) is not known.

The lower level question requires assessment of the probability that X and Y launched the missile. This is equivalent to many classical Bayesian word problems (e.g. Kahneman & Tversky, 1972; Casscells, Schoenberger & Graboys, 1978). However, in the present experiment this is substantially simpler because the chance of the missile being launched by either nation (before knowing of the successful detonation) is assumed equal. If there were some additional reason to suspect one nation more likely to have launched the missile, this question would have very similar properties to those classic problems, requiring integration of unequal priors (probability of launching a missile) with a diagnostic piece of information (the successful explosion and differential success). There are three reasons we assume an even prior for this element of the problem. Firstly, it allows us to observe how individuals deal with situations with no diagnostic information beyond that provided by the prior propensities, which has not been studied within this literature before. Secondly, it allows us to easily and naturally present the evidence in words only, without the need for including percentage values (Presenting a 50/50 chance precisely in words is considerably more straightforward than, for example, 70/30). This reduced need for arithmetic lessens one of the 'ecological validity' issues which have plagued the 'Bayesian word problem' literature (e.g. Cohen, 1981; Birnbaum, 1983; Welsh & Navarro, 2012). Finally, given that we were primarily interested in participant performance on the 'higher level' question, we wished to avoid unnecessary complication at this stage.

The 'higher level' question requires participants to integrate the successful detonation observation with the prior propensities to arrive at new propensity levels for both X and Y. This is a task which has not been studied in the Bayesian psychological literature so far. In the classic disease problem, it is roughly equivalent to updating the population disease proportion after assessing the individual patient. In the taxicab problem it is roughly equivalent to updating the proportion of cabs in the city which are green / blue, following assessment of the individual cab's most likely colour. In these problems it is either stated or assumed that the estimate provided for these propensities is stable. We examine a situation where this higher-level value is based upon a small number of observations and so a single new instance has a non-negligible impact.

---

[1] The rate of success observed by each nation in their six attempts may be considered by participants to be derived from an underlying 'true' propensity for success (Gigerenzer, Hoffrage & Kleinbölting, 1991).

Given that X and Y are exhaustive, mutually exclusive, and equally likely to launch the next missile, the estimation of each being the source of the missile (the 'lower level' question) is a relatively simple calculation. It follows from Bayes' theorem that the probability X was the source is

$$P(X\ is\ source) = \frac{P(X) \times 0.5}{\frac{1}{2}(P(X) + P(Y))} = \frac{P(X)}{P(X) + P(Y)}$$

where P(X) is X's prior propensity for success and P(Y) is Y's prior propensity.

The second, higher-level question, is non-trivial. A Bayesian network (BN) (see Figure 1) was constructed to model this. A BN is a directed graph whose nodes represent uncertain variables, and where an arc (or arrow) between two nodes depicts a direct causal or influential relationship (see Fenton & Neil [2012] for full details of BN's). In addition to the graph structure, each node has an associated probability table which defines the prior probability distribution for the associated variable, conditioned (where a node has parents) on its parent variables. Any time the state of a node is observed (e.g. the missile launch explodes successfully) the known value is entered into the BN and a propagation algorithm updates the probability distributions for all unobserved nodes. The 'Bayesian' in BN's is due to the use of Bayes' theorem in the underlying propagation algorithm.

The model in Figure 1 depicts the situation before observing the detonation, but with the information about previous missile launches. The central upper nodes give the probability distributions for successful detonation given an attempted missile launch by X and Y. These are updated from uniform priors automatically by the Bayesian propagation (in this case it simply uses the Binomial distribution assumption) in light of the data on previous launches (upper left and right nodes) and successful detonations (middle left and right nodes) by X and Y. The probability of each nation firing the next missile before we know if it successfully detonates or not, is modelled as a 50/50 chance (lower left node) as each have made an equal number of previous attempts. The central lower node depicts the probability that the next missile fired will detonate, given that we don't know who will launch it. This is therefore a conditional probability distribution combining P(X) and P(Y), conditioned upon the probability of each country firing the next missile (the lower left node).

