Support for a Deliberative Failure Account of Base-Rate Neglect: Prompting Deliberation Increases Base-Rate use

Natalie A. Obrecht (obrechtn@wpunj.edu)
William Paterson University, Department of Psychology, 300 Pompton Rd.
Wayne, NJ 07470 USA

Dana L. Chesney (dlchesney@gmail.com)
The Ohio State University, Department of Psychology, 1827 Neil Ave. Mall
Columbus, OH 43210 USA

Abstract

People often base judgments on stereotypes, even when contradictory base-rate information is provided. It has been suggested this occurs because people fail to engage or complete deliberative reasoning needed to process numerical base-rate information, and instead rely on intuitive reasoning. However, recent research indicates people have some access to this base-rate information even when they make stereotype judgments. Here we tested several hypotheses regarding these phenomena: A) People may believe stereotype information is more diagnostic; B) People may find stereotype information more salient; C) People have some intuitive access to base-rate information, but must engage in deliberation to make full use of it. Aligning with account C, and counter to account A, we found inducing deliberation generally increased the use of base-rate information. Counter to account B, inducing deliberation about stereotype information decreased use of stereotype information. Additionally, more numerate participants were more likely to make use of base-rate information.

Keywords: base rates; judgment; reasoning; inductive reasoning; dual process theory; mathematical cognition; individual differences

Introduction

It is widely accepted that humans have two systems of reasoning, one that is intuitive, and one that is deliberative (e.g. Evans, 2003; Kahneman, 2002; Stanovich, 1999; Sloman, 1996, 2014). The intuitive (Type 1) system automatically and quickly provides impressions about the world. The deliberative (Type 2) system instead requires one to engage in effortful, symbolic reasoning to reach conclusions. Intuitive thinking is less effortful and time consuming than deliberative processing, but sometimes at the cost of accuracy. Numerous studies have demonstrated intuitive thinking can lead to judgment biases: people providing intuitive answers often seem to ignore relevant information (e.g. Kahneman & Tversky, 1973; Tversky & Kahneman, 1983). For example, consider a problem from De Neys and Glumicic (2008):

> "In a study 1000 people were tested. Among the participants there were 997 nurses and 3 doctors. Paul is a randomly chosen participant of this study. Paul is 34 years old. He lives in a beautiful home in a posh suburb. He is well spoken and very interested in politics. He invests a lot of time in his career. Which is more likely? A) Paul is a nurse. B) Paul is a doctor."

Intuitively, Paul sounds like most people's stereotype of a doctor, while the base-rate information (997 nurses vs. 3 doctors) instead suggests Paul is a nurse. Most people make judgments in line with the stereotype information; that is, they judge that Paul is more likely to be a doctor.

Why do people fail to incorporate normatively relevant base-rates into their judgments? Many theorists attribute such reasoning errors to failures to initiate (Kahneman, 2002; Kahneman & Frederick, 2005), or carry out (Sloman, 1996, 2014) deliberative Type 2 processes. In such accounts, base-rates are not appreciated at an intuitive level, but rather can only impact judgments when deliberative Type 2 reasoning is engaged.

However, recent findings seem to contradict this account. People take longer to give their responses (De Neys & Glumicic, 2008) and are less confident in their answers (De Neys, Cromheeke, & Osman, 2011) when considering scenarios for which stereotype and base-rates are incongruent, (supporting different judgments, as in the example above) compared to scenarios in which they are congruent (supporting the same judgment, e.g. the above scenario if the sample instead had 997 doctors and 3 nurses). These findings conflict with the traditional explanation that base-rate neglect occurs because people do not process base-rates. Base-rate/stereotype congruency would not affect decision speed or confidence in stereotype responses unless participants who choose stereotype responses had some intuition regarding the base-rate information or did in fact engage in deliberation to make their choice. Thus, stereotype judgments are not simply due to a lack of Type 2 deliberation forestalling all access to base-rate information. Rather, they may occur either because (a) participants engage in deliberation but think stereotype information is more diagnostic than the base-rate information or (b) participants have intuitions regarding base-rate information, but their intuitions regarding stereotype information are simply more salient (Pennycook, Trippas, Handley, & Thompson, 2013). However, it may be that while the original deliberative failure account was incorrect regarding the specific mechanism – apparently people do seem to have some intuitive access to base-rate information (De Neys, 2013) ─ it is correct to the extent that the reason people give less weight to base-rate information is because they fail to engage in deliberative Type 2 reasoning.

