

Prompting deliberation increases base-rate use

Natalie A. Obrecht*

Dana L. Chesney†

Abstract

People often base judgments on stereotypes, even when contradictory base-rate information is provided. In a sample of 438 students from two state universities, we tested several hypotheses regarding why people would prefer stereotype information over base-rates when making judgments: A) People believe stereotype information is more *diagnostic* than base-rate information, B) people find stereotype information more *salient* than base-rate information, or C) even though people have some intuitive access to base-rate information, they may need to engage in *deliberation* before they can make full use of it, and often fail to do so. In line with the deliberative failure account, and counter to the diagnosticity account, we found that inducing deliberation by having people evaluate statements supporting the use of base-rates increased the use of base-rate information. Moreover, counter to the salience and diagnosticity accounts, asking people to evaluate statements supporting the use of stereotypes *decreased* reliance on stereotype information. Additionally, more numerate subjects were more likely to make use of base-rate information.

Keywords: base-rates, judgment, reasoning, inductive reasoning, dual process theory, mathematical cognition, numeracy, individual differences

1 Introduction

Consider this 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 a stereotypical doctor, but the base-rate information (997/1000) suggests Paul is a nurse. Most people make judgments in line with the stereotype and say Paul is more likely to be a doctor. Theorists attribute such lack of base-rate use to failures to initiate (Kahneman, 2002; Kahneman & Frederick, 2005), or carry out (Sloman, 1996, 2014) deliberative processes. Such accounts posit people lack intuitions about base-rates, so base-rates can affect judgments only when deliberative reasoning is engaged.

Recent findings seem to contradict this account. People take longer to respond (De Neys & Glumicic, 2008) and are less confident in their answers (De Neys, Cromheeke & Os-

man, 2011) when stereotypes and base-rates support different judgments (*incongruent*, like the above example) then when they support the same judgment (*congruent*, e.g., if the example scenario instead had 997/1000 doctors). Base-rate/stereotype congruency would not affect stereotype responders’ speed or confidence unless they had some intuition regarding the base-rates or had engaged in deliberation. Thus, stereotype judgments seem not to result from a lack of deliberation forestalling access to base-rates. Rather, they may occur because subjects deliberate, but think stereotype information is more diagnostic than the base-rate information (see De Neys & Franssens, 2009, for a discussion about diagnosticity).

Alternatively, although subjects have intuitions regarding base-rates, their intuitions regarding stereotypes may be more salient (Pennycook, Trippas, Handley & Thompson, 2013). Consistent with this, there are recency effects on judgments such that base-rate data are weighted more when they are presented after stereotype data (Krosnick, Li & Lehman, 1990). Finally, while the original deliberative failure account is flawed – people do seem to have some intuitive access to base-rate information (De Neys, 2013) — it may be correct to the extent that one reason people give less weight to base-rates than stereotypes is because they fail to deliberate. Consistent with this deliberative failure account, people make more normative use of base-rates (Schwarz, Strack, Hilton & Naderer, 1991) and probabilistic data (Kogler & Kuhberger, 2007) when problems are framed as statistical in nature.

To differentiate among these accounts, we had subjects judge scenarios like the Paul example above. We manipulated whether subjects evaluated statements that highlighted

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 NSF grant SES-1155924 which supported the second author. We thank Stacey Delos Santos for her assistance with reviewing stimuli and Ellen Peters for her support.

Copyright: © 2016. The authors license this article under the terms of the Creative Commons Attribution 3.0 License.

*Department of Psychology, William Paterson University, 300 Pompton Road, Wayne, NJ 07470, USA. Email: obrecht@wpunj.edu.

