Tasks That Prime Deliberative Processes Boost Base Rate Use

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

Dana L. Chesney (dlchesney@gmail.com)
St. John's University, Department of Psychology
8000 Utopia Parkway
Jamaica, NY 11439 USA

Abstract

Obrecht and Chesney (2016) contend that deliberation supports greater base rate use. In line with this, they found that prompting deliberation by evaluating arguments about the usefulness of base rate and/or stereotype data increased subsequent use of base rates in judgment tasks. However, an alternative account of these results is that the intervention increased base rate use merely by increasing the salience of base rate information, rather than by increasing deliberation. Here we examine these accounts in two experiments. Experiment 1 showed that participants prompted to deliberate by evaluating arguments used base rates more in subsequent judgements, compared to participants who were merely reminded of relevant information. Experiment 2 showed that participants prompted to deliberate by completing math problems prior to the judgment task also increased their base rate use. Taken together, these results support the theory that tasks that prompt deliberative processes increase normative use of base rates.

Keywords: statistical inference; judgment; decision making; base rates; normative behavior; priming; numeracy

Deliberation and Number Use

Many important decisions that people make involve numbers. However, a large body of work demonstrates that people often fall short of normative standards when using numerical information to make decisions (Kahneman & Tversky, 1972, 1973; Reyna, Nelson, Han, & Dieckmann, 2009; Schley, & Peters, 2014; Smith, McArdle, & Willis, 2010; cf. Gigerenzer, 2015). A canonical example of this is the phenomenon of base rate neglect (Tversky & Kahneman, 1974): people tend to underweight the overall frequency of an occurrence in the population when considering its likelihood. This phenomenon manifests itself in several ways. For example, many people, even medical students and experienced doctors, will grossly misunderstand the probability that a patient with a positive test result has a disease when given information about the accuracy of the test and the base rate of that disease in the population (Casscells, Schoenberger, & Graboys, 1978; Eddy, 1982). Moreover, people tend to give more weight to diagnostic or stereotypical information than to base rate information when making judgements (Kahneman & Tversky, 1972). This is typified by people's responses when making category judgements in cases where stereotype and base rate information appear to conflict. For example, consider the following scenario:

"In a study 1000 people were tested. Among the participants there were 3 doctors and 997 nurses. 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." (De Neys & Glumicic, 2008).

After reading this scenario, people will typically claim that it is more likely that Paul is a doctor, rather than a nurse, despite the fact that nurses vastly out number doctors in the given population. Here people give more weight to Paul's description, which sounds like a stereotypical doctor, than to the numeric base rate information. This is considered base rate neglect: although one may think doctors are more likely than nurses to match Paul's description this discrepancy is insufficient to counter the large base rate differences (De Neys & Glumicic, 2008). Therefore, normatively, participants should respond that Paul is more likely to be a nurse than a doctor.

The traditional explanation for non-normative judgments, including base rate neglect, is rooted in *dual-process theory*. According to this theory, humans have two categories of thinking processes: intuitive processes and deliberative processes (e.g., Evans, 2003; Kahneman, 2002; Stanovich, 1999; Sloman, 1996). Intuitive processes are swift and automatic, quickly providing impressions about the world. Deliberative processes instead require one to engage in effortful, symbolic reasoning to reach conclusions. Intuitive processing is less effortful and time consuming than deliberative processing, but sometimes at the cost of accuracy. In numerically-based judgment and decision making (JDM) tasks, such as the one shown above, people may underutilize numerical information because they rely on intuitive, rather than deliberative processes: deliberative processes may be required to fully integrate some numerical data into judgments (Kahneman, 2002).

However, a growing body of research has shown that this explanation is incomplete. Recent findings suggest that lay people do in fact have intuitions about some numbers, but fail to adequately incorporate numerical information when making judgments and decisions. For example, Pennycook, Trippas, Handley, and Thompson (2014) asked participants to judge category membership based on both diagnostic stereotype information and base rate information, such as in the doctor-or-nurse example above. However, they contrasted conditions where the stereotype information and base rate information were *congruent*, supporting the same judgment, with conditions where the stereotype information and base rate information were *incongruent*, supporting different judgements. The doctor-or-nurse example above is an incongruent scenario, as the stereotype information supports the doctor judgement, while the base rate information supports the nurse judgment. However, if the base rate was instead “3 nurses and 997 doctors”, the information would be congruent. Like De Neys and Glumicic (2008), Pennycook et al. (2014) found that participants typically based their judgements on stereotypes. However, they also found that participants were more confident in their stereotype judgements when these judgements were also supported by congruent base rates, compared to when base rates instead were incongruent with the stereotypical information. Pennycook et al. (2014) claimed that this demonstrates that people have intuitions about the base rate information, although these intuitions were weaker than those about stereotype information.