To differentiate between these accounts we had participants complete problems like the Paul example shown above. However, we manipulated whether participants were asked to evaluate arguments that support the stereotype answer and/or arguments that support the base rate answer before rendering their judgments regarding the group that an individual belongs to (e.g., Paul's profession). Evaluating these arguments should prompt participants to engage in Type 2 reasoning about that information and also increase its salience. As we outline below, different hypotheses make different predictions on how prompting such deliberation should impact participants' judgments.

Diagnosticity Hypothesis

If it is the case that people have intuitions about base-rate and stereotype information, but give the stereotype information more weight because they think it is more diagnostic, then prompting participants to reason about base-rates should not affect their rate of selecting the base-rate answer nor their confidence in their choice. Moreover, participants should rate arguments in favor of the stereotype answer as stronger than arguments in favor of the base-rate answer.

Salience Hypothesis

If it is the case that people have intuitions about base-rate and stereotype information, but give the stereotype information more weight because it is more salient, then increasing information salience should increase the use of that information. Thus, prompting participants to reason about base-rates should increase their rate of selecting, and their confidence in, the base-rate answer. Prompting participants to reason about stereotypes should increase their rate of selecting, and their confidence in, the stereotype answer. Base-rate and stereotype argument ratings should be similar.

Deliberative Failure Hypothesis

If it is the case that people select the stereotype rather than the base-rate answer on incongruent problems because they fail to carry out deliberative reasoning, then prompting deliberative reasoning about base-rates should cause participants to select the base-rate answer more often on such problems. This is because evaluating arguments should prompt deliberative thinking. Also, prompting this reasoning should cause participants who still give stereotype responses to incongruent problems to be less confident in their answers, as such reasoning would highlight the conflict between the stereotype and base-rate information. Participant's base-rate argument ratings should also be increased

Numeracy

Individuals who are higher in numerical ability are less susceptible to some judgment biases (Peters, Vastfjall, Slovic, Mertz, Mazzocco, & Dickert, 2006) and also are more likely to engage in deliberative thinking (Pennycook, Cheyne, Barr, Koehler, & Fugelsang, 2013). Therefore we included a numeracy measure to test whether the argument presentation manipulations would differentially affect participants as a function of numeracy.

Method

Participants

Undergraduate students at William Paterson University and The Ohio State University were recruited to participate in an online study for partial course credit. Recruitment goals of ~150 students per University were chosen to yield at least 20 participants per cell in the between subjects design, after accounting for a high anticipated incompletion rate. Of these, 137 of 146 recruited OSU students and 126 of 152 recruited WPU students completed the surveys for a final sample size of N=263.

Design

Participants completed inference tasks like the example above where they decided whether individuals belong to one group or another. Each participant read twelve scenarios in random order. In six scenarios the stereotype and base-rate information were congruent (e.g. Paul sounds like a doctor, and most of those sampled were doctors). In the other six, stereotype and base rate information were incongruent (e.g. Paul sounds like a doctor, but most of those sampled were nurses). Stimuli were taken from De Ney & Glumicic (2008), with a few minor updates to reflect current culture (e.g. Britney Spears was changed to Justin Bieber). Critically, we manipulated whether participants were given and asked to evaluate arguments that supported using the base-rate information and/or the stereotype information. Base-rate argument (evaluated vs. omitted) and stereotype argument (evaluated vs. omitted) were crossed between subjects yielding 4 cells in a 2x2 between-subjects design.

In order to hold the argument structure as constant as possible, the stereotype arguments simply repeated back the information that was given in the scenario. The base-rate arguments followed the format shown below, with slight variations to match the scenario. Note that for congruent problems the two arguments supported the same conclusion, while for incongruent problems they supported the opposite choice.

Example stereotype argument "Sam argues that Paul is very likely to be a doctor because Paul is 34 years old, lives in a beautiful home in a posh suburb, is well spoken and very interested in politics. Also, he invests a lot of time in his career."

Example base-rate argument "Sal argues that Paul is very likely to be a nurse because 997 out of the 1000 people in the sample are nurses; thus, the probability of randomly selecting a nurse is much higher than the probability of selecting a doctor."

For each scenario, where applicable, participants first rated the argument(s) they saw on a 1-7 scale from Extremely Strong to Extremely Weak (recall that which, if any, arguments were presented varied between subjects: neither, stereotype only, base-rate only, or both). Participants then judged which group the described individual belonged to using a 6 point scale; this scale allowed us to simultaneously obtain choice and confidence data.