The probability of the missile detonating is also modelled as a Boolean variable in the lower right node, with the 'true' value here simply being the mean of the central conditional distribution. This Boolean node allows us to make the observation in the problem that the missile has successfully detonated. Upon observing this, the BN automatically calculates the revised means to be 40.3% for P(X) and 65.1% for P(Y) (updated from 37.6% and 62.4% respectively before the observation).

A full description of the factors which influence the absolute change that the propensity of each nation undergoes as a result of the detonation observation cannot be included due to space constraints. For present purposes, given equal variance (and allowing for a 'truncated normal distribution' effect), the absolute change in success propensity for each nation after observing a successful detonation is identical i.e. as we see above, if P(X) increases by 2.7%, P(Y) also increases by 2.7%. This is not dependent upon the initial uniform prior assumption adopted here but is dependent upon the 50/50 assumption for the probability of X and Y launching the next missile.

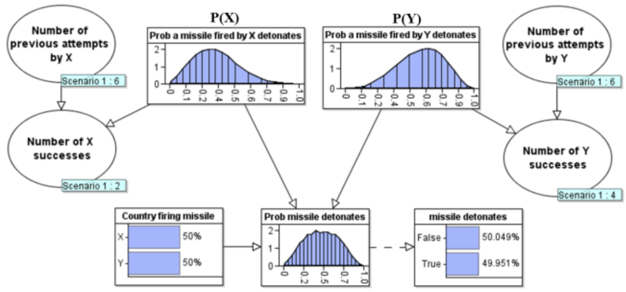

Figure 1. Bayesian Network Model of the problem

To our knowledge this problem is novel within the cognitive science literature. Expectations of how people will try to solve it can be derived from related work. Work on belief polarisation by Lord, Ross & Leper (1979), Rabin & Schrag (1999), and Fryer, Harms & Jackson (2016) has demonstrated that when asked to update an opinion (e.g. on the death penalty) based upon ambiguous evidence (e.g. an article) individuals frequently and erroneously perform a 'double update' where they first use their prior belief to interpret the ambiguous information and then use the interpreted form of the information, rather than the raw form, when updating. The present experiment differs from that paradigm in that we are evaluating propensities, not opinions, and the evidence we provide is 'objectively ambiguous' (there is no information other than the success of the explosion as to which nation launched the missile) while that paradigm uses 'subjectively ambiguous' evidence (an article on the topic which has no objective interpretation but instead a distribution of interpretations from 'pro' to 'con', with the mean being neutral). However, if this process generalises out of that paradigm, our participants may first interpret the evidence as being more likely to be Y based upon the prior propensities, and then use this interpreted form of the evidence when updating, leading to an increase in propensity for both nations, but a greater increase for Y.

Previous work also suggests a second erroneous process may be taken which would lead to an even greater exaggeration of the evidence in favour of Y. The problem introduces two important and distinct forms of probability. The first (each nation's prior probability of success), is based upon a set of observed frequencies (known as 'aleatory' probability). The second (the probability of the source of the missile explosion being X or Y) is a form of probability known as degree of belief, 'epistemic' probability, or 'single event' probability (Gigerenzer, 1994).

A large and extensive body of research over several decades (Gigerenzer & Hoffrage, 1995; Cosmides & Tooby, 1996; Macchi, 2000; Evans et al, 2000; Girotto & Gonzalez, 2001; Sloman, Over, Slovak & Stibel, 2003; Brase, 2008) has demonstrated consistently that many individuals have difficulty dealing with subjective beliefs about single events. More specifically to this context, several authors (e.g. Brase, Cosmides & Tooby, 1998; Cosmides & Tooby, 1996; Brase, 2008) have demonstrated that individuals will show a strong reluctance to 'carve up' a single indivisible unit (such as the missile explosion) into fractions based upon subjective belief. This may be particularly potent in this case as assessment is of a past event, which must in actuality have been caused wholly by either X or Y. These authors would therefore advocate presenting this problem with multiple unknown explosions. It is likely that solvers would ascribe some of those successful explosions to X, but the majority to Y, and roughly in proportion to their prior proficiencies. However, it is fundamental to the nature of this problem that assessment is made of a single event, and given such circumstances occurs in real settings, is worthy of study.