†Department of Psychology, St. John’s University. Dr. Chesney was at The Ohio State University during data collection.

stereotype information and/or base-rate information before making their judgments (e.g., judging Paul's profession). Prompting subjects to evaluate the strength of these statements should lead them to deliberate about the value of the information (i.e., base-rate and/or stereotype), thus increasing their base-rate use (see Trouche, Sander & Mercier, 2014 for an example of how evaluating statements can increase correct responses in reasoning tasks). Additionally, we tested whether individuals may need the usefulness of base-rate information "spelled out" for them before they will make use of it, even when the information has been highlighted. Thus, we additionally manipulated the amount of "deliberative support" these statements provided: while some subjects read statements that merely reiterated (and thus highlighted) information in the original scenario, others read statements that also contained an explanation of why the information was useful. Finally, we evaluated subjects' numerical ability (*numeracy*).

1.1 Theoretical predictions

Different hypotheses yield different predictions on how prompting deliberation should affect judgments.

Diagnosticity. If people give stereotypes more weight because they think they are more diagnostic than base-rates, then prompting subjects to deliberate about base-rates should not affect judgments. Additionally, prompting subjects to deliberate about stereotypes should either not affect judgments or should increase stereotype use. Also, subjects should generally rate stereotype statements as stronger than base-rate statements.

Salience. If people give stereotype information more weight because it is more salient, then increasing base-rate or stereotype information's salience should increase its use. Thus, evaluating base-rate statements should increase subjects' base-rate judgments (e.g., "Paul is a nurse"); and rating stereotype statements should similarly affect stereotype judgments (e.g., "Paul is a doctor"), unless stereotype use is already at ceiling. There are no a-priori predictions regarding perceived statement strength.

Deliberative failure. If people make stereotype rather than base-rate judgments for incongruent scenarios because they fail to deliberate, then prompting deliberation about base-rates should cause subjects to make more use of base-rates in judgments on such problems: Deliberation would highlight the conflict between the stereotype and base-rate information. We had no a-priori predictions regarding perceived statement strength.

Degree of deliberative support. How might the structure of the statements provided affect judgments? Subjects may merely need prompting to deliberate about previously provided information to increase its use. Alternatively, they may need to be told why information is relevant before it affects judgements. To test this, we varied whether the statements subjects evaluated explicitly stated why the information was relevant.

Numeracy. More numerate individuals are less susceptible to some judgment biases (Peters, Vastfjall, Slovic, Mertz, Mazzocco & Dickert, 2006) and are more likely to engage in deliberation (Pennycook, Cheyne, Barr, Koehler & Fugelsang, 2013). Since base-rates are numeric, it seems particularly likely that individual differences in numeracy might affect how base-rates are used. Therefore we included a numeracy measure to test whether the statement presentation manipulations would differentially affect subjects as a function of numeracy.

2 Method

2.1 Subjects

Undergraduate students at William Paterson University ($N=192$) and The Ohio State University ($N=246$) participated for course credit (total $N=438$). We excluded 29 additional surveys from consideration: 26 with incomplete judgement data plus 3 surveys identified as second attempts by subjects already in the sample. The same design and randomization was used at both universities.

2.2 Design

In an online survey, subjects completed an inference task. Task conditions were varied in a $2(\text{congruency}) \times 2(\text{base-rate-statement}) \times 2(\text{stereotype-statement}) \times 2(\text{statement-structure})$ mixed-model design.

Each subject judged twelve scenarios like the "Paul" example above. In six, the stereotype and base-rate information were *congruent* (e.g., Paul sounds like a doctor, and most of the population were doctors). In the other six, stereotype and base-rate information were *incongruent* (e.g., Paul sounds like a doctor, but most of the population 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). The text of all scenarios is available in the Supplement. Scenario order was randomized between subjects.

Critically, after reading the scenarios, but before making their judgments, we manipulated whether subjects were asked to evaluate statements that supported using base-rates and/or the stereotype information. Base-rate statement

(*evaluated or omitted*) and stereotype statement (*evaluated or omitted*) were crossed between subjects. For subjects who evaluated both base-rate and stereotype statements, we counterbalanced presentation order between subjects, resulting in two additional conditions (one each for the *explanation* and *reiteration* conditions explained below). Subjects were randomly assigned to one of the 10 resulting groups. However, the counterbalancing conditions were collapsed for the purposes of data analysis, yielding 8 groups in the final design.