Obrecht and Chesney (2016) explored an alternative take on this phenomenon, which sought to combine these findings with the more standard deliberative failure account. They proposed that, while people may have intuitions about base rates, these intuitions are relatively weak. Deliberation makes people more aware of the importance of base rate information. Thus base rates are underweighted in part because people fail to deliberate, instead relying on intuitive impressions. It follows that this underweighting of base rates could be at least partially remedied by prompting deliberation. To test this experimentally, they randomly assigned participants - after reading scenarios like the doctor-or-nurse example above – to evaluate arguments that supported the use of base rate and/or stereotypical data in decision making:

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.”

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.”

Critically, participants were asked to rate each argument on a 1 (Extremely Strong) to 7 (Extremely Weak scale). This was intended to prompt deliberation about the value of the base rate and/or stereotype data.

After reading the scenarios and rating their assigned arguments, participants were ask to judge category membership (e.g., doctor or nurse?). Participants who evaluated such arguments prior to making their judgments generally used base rates more than those who did not (e.g., they were more likely to think that Paul is a nurse). Even when participants only considered arguments in favor of using stereotypical descriptions they nevertheless used base rates more compared to a control group that didn’t evaluate any arguments (Obrecht & Chesney, 2016). Obrecht and Chesney (2016) contend this is because argument evaluation prompts participants to deliberate, supporting the notion that deliberative thinking indeed leads to greater base rate use. However, one limitation of this study is that there was no manipulation check to establish that reading the arguments in fact increased deliberation. Thus, an alternative explanation for the increased use of base rates is that evaluating the arguments merely refreshed participants’ recollection of the information from the scenarios that they could use to make their decision. Such a reminder might increase the salience of base rate information, and thus the salience of base rate intuitions, without deliberation. Thus, it might be the case that it is sufficient to remind participants about of the information, making base rates salient to intuitive processes. In this case, deliberation may not be necessary to increase base rate use.

A second issue left unresolved by Obrecht and Chesney (2016) is whether induced deliberation must be specifically about base rate information in order for it to increase base rate use. It is possible that simply prompting deliberative number use in general could increase people’s use of base rates. Prior studies have suggested that ‘numerical priming’, such as completing number-based tasks, improves performance on subsequent JDM tasks. For example, Hsee and Rottenstreich (2004) found that people tend to use numbers non-linearly when allocating resources: for example, a person who would give $5.00 to save one panda might only be willing to give $6.00 to save two, and $7.00 to save four, rather than increasing donations linearly with the number of pandas. Hsee and Rottenstreich (2004) attribute this to a lack of deliberative processing. They propose that people weight values more steeply, and in a more linear fashion, when engaged in deliberation using calculation, but instead weight values in a shallow, curvilinear manner, when engaged in intuitive processing. To test this, they randomly assigned participants to complete math problems before making their allocation judgements. Indeed, participants who completed this numerical prime were more linear in their allocations than those who did not. Given Hsee and Rottenstreich’s (2004) finding that numerical priming can increase deliberative calculation in a JDM task, it is possible that such priming may increase deliberation about numerical information more generally. By increasing general deliberation about numbers, such priming might increase base rate use.

Here, we address these questions in two experiments. In Experiment 1, we examine whether argument evaluation increases base rate use by prompting deliberation, or whether simply reviewing the scenario information is sufficient to increase base rate use. We note that these possibilities are not mutually exclusive. It is possible that reviewing base rate information increases base rate use and that inducing deliberation about base rates yields an additional increase. We test these hypotheses by comparing participants’ base rate use in conditions when they a) evaluate arguments, b) evaluate whether recalled information is accurate, and c) receive no intervention. We also examine response times in this study: longer response times when making category judgments may indicate deliberative processes are taking place. In Experiment 2 we examine whether numerical priming also increases base rate use. Participants were either assigned to complete math problems before (primed) or after (control) the group membership judgment task. To foreshadow our results, both experiments support the hypothesis that priming people to engage in deliberative processes, whether by evaluating arguments or completing math problems, leads to more normative use of base rates.

Experiment 1

Method

Participants 180 undergraduate students from William Paterson University and St. John's University participated for course credit or extra credit.