Example group judgment "Do you think Paul is a nurse or a doctor? Please select one of the following:". Participants rated their confidence that Paul was a nurse or a doctor on a 6 point scale (1-very confident Paul is a nurse to 6-very confident that Paul is a doctor).

Counterbalancing For participants who were given both base-rate and stereotype arguments, the order of these arguments was counterbalanced between subjects, with half always evaluating the base-rate argument first, and half always evaluating the stereotype argument first. We also counterbalanced scenario congruency by alternated which six of the scenarios were congruent or incongruent between subjects. In the example above, Group A, would read the incongruent version of the scenario ("997 nurses and 3 doctors") while Group B would read the congruent version of the scenario ("997 doctors and 3 nurses"). The stereotypical category name was listed first in half of the scenarios.

Numeracy Participants completed an 8-item Objective Numeracy Scale (ONS) (Weller, Dieckmann, Tusler, Mertz, Burns, & Peters, 2013) and other individual difference measures not discussed here.

Procedure Participants completed the experiment online. Participants were randomly assigned to one of the 10 between subjects conditions: 2 (story/congruency pairing) x 2 (base-rate argument) x 2 (stereotype argument) + 2 (argument order among those evaluating both). All participants responded to 12 scenarios (6 congruent, 6 incongruent). For each scenario, they first evaluated their assigned argument(s) (neither, stereotype, base-rate, or both), then judged group membership. Finally, they completed the numeracy measure.

Results

Coding

The 1-6 scale group responses were recoded to create 2 separate choice variables indicating the proportion of stereotype responses for congruent and incongruent conditions, and 4 separate confidence variables indicating confidence in stereotype and non-stereotype answers in the congruent and incongruent conditions (6 variables total). Choice was coded as 1 or 0, with 1 indicating an answer that was consistent with the stereotype (e.g. 0 = nurse, 1 = doctor). On incongruent trials this indicated base-rate neglect, while on congruent trials this was the normative response, matching both the stereotype and base-rate information. Confidence was coded on a 1 to 3 scale (e.g., 1: slightly; 2: moderately; 3: very). These values were then averaged to yield the 6 variables for each participant. Argument ratings were coded by separately averaging the 6 base-rate and/or 6 stereotype argument evaluations in each congruency condition.

Preference For Stereotype Over Base-Rates

Participants typically gave the stereotype answer (M = .77, SE = .02), but chose this response more frequently in congruent compared to incongruent scenarios (congruent: M = .86, SE = .01; incongruent: M = .68, SE = .02, paired samples $t(262) = 7.87$, $p < .001$, d = .49, r = -.01). This same effect held when examining only the subset of participants who did not evaluate any arguments (congruent: M = .94, SE = .01; incongruent: M = .79, SE = .04, $t(54)$ = 3.67, $p < .001$, d = .49, r = -.17).

Deliberation Increased Base-Rate Use

We used a mixed model ANOVA with congruency (congruent, incongruent) as a within subjects factor, and base-rate argument (evaluated, omitted) and stereotype argument (evaluated, omitted) as between subjects factors predicting the proportion of stereotype responses. As expected, there was a main effect of congruency such that stereotype responses were more common in the congruent conditions, as noted above ($F(1, 259) = 49.2$, $p < .001$, η^2_p = .160). As predicted by the deliberative failure account and salience account, but inconsistent with the diagnosticity account, participants who evaluated base-rate arguments were less likely to select the stereotype response than those who did not ($F(1, 259) = 10.4$, $p = .001$, η^2_p = .039; base-rate argument omitted: M = .82 SE = .02; base-rate argument evaluated: M = .74, SE = .02). Also there was a main effect of stereotype arguments ($F(1, 259) = 7.1$, $p = .008$, η^2_p = .027). Contrary to the salience account, participants who evaluated the stereotype argument were actually less likely to go on to select the stereotype response. This was qualified by an interaction between base-rate argument and congruency ($F(1, 259) = 13.3$, $p < .001$, η^2_p = .049) such that the effect of evaluating base-rate arguments was only found in the incongruent scenarios – scenarios where base-rates suggested an answer different from the stereotype (see Figure 1). No other effects were found (all ps > .1)

Base-Rate Arguments Rated Stronger Than Stereotype Arguments Use

Contrary to the diagnosticity hypothesis, participants rated the base-rate arguments to be stronger than the stereotype arguments. This effect was statistically significant among participants who viewed both arguments (*N* = 104, base-rate evaluation: *M* = 5.26 *SE* = .12; stereotype evaluation: *M* = 4.26, SE = .13; paired samples $t(103) = -5.78$, $p < .001$, $d = .57$, $r = .044$) and marginally significant between groups that evaluated only one argument type (base-rate evaluation: *M* = 5.16, *SE* = .16, *N* = 52; stereotype evaluation: *M* = 4.79, *SE* = .12, *N* = 52; $t(102) = 1.86$, $p = .065$, $d = .37$).