If individuals do indeed represent the single missile explosion categorically, rather than probabilistically then it is plausible that the event will be ascribed to the more likely cause (to Y). Relative to normative Bayesian standards, as automatically calculated in the BN of Figure 1, this exaggerates the weight of the evidence and would lead to increasing only P(Y), and not increasing P(X) at all.

The application of either of the above theorised errors would increase prior beliefs out of proportion with the weight of the evidence, acting as a form of confirmation bias (Kunda, 1990; Nickerson, 1998; Hahn & Harris, 2014) or, acting to polarise two individuals with different prior beliefs (Lord, Ross & Leper, 1979, Rabin & Schrag, 1999; Fryer, Harms & Jackson, 2015). The double update error (the use of the probabilistic interpreted form of the evidence) would show a lesser departure from normative standards than the categorical error (the use of the categorical interpreted form of the evidence).

This experiment firstly aims to empirically test how people tackle this problem, using a version of the scenario introduced at the start of this paper, where the propensities of two nations for detonating missiles must be updated upon observation of one instance of such a detonation (and where one nation is the historically more common cause of detonations). We predict that at least two specific non-normative responses will occur. Firstly, some participants will choose to update both nations, but to update Y more. This is commensurate with the 'double updating' error observed within the polarisation literature. Secondly, some participants will choose to update Y only. This is based upon previous research demonstrating many individuals' reluctance to divide up an indivisible event based upon subjective probability.

Secondly, this experiment aims to compare response types between four conditions to test the cognitive processes behind the two proposed errors. While the control condition asked participants nothing about the new observed explosion, the categorical condition asked participants to choose which nation was the most likely source and the two probabilistic conditions asked participants to represent the probability of each nation being the source on a single scale (condition three) and two separate scales for each nation (condition four). In regards to the probabilistic error, we hypothesized that the control condition would show a lower number of these errors than the other three conditions. This is because assessing which nation is the cause of the explosion is not necessary to solve the problem (both nations increase equally regardless of who is the more likely cause). If fewer participants therefore do this in the control condition (because they are not prompted to), there will be less opportunity for them to carry this interpretation through to the updating phase. In regards to the categorical error, which is theorised based on previous work to occur at the point the evidence is assessed (i.e. the evidence is represented as 'Y launched the missile' instead of probabilistically), we hypothesize that the categorical condition will show an increased number of categorical errors, while the two probabilistic conditions will show a decreased number, relative to the control.

Method

Participants

Two hundred and fifty-five participants (50.2% female, mean age = 37.9 [SD=11.7]) were recruited from Amazon MTurk and were compensated $6.50 per hour.

Design, Procedure & Materials

A between-subjects design was employed with four condition. All participants were presented with the background information including the six observations for each nation (named 'Oclar' and 'Trubia'). They were then asked to provide an estimate of each nation's proficiency with missiles in percentage form on a sliding scale (one for each nation).

All participants were then presented the information regarding the successful missile explosion of unknown source. Unlike the example above, participants were given prior statistics of 1/6 for X and 4/6 for Y as piloting showed that the use of 2/6 and 4/6 encouraged the unwarranted assumption that there were 6 total observations, 2 for X and 4 for Y. The only difference between the four conditions then occurred at this point. The control condition (N = 66) were asked no question about the evidence. The categorical condition (N = 72) were asked to choose which of the two nations was most likely to have launched the missile. Participants in the first probability condition (N = 59) were required to represent their belief of which nation launched the missile on a single sliding scale ranging from -5 ('Definitely X) through zero ('Equal Chance') to +5 ('Definitely Y). The second probability condition (N = 58) required participants to indicate on separate sliding scales for each nation the percentage chance that they launched the missile.