Design This study used a 3(intervention: argument, recall, or control) x 2(congruency: congruent, incongruent) mixed model design.

Procedure The study was run online using the Qualtrics survey platform. Undergraduates read 12 scenarios, including the doctor-or-nurse example shown above, describing a population of 1000 people. In each scenario most of the people belonged to one group (e.g., nurses) and a few belonged to the other group (e.g., doctors). Additionally, a description of an individual drawn from the population was given. The description was stereotypical of one of the two groups in the population. Scenarios were based on those used by De Neys & Glumicic (2008), with minor updates to reflect current culture (see Obrecht & Chesney, 2016).

Congruency conditions In half of the scenarios the base rate and stereotype information was incongruent, meaning they supported different group judgments. In the doctor-or-nurse example above, the base rates suggest Paul is a nurse, while the stereotypical description suggests he is a doctor. In the other half of the scenarios the base rates and stereotypes were congruent such that both base rates and stereotypes supported the same group judgment (e.g. a congruent version of the above scenario would instead have 3 nurses and 997 doctors).

Intervention conditions Participants were randomly assigned to one of three between-subjects conditions: argument, recall, or control. For the argument condition, we followed the same procedure as that used by Obrecht and Chesney (2016), except that scenarios and questions were presented on separate webpages to better allow for the collection of response time data. Participants first read the scenarios presented on one webpage, and then, on a subsequent webpage, rated the strength of the arguments, like those given in the introduction, that tout the usefulness of base rate and/or stereotype information.

Recall condition participants also began each trial by reading a scenario, but then on a subsequent webpage, were assigned to evaluate the accuracy of "recall statements" about base rates and/or stereotype data. For example:

Base rate recall: "Sal reads the above description and then later recalls that 996 out of the 1000 people in the sample were women. Please rate the accuracy of Sal's recall." (1-Extremely Strong to 7-Extremely Weak).

Stereotype recall: "Sam reads the above description and then later recalls that the person described is finishing a degree in engineering, and likes to go out cruising with friends while listening to loud music and drinking beer. Please rate the accuracy of Sam's recall."

We designed the length and structure of these recall statements to closely match the corresponding base rate and stereotype arguments.

In the control condition, participants simply read the scenarios. They did not evaluate any additional statements.

After completing the intervention (argument or recall statement evaluations; if applicable) participants made each category membership judgement on a separate webpage. They rated their confidence that the individual in the scenario belonged to one group or the other using a 6-point scale (e.g. 1: "Very confident that Paul is a doctor" [stereotype] to 6: "Very confident that Paul is a nurse" [base rate]). Higher numbers indicate greater base rate use. Response times were measured based on the amount of time participants viewed this page before responding.

Note: As in Obrecht and Chesney (2016), we also randomly varied between-subjects whether argument and recall participants evaluated only stereotype statements, only base-rate statements, or both. However, we collapsed across these conditions in the current analyses.

Analysis and Results

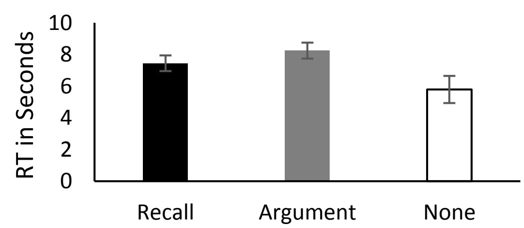

Figure 1: Marginal mean reaction times by statement condition in Experiment 1. Standard error bars are shown.

Response times As can be seen in Figure 1, participants who evaluated statements (argument or recall) took longer to judge which category an individual belonged to compared to the control group. A 2(congruent vs. incongruent) x 3 (argument, recall, none) repeated measures ANOVA uncovered a significant main effect of statement type on mean response time in seconds ($F(2, 175)=3.075$, $p=.049$, $\eta_p^2=.034$; Marginal means: Argument: $M=7.445$, $SE=.495$, Recall: $M=8.242$, $SE=.505$, None: $M=5.789$, $SE=.857$). No other effects were significant. Note: Due to experimenter error, one congruent recall question was dropped from this analysis for 14 participants who saw both stereotype and base rate statements.

Base rate use Since Obrecht and Chesney (2016) found the effect of argument evaluation was only relevant in the incongruent conditions (in which base rates and stereotypes conflicted), we focused on the 6 incongruent scenarios for the analysis of base rate use in judgements.