The content of both the base-rate and stereotype statements was manipulated between subjects to provide either a reiteration of information with an *explanation* or just a *reiteration* of information. Statements with explanations provided both a restatement of previously given information (i.e., base-rates and/or stereotypical descriptions) followed by an explanation of why the information was relevant. Reiteration statements just restated the information; the explanation was omitted. For base-rate statements, the explanation asserted that a randomly selected individual was more likely to come from the category with more people. For stereotype statements, the explanation asserted that the description was more likely to fit a person from the stereotypical category, than one from the other category.

Example base-rate statement (*explanation in italics*).

“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.*”

Example stereotype statement (*explanation in italics*).

“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. *This description is more likely to fit a random doctor than a random nurse.*”

2.3 Scales

Subjects rated the statements they saw (if any) on a 1–7 scale where 1 = *Extremely Strong* and 7 = *Extremely Weak*.

Subjects judged group membership on a 6 point scale that allowed us to simultaneously obtain judgment and confidence data:

Do you think Paul is a doctor or a nurse? Please select one of the following:

- 1-Very confident that Paul is a doctor [Note: Strong stereotype response]
- 2-Moderately confident that Paul is a doctor
- 3-Slightly confident that Paul is a doctor
- 4-Slightly confident that Paul is a nurse

5-Moderately confident that Paul is a nurse

6-Very confident that Paul is a nurse [Note: Strong base-rate response]

2.4 Numeracy

Subjects completed an 8-item Objective Numeracy Scale (ONS) (Weller, Dieckmann, Tusler, Mertz, Burns & Peters, 2013). The text of this scale is available in the Supplement.

2.5 Procedure

Subjects at both universities completed the experiment online. They were randomly assigned to one of the 10 between subjects conditions: 2(base-rate-statement)×2(stereotype-statement)×2(argument-structure) + 2 (order-counterbalancing). All subjects responded to 12 scenarios (6 congruent, 6 incongruent) presented in different random order for each subject. For each scenario, subjects first read the scenario description (like the “Paul” example above), then they evaluated their assigned statement(s) (*neither, stereotype, base-rate, or both*), and finally they judged group membership. After the twelve scenario judgments were completed, subjects took the numeracy measure and provided demographic information.

3 Results

3.1 Coding

The 1–6 scale group-membership responses were averaged across the six congruent and six incongruent scenarios for each subject to create two “judgments” indicating the strength of base-rate responses for incongruent scenarios and the strength of stereotype/base-rate responses for congruent scenarios. In both cases, higher numbers imply more base-rate use. In the incongruent condition, lower numbers indicate more stereotype use. Base-rate and stereotype statement ratings were separately averaged over the 6 base-rate and/or 6 stereotype statement ratings in each congruency condition. Numeracy assessment scores equaled the total number of questions answered correctly out of eight. Non-responses were scored as incorrect. Six subjects who skipped the numeracy assessment were excluded from analyses involving numeracy. Numeracy scores were normally distributed around 4.28 ($SE=.10$, $range=0-8$; $Skewness=-.04$, $SE=.12$; $Kurtosis=-.66$, $SE=.23$).

3.2 Deliberation increased base-rate use

We tested two regression models, one for responses to incongruent scenarios, and another for responses to congruent scenarios. Base-rate-statement (given, omitted)

and stereotype-statement (given, omitted), numeracy (0–8 scale), and all of their interactions were entered as predictors and mean 6 point judgement rating for incongruent scenarios as the outcome. Statement content (reiteration only vs. reiteration with explanation) and its interactions are omitted here, as multiple analyses failed to show any significant effects of explanations on judgments. For incongruent scenarios, we found significant effects of stereotype statement ($\beta = .123, p = .010$), base-rate statement ($\beta = .182, p < .001$), and numeracy ($\beta = .125, p = .010$), as well as an interaction among all three factors ($\beta = .103, p = .033$). No other effects were significant.