All participants were then asked to indicate whether they would like to change their proficiency estimate for either X or Y, based on this new information. This response was recorded via two 7-point likert scales, one for each nation (e.g. Increase A LOT X's proficiency; Increase SOME X's proficiency; Increase A LITTLE X's proficiency; Make NO CHANGE to X's Proficiency; etc). This method was chosen rather than including another percentage sliding scale for several reasons. Firstly, we wanted to capture intuitive responses, rather than encouraging a numerical approach. Secondly, many previous experiments have had to set an arbitrary amount of change which 'counts' (e.g. Krynski & Tenenbaum, 2007) and this issue is further compounded by the fact that a digital slider is somewhat 'noisy' and so participants can struggle to get it to exactly the point they wish. Asking directly whether the participant wish to change their response avoids these issues. After every response participants were asked to provide their reasoning in an open-ended text box.

Piloting uncovered a third, non-theorised error and prompted an additional change to the experiment design. A substantial portion of participants chose to make no change to either nation. To understand this response further, participants who provided this were asked on the next page to endorse one of three statements to explain their reasoning: (1) An observation cannot change a fixed propensity; (2) One observation has a negligible effect; (3) We do not know who launched the missile.

The materials for this study and raw data have been made available at https://osf.io/4qanj/

Results

Overall, in the control condition, 13.89% of participants chose to increase the success propensity of both nations equally, which is, at least qualitatively, the normative response. However, 18.06% chose to make no change to either estimate *and* endorsed the 'We do not know who launched the missile' option (20.83% overall made no change). Furthermore, 12.5% increased their estimate of both nations, while increasing Y more (theorised to indicate the double update error). Finally, the dominant response at 34.72% was to increase the estimate of Y only, choosing 'no change' for X (theorised to indicate the categorical error).

To determine the impact of the three condition types (1=Categorical, 2=Control, 3=Combined Probabilistic) on those main response types, three binary logistic regressions were run. No significant effect was seen on the double update error (Wald $\chi^2(2)$ = 2.345, p=.310), the categorical error (Wald $\chi^2(2)$ = 1.126, p=.570), or the no change response (Wald $\chi^2(2)$ = 3.651, p=.161).

Regarding the assessment of the who launched the missile, all but one participant in both probabilistic conditions provided a 'probabilistic' representation of the chance of it being X / Y. By this it is meant that none represented the situation as being 'Certainly Y' or '100% Y / 0% X'. In fact the mean ratings in condition four (the only condition where this assessment was possible) were very close to the normative answer for this question (77.7% for Y (normative 80%) and 22.1% for X (normative 20%)). In condition one (the categorical condition), 87.88% of individuals chose Y as the nation most likely to have launched the missile. Within condition three, a linear regression was run to test the relationship between participants' ratings on the uni-dimensional scale representing the probability that each nation launched the missile and later committing the categorical error (B=.780, t = 1.727, p=0.90, *two-tailed*). Within condition four, two linear regressions were performed to test the relationship between committing the categorical error and their estimate that Y launched the missile (B=6.890, t=1.824, p=0.73, *two-tailed*), and secondly, their estimate that X launched the missile (B=-10.223, t=2.274, p=.027, *two-tailed*). Individuals who committed the categorical error are represented as orange circles in the diagrams below (all other answers as grey circles): a tendency towards a more Y-biased representation of the situation can be seen for those committing the categorical error.

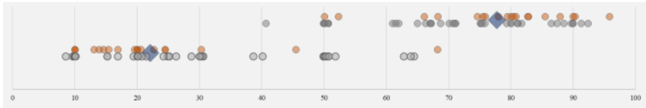

Figure 2. Participant ratings from condition four on separate scales for the probability that X (bottom row) and Y (top row) launched the missile. Orange dots represent those committing the categorical error, grey dots other answers and the blue diamond the mean.

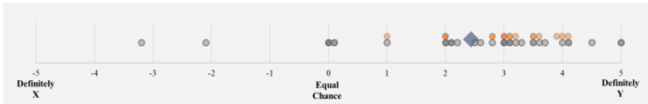

Figure 3. Participant ratings from condition three on a single scale for the chance that the missile was launched by X or Y.