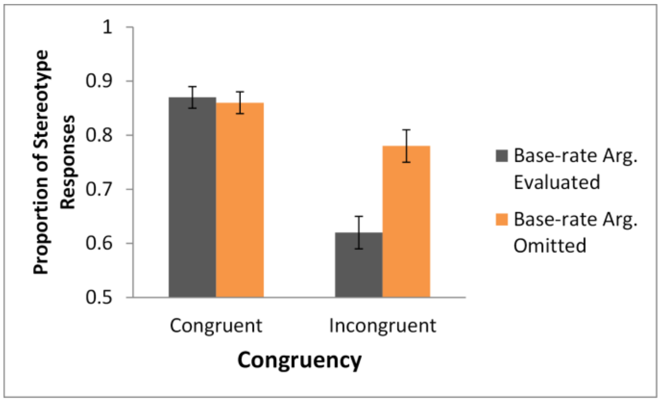

Figure 1: Proportion of stereotype responses as a function of congruency and base-rate argument evaluation. Standard errors bars are shown.

Evaluating Arguments Inconsistent With Later Judgments Reduced Choice Confidence Use

We tested whether the arguments people evaluated affected their confidence in their subsequent choices. We first looked at participants' confidence in their base-rate choices on incongruent scenarios via a 2 (base-rate argument: evaluated, omitted) × 2 (stereotype argument: evaluated, omitted) between-subjects ANOVA. Participants who evaluated the stereotype argument were less confident in their subsequent base-rate choices compared to participants for whom the stereotype argument was omitted ($F(1, 186) = 3.93$, $p = .049$, $\eta^2_p = .021$; stereotype evaluated: *M* = 1.48, *SE* = .06, stereotype omitted: *M* = 1.67, SE = .07). No other effects were uncovered (all *p*s > .1).

In a similar analysis we tested whether base-rate argument and stereotype argument predicted confidence in stereotype choices on these incongruent conditions. Participants who evaluated base-rate arguments were less confident in their subsequent stereotype choices than participants for whom this argument was omitted ($F(1, 242) = 12.33$, $p < .001$, $\eta^2_p = .048$; base-rate evaluated: *M* = 1.73 *SE* = .05; base-rate omitted: *M* = 1.99 *SE* = .05). There were no other significant effects (all *p*s > .1). These results show that people's confidence is reduced when they have evaluated arguments inconsistent with their decision. However, their confidence is not increased by evaluating arguments that are consistent with their choice. Two parallel analyses for the congruent trials found only that evaluating the stereotype argument reduced confidence in the 'correct' stereotype/base-rate answer ($F(1, 256) = 8.87$, $p = .003$, $\eta^2_p = .033$; stereotype evaluated: *M* = 1.99, *SE* = .05; stereo-type omitted: *M* = 2.21, *SE* = .06; all other *p*s > .1).

Numeracy Predicted Judgments Use

Participants' numeracy scores equaled the total number of questions they answered correctly on the 8 item numeracy assessment. Non-responses were scored as incorrect. Scores ranged from 0 to 8, and were normally distributed around 4.11 (*SE* = .11) with no significant skew.

Choice In order to examine whether the effects of arguments differ for people at different numeracy levels we conducted a regression analysis with stereotype argument (evaluated = 1, omitted = 0), base-rate argument (evaluated = 1, omitted = 0), and numeracy score and their interactions as predictors of the proportion of stereotype responses. All variables and their interactions were mean centered. Analyses were run separately for congruent and incongruent scenarios.

For the incongruent scenarios, this analysis showed base-rate argument ($\beta = -.27$, $p < .001$) and the interaction among numeracy, base-rate argument, and stereotype argument ($\beta = -.15$ $p = .014$) significantly predicted choices. No other effects were significant (all *p*s > .06). Generally, evaluating the base-rate argument decreased the chances that participants chose the stereotype answer in incongruent problems. However, for more numerate participants, this effect is only seen when the stereotype argument was also evaluated, while for less numerate participants the effect of the base-rate argument was more pronounced when the stereotype argument was omitted. It could be that more numerate participants consider base-rate arguments without prompting, but that the importance of base-rate information is further appreciated when contrasted with stereotype arguments. In contrast, less numerate participants may benefit from being directed to think about base-rate information, but this may be overwhelmed by stereotype information.