As predicted by the deliberative failure and salience accounts, but *inconsistent* with the diagnosticity account, subjects who evaluated base-rate statements favored base-rate responses more than those who did not (base-rate statement omitted: $M = 2.57, SE = .07$; base-rate statement given: $M = 2.91, SE = .06$).

Contrary to the salience and diagnosticity accounts, subjects who evaluated stereotype statements actually showed *less* confidence in the stereotype response (stereotype statement omitted: $M = 2.61, SE = .07$; stereotype statement given: $M = 2.86, SE = .06$).

Also, more numerate subjects were more likely to use base-rates (median split; lower numeracy: $M = 2.64, SE = .06$, higher numeracy: $M = 2.84, SE = .07$), but generally showed less benefit from being prompted to deliberate (lower numeracy: no statements given: $M = 2.15, SE = .14$ vs. just base-rate: $M = 2.85, SE = .13$, just stereotype: $M = 2.63, SE = .15$, both: $M = 2.91, SE = .10$; higher numeracy: no statements given $M = 2.76, SE = .15$, just base-rate: $M = 2.68, SE = .17$, just stereotype: $M = 2.72, SE = .14$, vs. both given: $M = 3.19, SE = .11$). This would be expected if more numerate subjects are more likely to recognize the value of base-rates without prompting, but need the contrast of base-rate and stereotypical information to further increase their base-rate use.

We ran the same regression model predicting responses to congruent problems and found significant effects of stereotype statement ($\beta = -.112, p = .017$) and numeracy ($\beta = .280, p < .001$). Evaluating stereotype statements *decreased* subjects' confidence in the normative base-rate/stereotype responses (stereotype statement omitted: $M = 4.79, SE = .06$; stereotype statement evaluated: $M = 4.59, SE = .05$). Subjects who scored higher in numeracy gave ratings more in line with both base-rate/stereotype responses than those lower in numeracy (median split; lower numeracy: $M = 4.50, SE = .05$, higher numeracy: $M = 4.88, SE = .06$). No other effects were significant.

3.3 Base-rate statements rated stronger than stereotype statements

Contrary to the diagnosticity hypothesis, a 2(statement-type) \times 2(statement-structure) mixed ANOVA found that

subjects who viewed both statement types rated base-rate statements as stronger than stereotype statements ($F(1,176) = 23.87, p < .001, \eta^2_p = .119$; base-rate statement: $M = 2.68, SE = .09$; stereotype statement: $M = 3.32, SE = .08$; recall, smaller numbers indicate stronger ratings). Also, statements that provided explanations were rated stronger than those without explanations ($F(1,176) = 8.60, p = .004, \eta^2_p = .047$; explanation given: $M = 2.83, SE = .08$; explanation omitted: $M = 3.16, SE = .08$). The interaction was n.s. ($p > .7$). A parallel 2 \times 2 between-subjects ANOVA with subjects who only viewed one type of statement showed that base-rate statements were in the direction of being rated as stronger than stereotype statements, although the difference was n.s. ($F(1,169) = 2.30, p = .131, \eta^2_p = .013$; base-rate statement: $M = 3.14, SE = .13$; stereotype statement: $M = 3.42, SE = .13$). Statements with explanations were again rated significantly stronger ($F(1,169) = 7.30, p = .028, \eta^2_p = .028$; reiteration and explanations: $M = 3.08, SE = .13$; reiteration only: $M = 3.49, SE = .13$). Again, the interaction was n.s. ($p > .7$).