The same set of regressions were run for the double update error and showed no significant relationship with condition three ratings (B=1.205, t=1.031, p=.307, *two-tailed*), condition four Y ratings (B=6.419, t=.1205, p=.233, *two-tailed*), or X ratings (B=5.549, t=.710, p=.481, *two-tailed*). The no change error also showed no relationship with condition three ratings (B=.372, t=.808, p=.422, *two-tailed*), condition four Y ratings (B=-5.076, t=-.1206, p=.233, *two-tailed*), or X ratings (B=-.129, t=-.025, p=.980, *two-tailed*).

An exploration of the think aloud data for the final posterior propensity update question (where participants are asked to justify their response) for 'double update' and 'categorical' errors was undertaken by the first author. Little could be discerned from the double update error data, with only three individuals clearly indicating that a greater probability of the missile being launched by Y was the reason for their answer. In terms of the categorical error however, 28 / 88 individuals committing this error showed either a clear belief, or a 'best guess' that Y was the one who launched the missile and six even stated that Y now has five out of seven successful attempts (and by implication that this was their reason for updating their propensity only). These responses can be seen in the public dataset.

Discussion

No difference was observed between the control condition and the categorical / probabilistic conditions in frequency of double update errors. This error is theorised to operate by individuals first determining the probabilities of who launched the missile based on the priors, then integrating this interpreted form of the evidence with the priors again to calculate the posterior propensities (i.e. using the priors twice). The control condition did not require participants to assess who launched the missile, suggesting that individuals in this condition spontaneously did this in their attempt to solve the problem, and did so in equal numbers to those who were prompted. This may suggest either that many participants genuinely saw this as the correct way to approach the problem, or perhaps that another process is responsible for the error. However, it tentatively appears that some individuals on this problem do commit a very similar 'double updating' error to that observed in the polarisation literature (e.g. Lord, Ross & Leper, 1979; Rabin & Schrag, 1999; Fryer, Harms & Jackson, 2015).

The categorical was theorised to be due to individuals being unwilling to carve up a single event based on subjective probability (e.g. Brase, Cosmides & Tooby, 1998) and so it was theorised that encouraging them to represent the evidence categorically would make this error more likely (and probabilistically, less). However, the categorical condition did not produce more categorical errors than the control condition and the probabilistic condition did not produce less. Furthermore, in the two probabilistic conditions, all but one individual represented the evidence probabilistically, with estimates for both X and Y being close to the normative answer. However, a large proportion of these individuals (roughly one third) who willingly provide a subjective probabilistic representation of the evidence then go on to act *as if* they have a categorical representation ('Y launched the missile') during updating (by increasing the propensity of Y only). Furthermore, the more certain a participant was that Y launched the missile, the more likely they were to make the categorical error. This finding does not fit neatly with the proposal derived from work by Brase, Cosmides and Tooby (1998) that individuals find parsing events in a non-holistic manner based on subjective belief difficult. It is not clear why, on this view, an individual would be more content to parse the event as a 70/30 split (i.e. to update both X and Y) without falling into categorical thinking ('Y launched the missile') but would be more likely to do so with a 90/10 split.

The think aloud data for the categorical error indicated many individuals stating a belief that Y launched the missile but was unable to discern if individuals believed this with certainty or were just making a 'best guess'. Many participants making this error however used non-certain terms such as 'think', 'believe', 'assume', 'likely' (that Y launched the missile) etc providing further tentative evidence that individuals may be using some form of rounding up heuristic. This fits with a 1973 theory by Gettys, Kelly and Peterson who theorised that in a multistage inference (where the output from step one (who launched the missile?) is used in step two (what are the new propensities?)) individuals will often round up the value from step one (or, use their 'best guess') before its use in step two. No theorised cognitive process (beyond presumed computational simplicity) was provided by the authors.