The regression looking at the congruent trials showed an effect of numeracy ($\beta = .31$, $p < .001$) such that participants higher in numeracy were more likely to pick the choice that reflected both the stereotype and base-rate answer compared to less numerate participants. Also, there was an effect of stereotype argument ($\beta = -.17$, $p = .003$) and an interaction between stereotype and base-rate argument ($\beta = .16$, $p = .006$) such that participants were less likely to select the stereotype/base-rate answer when the stereotype argument was evaluated, especially when the base-rate argument was omitted (no arguments: *M* = .94, *SE* = .01; stereotype argument: *M* = .78, *SE* = .04; base-rate argument: *M* = .88, *SE* = .03; both arguments: *M* = .86, *SE* = .02). Evaluating the stereotype argument actually decreases people's choice of that intuitive response even though in congruent problems base-rates also support the stereotype answer. This perhaps suggests a belief that relying on stereotype information is 'bad'.

Argument ratings A regression analysis with numeracy as the predictor and base-rate argument rating as the outcome showed that participants higher in numerical ability rated the base-rate argument as stronger than did those lower in numerical ability ($\beta = -.312$, $p < .001$). No such relationship was found when the regression was run on the stereotype ratings.

Discussion

Our results are consistent with the Deliberative Failure hypothesis, that people give more weight to stereotype information than base-rate information in part because they do not spontaneously engage in deliberative Type 2 reasoning. Participants who were prompted to deliberate about base-rate information (i.e. by evaluating arguments about base-rate information) both chose the base-rate answer at a greater rate than those who did not, and were less confident in their response on the occasions they chose the non-base-rate (stereotype) answer. This cannot be attributed to argument evaluation simply making the base-rate answers more salient, as evaluating the stereotype argument also increased the chances of selecting the base-rate answer. It appears that prompting people to engage in deliberation, regardless of which answer is made salient, increases their use of base-rate information. These findings stand in contrast to the Diagnosticity hypothesis that people spontaneously process and use base-rate information, but believe stereotype information is more diagnostic of group membership. If that were the case, then inducing deliberation should not have affected participants' choices.

Moreover, participants explicitly rated base-rate arguments as stronger than the stereotype arguments. Indeed, it appears that reasoning about stereotype information caused participants to give stereotypes less weight. Thus we believe that while people do at some basic level process base-rates (De Neys & Glumicic, 2008; Pennycook et al., 2013), they do not appear to fully appreciate their value without deliberation. These data suggest that interventions to increase deliberation may increase people's use of statistical information such as base-rates.

Individuals' numeracy also seemed to influence their choices. When base-rates and stereotypes were consistent, participants higher in numeracy selected the base-rate response more frequently than those lower in numeracy. On incongruent trials, more numerate participants picked the base-rate answer more often when both the stereotype and base-rate arguments were given. Less numerate participants were simply more likely to select the base-rate response when they evaluated the base-rate argument, regardless of whether the stereotype argument was given. More numerate participants also rated the base-rate argument to be stronger than less numerate participants. These results are consistent with previous research showing that more numerate people tend to make greater use of numbers when making decisions (Peters et al., 2006) while less numerate people may require an intervention that promotes numerical use (Obrecht, 2010). It is currently unclear whether this is due to more numerate participants being more likely to deliberate spontaneously (Pennycook et al., 2013), or due to more numerate participants having stronger intuitions about numbers (Schley & Peters, 2014). Further research is needed to address this issue.

Although these results show support for the deliberative failure account, we cannot conclude this is the only factor that accounts for failure to consider base-rates and other normatively relevant data. People's choices generally still favored the stereotype answer when it was pitted against base-rate data, even when Type 2 thinking was prompted by the argument evaluations. We suspect that people appreciate the information that base-rates provide and see it as formally stronger evidence than the stereotype information. However, it may be that the value people give the stereotype information is not captured by their ratings of argument strength. Following Thompson's (2009) view, people experience a System 1 based "feeling of rightness" that is stronger when an answer more easily comes to mind, like the answer that Paul must be a doctor, and is not counteracted by the System 2 processes that indicate that base-rate information may provide superior information.