We checked whether subjects still rated base-rate statements as stronger than stereotype statements when looking only at statement ratings made for scenarios where subjects made judgments consistent with stereotypes (e.g., when subjects reported slight to strong confidence in the stereotype response). The pattern held for congruent scenarios, where base-rates and stereotypes support the same outcome (paired *t*-test among subjects rating both statements: $t(176) = 7.52, p < .001, d = .57$, base-rate statement: $M = 2.31, SE = .09$, stereotype statement: $M = 3.25, SE = .10, r_{ST \times BR} = .14$; independent *t*-test among subjects rating only one statement: $t(170) = 3.62, p < .001, d = .55$, base-rate statement: $M = 2.51, SE = .14$, stereotype statement: $M = 3.22, SE = .14$). However, for the incongruent scenarios, where base-rates and stereotypes support different outcomes, subjects rated stereotype statements as stronger, consistent with their own judgments (paired: $t(166) = 2.73, p = .007, d = .21$, base-rate statement: $M = 3.27, SE = .12$, stereotype statement: $M = 2.84, SE = .09, r = -.17$; independent: $t(164) = 4.44, p < .001, d = .69$ base-rate statement: $M = 4.02, SE = .18$, stereotype statement: $M = 3.06, SE = .13$).

3.4 Higher numeracy predicts stronger base-rate statement ratings

Regressions predicting base-rate and stereotype statement ratings from numeracy, statement structure, and their interaction found that more numerate subjects gave stronger base-rate statement ratings ($\beta = -.36, p < .001$). No significant relationship was found between numeracy and stereotype statement ratings ($\beta = .08, p = .178$), but both base-rate ($\beta = -.13, p = .024$) and stereotype statements ($\beta = -.16, p = .012$) were rated stronger when they provided explanations. Interactions were n.s. ($ps > .4$)

4 Discussion

Our results are consistent with the deliberative failure hypothesis. It appears that people give more weight to stereotype information than to base-rate information in part because they do not spontaneously engage in deliberative reasoning. Subjects who were prompted to deliberate about base-rate information (i.e., by evaluating base-rate statements) made more use of base-rates in their judgments. This cannot be attributed to statement evaluation simply making the base-rate information more salient, as evaluating stereotype statements did not similarly increase rates of stereotype judgements. On the contrary, evaluating stereotype statements significantly *decreased* stereotype use. These findings also stand in contrast to the diagnosticity hypothesis, which claims people believe stereotype information is more diagnostic of group membership than base-rate data. Under the diagnosticity hypothesis, inducing deliberation should not have affected subjects' choices, or perhaps may have shifted them towards the use of stereotypes. Moreover, subjects explicitly rated base-rate statements as stronger than the stereotype statements, even in the congruent condition (where these statements supported the same judgements). Although we did not predict a-priori that inducing deliberation about stereotype would decrease stereotype use, this outcome is fully in line with the deliberative failure hypothesis. If people believe stereotype information is less useful when they think about it, then inducing people to deliberate about stereotypes should result in less use of stereotype information.

Our results indicate higher numeracy is predictive of greater base-rate use. 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 benefit from interventions that promote use of numbers (Obrecht, 2010). It is currently unclear whether this is due to more numerate subjects being more likely to deliberate spontaneously (Pennycook, et al., 2013), or due to more numerate subjects having stronger intuitions about numbers (Schley & Peters, 2014).

We conclude that, while people may have spontaneous intuitions about base-rates (De Neys & Glumicic, 2008; Pennycook, et al., 2013), some, particularly the less numerate, do not appear to fully appreciate the value of base-rate information without deliberation. Interestingly, it seems just prompting deliberation was sufficient to increase base-rate use in our sample. Explicitly providing explanations about base-rate information's relevance yielded stronger statement ratings, but did not significantly increase base-rate use beyond that seen when evaluating statements without explanations. It appears that individuals who make use of base-rate information do not typically require an explicit account of why that information is useful.

Although these results show support for the deliberative failure account, we cannot conclude this is the only factor that accounts for neglect of base-rates and other normatively relevant data (e.g., see Barbey & Sloman, 2007 for an extensive review). People's choices generally still favored stereotype judgments when they were pitted against base-rate information, even when prompted to deliberate. While the stronger average base-rate statement ratings indicate that people appreciate base-rates and see them as formally stronger evidence than stereotype information, it may be that the value people give the stereotype information is not captured by their ratings of statement strength (e.g., see Thompson, 2009, regarding intuitive "feelings of rightness").