We found the mean of participants' group membership rating judgments for the six incongruent scenarios. Higher values indicate greater base rate use, while lower values indicate more reliance on stereotypical descriptions. As shown in Figure 2, there was a main effect of statement type such that participants in the argument condition used base rates more (*M*=2.631, *SE*=.100, *N*=79) than those in the recall (*M*=2.333, *SE*=.107, *N*=75, *t*(150.8)=2.033, *p*=.044, *Cohen's d*=.328) or control conditions (*M*=2.296, *SE*=.125, *N*=27; *t*(61.711)=2.088, *p*=.041, *Cohen's d*=.400)

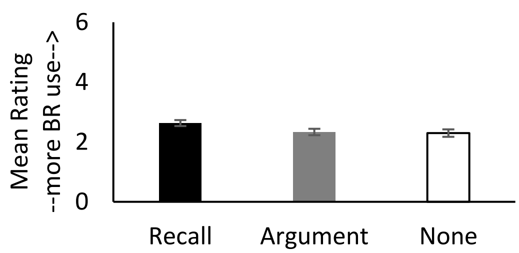

Figure 2: Mean rating by condition, for incongruent trials in Experiment 1. Standard error bars are shown.

Discussion

These data are consistent with Obrecht and Chesney's (2016) account that argument evaluation prompts deliberation, and this deliberation subsequently increases base rate use. Participants in the argument condition indeed used base rates more than controls. In contrast, no such benefit was seen for the recall participants, who were merely reminded of the scenarios' information, not prompted to deliberate. Additionally, participants' in the experimental condition took longer to make their group judgments than controls, which suggests they were more likely to engage in deliberative processes.

Experiment 2

Method

Participants 217 undergraduate students from William Paterson University and St. John's University participated for course credit.

Design This study used a 2(intervention: priming vs. control) x 2(congruency: congruent vs. incongruent) mixed model design.

Procedure As in Experiment 1, this experiment was conducted online using the Qualtrics platform. Participants were randomly assigned to complete a math assessment either before (primed) or after (control) the category membership judgment task.

The category membership judgement task consisted of the 12 scenarios (6 congruent and 6 incongruent) that were used in Experiment 1. However, the judgment questions were asked on the same webpage as the scenario presentation, as we were not intending to collect precise RT data.

The math assessment required calculation; participants completed the Objective Numeracy Scale (ONS; Weller, et al., 2013). This scale consists of 8 problems that required participants to use basic math skills, such as transforming numbers into different scales, for example:

- In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?
- If the chance of getting a disease is 10%, how many people would be expected to get the disease? ___ out of 1000

Analysis and Results

As can be seen in Figure 3, priming participants gave greater weight to base rate information than did control participants (2 (primed, control) x 2 (congruent vs. incongruent) repeated measures ANOVA: *F*(1,215)=4.256, *p*=.040, η_p^2=.019). This effect was specific to incongruent items: Primed participants made more use of base rates (*M*=2.453, *SE*=.099, *N*=99) than control participants (*M*=2.220, *SE*=.078, *N*=118) on the incongruent items (*t*(215)=1.943, *p*=.053, *Cohen's d*=.265), but on the congruent items, base rate use did not differ (*M* = 4.705, *SE* = .080, *N*=99 for primed vs. *M*=4.651, *SE*=.069, *N*=118 for control; *t*(215)=.516, *p*=.606, *Cohen's d*=.070). This is consistent with our prediction; it is only in the incongruent trials that intuitively processed stereotype information supports different judgements than deliberatively processed base rate information. Therefore, one would expect a shift toward deliberative processing to be seen most clearly in incongruent conditions. The expected effect of congruence was also seen (*F*(1,215)=590.384, *p*<.0005, η_p^2=.764; Marginal means: Congruent: *M*=4.678, *SE*=.053; Incongruent: *M*=2.337, *SE*=.060). The interaction was n.s.

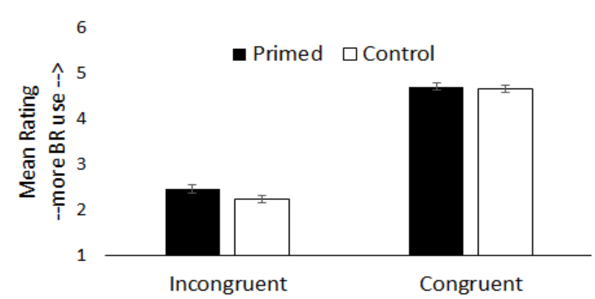

Figure 3: Mean rating by condition, for congruent and incongruent trials in Experiment 2. Standard error bars are shown.