We also note that one may have mathematically sound reasons for, apparently, giving more weight to stereotype information than base-rates in some cases. For example, individuals may implicitly assign a very low probability to the chances of Paul being a nurse, given his description, compared to the probability of him being a doctor. Suppose one feels only 1 in 1000 nurses would match Paul's description, while 999 of 1000 doctors would match the description. In this case, even given the "3 doctors to 997 nurses" base-rate, the odds that Paul is a doctor rather than a nurse are about 3 to 1. This could explain why participants persist in choosing the stereotype responses despite rating the base-rate argument as stronger. Indeed, past research has shown that people do not merely rely on explicitly provided information, such as base-rates, when making judgments, but also consider implied probabilistic information. For example, people give more weight to sample size when averaging explicitly provided sample means as the implied likelihood that those samples must have come from the same population increases (Chesney & Obrecht, 2011, 2012; Obrecht & Chesney, 2013). Although not typically considered normative, this represents some sophisticated reasoning in which information is evaluated in light of prior beliefs.

One limitation of this work is that we did not have measures to confirm that our argument manipulations indeed increased deliberation and salience. A follow up study is needed to confirm whether this was the case, or whether the effect of our manipulation might be attributed to some other causal pathway.

In sum, it appears that people do appreciate the importance of base-rates and are able to make better use of them when prompted to deliberate on the value of this information.

Acknowledgments

This research was partially supported by a Summer Research Stipend awarded to the first author by the Research Center for the Humanities and Social Sciences at William Paterson University and by NSF grant SES-1155924 funding the second author. We thank Stacey Delos Santos for her assistance with reviewing stimuli. We thank Ellen Peters for her support.

References

Chesney, D. L. & Obrecht, N.A. (2011). Adults are sensitive to variance when making likelihood judgments. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 3134-3139). Austin, TX: Cognitive Science Society.

Chesney, D. L., & Obrecht, N. A. (2012). Statistical judgments are influenced by the implied likelihood that samples represent the same population. M*emory & Cognition, 40,* 420-433. doi: 10.3758/s13421-011-0155-3

De Neys, W. (2013). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning, 20*, 169-187.

De Neys, W. & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition, 106*, 1248-1299.

De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PLoS ONE, 6*, e15954. doi:10.1371/journal.pone.0015954.

Evans, J. St. B. T. (2003). In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences, 7*, 454-459.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 513-525. doi: 10.1037/0096-1523.14.3.513

Kahneman, D. (2002). Maps of bounded rationality: A perspective on intuitive judgment and choice. In T. Frangsmyr (Ed.), Les Prix Nobel 2002 [Nobel Prizes 2002]. Stockholm, Sweden: Almquist & Wiksell International.

Kahneman, D. & Frederick, S. (2005). A model of heuristic judgment. In K.J. Holyoak & R.G. Morrison [Eds.] *The Cambridge Handbook of Thinking and Reasoning*. Cambridge University Press. 267-293.

Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251.

Obrecht, N. A. & Chesney, D. L. (2013). Sample representativeness affects whether judgments are influenced by base rate or sample size. *Acta Psychologica, 142,* 370-382. doi: 10.1016/j.actpsy.2013.01.012.

Obrecht, N. A. (2010). *Sample size weighting in probabilistic inference*. (Doctoral dissertation). Rutgers University, New Brunswick.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C.K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, 17, 408-414.

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2013). Cognitive style and religiosity: The role of conflict detection. *Memory & Cognition, 42*, 1-10.

Pennycook, G., Trippas, D., Handley, S. J., & Thompson, V. A. (2013) Base rates: Both neglected and intuitive. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 544-554.Schley, D. R. & Peters, E. (2014). Assessing "economic value": Symbolic number mappings predict risky and riskless valuations. *Psychological Science, 25*, 753-761.

Sloman, S. A. (1996). The empirical case for type systems of reasoning. *Psychological Bulletin, 19*, 3-22.

Sloman, S. A. (2014). Two systems of reasoning, an update. In Sherman, J., Gawronski, B., & Trope, Y. (Eds.). *Dual process theories of the social mind*. Guilford Press.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning.* Mahwah, NJ: Erlbaum.

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In J. Evans and K. Frankish (Eds.), *Two Minds: Dual Processes and Beyond.* Oxford University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293-315.

Weller, J., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch Analysis approach. *Journal of Behavioral Decision Making, 26*,198-212. doi:10.1002/bdm.1751