One limitation of this work is that we assume, but do not separately confirm, that having subjects evaluate statements prompts deliberation. This could be confirmed in future research, possibly looking at response times. Also, the rate of incongruent descriptions was necessarily higher than it would be in the real world. This may have resulted in subjects responding differently than they would have responded in a more realistic situation (see Koehler, 1996, for discussion). Indeed, past research has shown that people do not merely rely on explicitly provided information, such as base-rates, but also consider implied probabilistic information (e.g., the real-world likelihood of the scenario occurring, see Chesney & Obrecht, 2011, 2012; Obrecht & Chesney, 2013). Future studies with greater ecological validity are needed to address this issue. In sum, it appears that people do appreciate the importance of base-rates and make better use of them when prompted to deliberate. However, this is not sufficient to fully overcome the power of stereotypes.

References

- Barbey, A. K. & Sloman, S. A. (2007). Base-rate neglect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*, 30, 241–254. <http://dx.doi.org/10.1017/S0140525X07001653>.
- 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* (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. *Memory & Cognition*, 40, 420–433. <http://dx.doi.org/10.3758/s13421-011-0155-3>.
- De Neys, W. (2013). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking*

- & Reasoning, 20, 169–187. <http://dx.doi.org/10.1080/13546783.2013.854725>
- De Neys, W. & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, 106, 1248–1299. <http://dx.doi.org/10.1016/j.cognition.2007.06.002>.
- De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PLoS ONE*, 6, e15954. <http://dx.doi.org/10.1371/journal.pone.0015954>.
- De Neys, W., & Franssens, S. (2009) Belief inhibition during thinking: Not always winning but at least taking part. *Cognition*, 113, 45–61. <http://dx.doi.org/10.1016/j.cognition.2009.07.009>.
- 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.
- Koehler, J. J. (1996). The base rate fallacy reconsidering: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, 19, 1–17. <http://dx.doi.org/10.1017/S0140525X00041157>.
- Kogler, C. & Kuhberger, A. (2007). Dual process theories: A key for understanding the diversification bias? *Journal of Risk & Uncertainty*, 34, 145–154. <http://dx.doi.org/10.1007/s11166-007-9008-7>.
- Krosnick, J. A., Li, F., & Lehman, D. R. (1990). Conversational conventions, order of information acquisition, and the effect of base rates and individuating information on social judgments. *Journal of Personality and Social Psychology*, 59, 1140–1152. <http://dx.doi.org/10.1037/0022-3514.59.6.1140>.
- 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. <http://dx.doi.org/10.1016/j.actpsy.2013.01.012>.
- Obrecht, N. A. (2010). *Sample size weighting in probabilistic inference* (Doctoral dissertation). Rutgers University, New Brunswick.
- 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. <http://dx.doi.org/10.3758/s13421-013-0340-7>.
- 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. <http://dx.doi.org/10.1037/a0034887>.
- Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, 17, 408–414. <http://dx.doi.org/10.1111/j.1467-9280.2006.01720.x>.
- Schley, D. R. & Peters, E. (2014). Assessing “economic value”: Symbolic number mappings predict risky and riskless valuations. *Psychological Science*, 25, 753–761. <http://dx.doi.org/10.1177/0956797613515485>.
- Schwarz, N., Strack, F., Hilton, D., & Naderer, G. (1991). Base rates, representativeness, and the logic of conversation: The contextual relevance of “irrelevant” information. *Social Cognition*, 9, 67–84. <http://dx.doi.org/10.1521/soco.1991.9.1.67>.
- Slooman, S. A. (1996). The empirical case for type systems of reasoning. *Psychological Bulletin*, 119, 3–22. <http://dx.doi.org/10.1037/0033-2909.119.1.3>.
- Slooman, 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.
- 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.
- Trouche, E., Sander, E., & Mercier, H. (2014). Arguments, more than confidence, explain the good performance of reasoning groups. *Journal of Experimental Psychology: General*, 143, 1958–1971. <http://dx.doi.org/10.1037/a0037099>.
- 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. <http://dx.doi.org/10.1002/bdm.1751>.