We also examined participants' ONS scores (total correct out of 8 items). Primed participants, who completed the ONS task first, had significantly higher ONS scores (*M*=4.172, *SE*=.178, *N*=99) than control participants, who did the task second (M=3.686, SE=.148, N=118; $t(215)$=2.115, p=.036, *Cohen's d*=.288). This discrepancy cannot be explained by sex differences in ONS. Males scored slightly higher (M=4.28, SE=.197, N=69) than females (M=3.76, SE=.142, N=145) on the ONS overall ($t(212)$ = 2.096, p = .037), but made up a larger proportion of the control (41 of 118 male), compared to the primed (28 of 99 male), sample.

Discussion

As was the case in Experiment 1, these results support the hypothesis that prompting deliberation can increase base rate use. Moreover, the data provide initial experimental evidence that completing calculation tasks can prime subsequent use of numbers when making judgments. The results extend the findings of Hsee and Rottenstreich (2004) by showing that not only can calculation tasks improve the use of numbers on judgement tasks that require calculation (i.e., double the donation to help double the pandas), but they can also improve the use of number on tasks where numerical base rate information merely needs to be taken into consideration: identifying that having many more nurses than doctors in a population makes the random selection of a doctor less likely does not involve calculation per se.

An unanticipated result of this study was that scores on the calculation task were significantly lower for control participants than for priming participants. This difference does not appear to be explained by sex differences, as we had more males than females in the control group, yet males scored higher on the calculation task overall. Thus, scores on the calculation task may susceptible to order effects. This result is concerning, as many researchers use short calculation tasks, like the ONS task we used here, to gauge individual differences in people's underlying numerical ability (i.e., *numeracy*). It is therefore somewhat problematic that completing a 5-10 minute judgement task is sufficient to affect numeracy task performance. The effect size here may appear small, *Cohen's d* = .288, but as a point of comparison, this benefit is more than half of the benefit to scores resulting from the practice effect, when people repeat this task, *Cohen's d* = .555 (Chesney, Bjälkebring, & Peters, 2015). We recommend that researchers seeking to measure individual differences in numeracy carefully track the impact of design features to insure order effects do not inadvertently influence their results.

General Discussion

The results presented here indicate that interventions that prompt people to engage in deliberative processes lead to greater base rate use. This pattern of results was found both when participants were prompted to deliberate by evaluating the strength of arguments, and also when they were prompted to deliberate by completing math problems. These data are consistent with the theory that deliberative processes promote improved judgments. Thus, the current results bolster the theory that failure to deliberate underlies many non-normative decisions (Kahneman, & Tversky, 1972).

Limitations

While the data presented here generally support the idea that inducing deliberation increases base rate use, there are some caveats. The results of Experiment 1 were somewhat inconsistent. As predicted, participants who evaluated arguments showed more base rate use and also took longer to make category judgements than did participants in the control group. Further, participants in the recall condition performed similarly to the control group regarding base rate use. However, the recall group had longer RTs when making their category judgments. This is perplexing if longer RTs indeed indicate deliberation, and deliberation increases base rate use. If, as the RTs suggest, recall participants were deliberating more than control participants, then it follows that the recall group should have used base rates more than controls. Possible explanations for this inconsistency are that the RTs were not fully capturing deliberation, or that participants in the recall condition were not deliberating about the right data. That is, we cannot assume that longer RTs in the recall and argument conditions mean that participants were thinking about the same things. It could be that people in the recall condition took longer because they were working to recall the information provided in the scenario, rather than deliberating on the arguments regarding the usefulness of the data. This possibility is particularly of interest because it necessitates that we consider deliberation as a resource that can be directed along different pathways, rather than as a monolithic construct that is either present or not present. Further research is needed to investigate these possibilities.

A second issue that must be addressed is that in Experiment 2, participants in the control group, who completed the calculation task second, had lower numeracy scores on that task than participants in the priming group, who completed the calculation task first. We interpret this to indicate that the task order influenced numeracy scores. However, an alternative explanation would be that, by random chance, the primed participants were more numerate than the control participants. This is of particular concern, as it may affect the interpretation of our results. More numerate individuals tend to make more normative judgements in general (e.g., Peters, Västfjäll, Slovic, Mertz, Mazzocco, & Dickert, 2006), showing less susceptibility to some judgment biases compared to less numerate individuals. Indeed, more numerate individuals have been shown to make more use of base rates on the same category judgement tasks we use here (Obrecht & Chesney, 2016). Thus, an alternative explanation of our finding that the priming participants made more use of base rates is that - by chance - they happened to be more numerate than the controls. Therefore replication is needed to confirm that these results are not an artifact and that indeed, calculation priming does improve base rate use.