It should be noted that the correct solution of this problem is in fact not a multistage inference, as the interpretation of the evidence is not needed. The use of the interpreted form of the evidence is the theorised process leading to the double update error. It appears therefore that the categorical error builds upon and exaggerates this error, by, after interpreting the evidence, rounding it up, and using this rounded form in the final inference. Given this connection between these two errors, an intervention to reduce categorical errors directly by e.g. discouraging rounding may see a corresponding increase in double update errors.

In terms of the 'no change' error, the overwhelming endorsement of the 'We don't know who launched the missile' statement provides some insight into why this response occurs, but further research is needed to determine why these individuals do not instead choose the normative response of updating both nations' propensities (indeed 'We don't know who launched the missile' could also be considered a rationale for the normative response).

Both the double update and categorical errors unjustifiably increase one's prior belief based on entirely ambiguous evidence, leading to confirmation bias, or polarisation. Greater understanding of the cognitive processes underlying them should therefore be an important avenue for future study.

Acknowledgements

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), under Contract [2017-16122000003]. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. Funding was also provided by the ERC project ERC-2013-AdG339182-BAYES_KNOWLEDGE and the Leverhulme Trust project RPG-2016-118 CAUSAL-DYNAMICS.

References

Birnbaum, M. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *The American Journal of Psychology, 96(1),* 85–94.

Brase, G., Cosmides, L., & Tooby, J. (1998). Individuation, counting, and statistical inference: The role of frequency and whole-object representations in judgment under uncertainty. *Journal of Experimental Psychology-General, 127(1),* 3–21.

Brase, G. (2008). Frequency interpretation of ambiguous statistical information facilitates Bayesian reasoning. *Psychonomic Bulletin & Review, 15(2),* 284–289.

Casscells, W., Schoenberger, A., & Graboys, T. (1978). Interpretation by physicians of clinical laboratory results. *The New England Journal of Medicine, 299(18),* 999–1001.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58,* 1–73.

Evans, J., Handley, S., Perham, N., Over, D., & Thompson, V. (2000). Frequency versus probability formats in statistical word problems. *Cognition, 77,* 197–213.

Fenton, N., & Neil, M. (2012). *Risk assessment and decision analysis with Bayesian networks.* CRC Press.

Fryer, R., Harms, P., and Jackson, M., (2016). *Updating Beliefs When Evidence is Open to Interpretation: Implications for Bias and Polarization.* Available at SSRN: https://ssrn.com/abstract=2263504

Gettys, C. F., Kelly, C., & Peterson, C. R. (1973). The best guess hypothesis in multistage inference. Organizational Behavior and Human Performance, 10(3), 364–373.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important (and vice versa) in G. Wright & P. Ayton (Eds.). (1994). *Subjective probability* (pp. 129-161). Chichester: Wiley

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102(4),* 684–704.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. Psychological review, 98(4), 506.

Girotto, V., & Gonzalez, M. (2001). Solving probabilistic and statistical problems: A matter of information structure and question form. *Cognition, 78(3),* 247–276.

Hahn, U., & Harris, A. J. (2014). What does it mean to be biased: Motivated reasoning and rationality. *The psychology of learning and motivation, 61,* 41-102.

Kahneman, D., & Tversky, A. (1972). On prediction and judgment. *Oregon Research Institute Bulletin, 12(4),* 10.

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

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(11),* 2098.

Macchi, L. (2000). Partitive Formulation of Information in Probabilistic Problems: Beyond Heuristics and Frequency Format Explanations. *Organizational Behavior and Human Decision Processes, 82(2),* 217–236.

Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology, 2(2),* 175.

Rabin, M., & Schrag, J. (1999). First impressions matter: A model of confirmatory bias. *The Quarterly Journal of Economics, 114(1),* 37-82.

Sloman, S., Over, D., Slovak, L., & Stibel, J. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes, 91(2),* 296–309.

Welsh, M., & Navarro, D. (2012). Seeing is believing: Priors, trust, and base rate neglect. *Organizational Behavior and Human Decision Processes, 119(1),* 1–14.