Future Directions

These results have the potential to aid the development of number-based decision aids. Our results demonstrate that short interventions can improve people's use of numerical information. This is particularly important as many critical life decisions involve numbers. Unfortunately, many people in the US have difficulty making use of this information (Kutner, Greenberg, Jin, Boyle, Hsu, & Dunleavy, 2007; Lipkus, Samsa, & Rimer, 2001). More research is needed regarding how people make use of numerical information when making decisions, and how to help improve these decisions. In the long term, this improved understanding could lead to the development of tailored interventions to benefit individuals' judgment and decision making skills, and thus, in turn, their life outcomes.

Additionally, our unanticipated finding that order influenced ONS scores may be useful for future research. It is implausible that the participants' underlying numerical ability (i.e., *trait numeracy*) changed over the course of the study. Therefore, the manipulation must have affected their tendency to use numerical information in the moment (*state numeracy*). Thus, some aspects of numeracy may be susceptible to brief interventions. This might be leveraged to test for causal effects of numeracy.

Conclusions

The results of the current study lend further support to the theory that people's lack of number use in judgements, at least in part, is due to a lack of deliberation. Moreover, our data show that short interventions can improve the use of numbers in judgement tasks. These results may be useful in the development of future decision aids.

Acknowledgments

We thank the RAs that helped us with this research: C. Brunswick, L. Guo, M. Mahida, R. Salim, & M. Shoaib.

References

Casscells, W., Schoenberger, A., & Grayboys, T. (1978). Interpretation by physicians of clinical laboratory results. New England Journal of Medicine, 299, 999–1000. doi:10.1056/NEJM197811022991808

Chesney, D. L., Bjälkebring, P., & Peters, E. (2015). How to estimate how well people estimate: Evaluating measures of individual differences in the approximate number system. *Attention, Perception, & Psychophysics, 77,* 2781–2802. doi:10.3758/s13414-015-0974-6

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. Cognition, 106, 1248-1299. doi:10.1016/j.cognition.2007.06.002

Eddy, D.M., (1982). Probabilistic reasoning in clinical medicine: problems and opportunities, in: Judgment under Uncertainty: Heuristics and Biases, D. Kahneman, P. Slovic, and A. Tversky, eds., Cambridge University Press, Cambridge, pp. 249–267.

Evans, J. St. B. T. (2003). In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences, 7*, 454-459. doi:10.1016/j.tics.2003.08.012

Gigerenzer, G. (2015). On the supposed evidence for libertarian paternalism. *Review of Philosophy and Psychology, 6*, 361-383

Hsee, C. K. & Rottenstreich, Y. (2004). Music, pandas and muggers: On the affective psychology of value. *Journal of Experimental Psychology: General, 133, 23*-30. doi:10.1037/0096-3445.133.1.23

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. & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430-454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-25l. doi:10.1037/h0034747

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making, 21,* 37–44. doi:10.1177/0272989X0102100105

Obrecht, N. A., & Chesney, D. L. (2016). Promoting deliberation increases base rate use in judgments. *Judgment and Decision Making, 1*, 1-6.

Pennycook, G., Trippas, D., Handley, S.J., & Thompson, V.A. (2014). Base rates: Both neglected and intuitive. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 40*, 544-54. doi:10.1037/a0034887

Peters, E., Västfjäll, D., Slovic, P., Mertz, C.K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science, 17*, 407-413. doi:10.1111/j.1467-9280.2006.01720.x

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin, 135,* 943–973. doi:10.1037/a0017327

Schley, D.R. & Peters, E. (2014). Assessing "economic value": Symbolic number mappings predict risky and riskless valuations. *Psychological Science, 25*, 753-761. doi:10.1177/0956797613515485

Sloman, S. A. (1996). The empirical case for type systems of reasoning. *Psychological Bulletin, 19*, 3-22. doi:10.1037/0033-2909.119.1.3

Smith, J. P., McArdle, J. J., & Willis, R. (2010). Financial decision making and cognition in a family context. *The Economic Journal, 120,* F363– F380. doi:10.1111/j.1468-0297.2010.02394.x

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

Tversky, A. & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. Science, 185, 1124-1131. doi:10.1126/science.185.4157.1124

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