Reasoning with base rates is routine, relatively effortless, and context dependent

Gordon Pennycook · Valerie A. Thompson

Abstract We tested models of base rate "neglect" using a novel paradigm. Participants ($N = 62$) judged the probability that a hypothetical person belonged to one of two categories (e.g., nurse/doctor) on the basis of either a personality description alone (NoBR) or the personality description and a base rate probability (BR). When base rates and descriptions were congruent, judgments in the BR condition were higher and more uniform than those in the NoBR condition. In contrast, base rates had a polarizing effect on judgments when they were incongruent with the descriptions, such that estimates were either consistent with the base rates or discrepant with them. These data suggest that the form of base rate use (i.e., whether base rates will be integrated with diagnostic information) is context dependent. In addition, judgments made under instructions to respond intuitively were influenced by the base rates and took the same length of time in the two conditions. These data suggest that the use of base rates is routine and effortless and that base rate "neglect" is really a mixture of two strategies, one that is informed primarily by the base rate and the other by the personality description.

Keywords Base rate neglect · Dual-process theory · Probability estimates · Conflict detection

G. Pennycook (✉)
Department of Psychology, University of Waterloo,
200 University Avenue West,
Waterloo, ON, Canada N2L 3G1
e-mail: gpennyco@uwaterloo.ca

V. A. Thompson
University of Saskatchewan,
Saskatoon, SK, Canada

Making successful judgments often requires consideration of prior probabilities or base rates; for example, a decision to purchase a particular make of car may be based on statistical information about its repair record. However, when such information is placed in the context of salient but often less reliable information (e.g., Uncle Joe's dissatisfaction with his version of it), prior probabilities may be undervalued (Kahneman & Tversky, 1973). Despite dozens of studies examining base rate neglect in both applied and experimental settings (see Barbey & Sloman, 2007, for a review), there is not yet a consensus on the cognitive mechanisms that produce it.

The goal of the present study is to test three accounts of base rate neglect, using a novel paradigm. We presented problems in a standard format, in which participants were provided with the base rate probabilities of category membership and a personality description that clearly favored one category over the other (De Neys & Glumicic, 2008; Kahneman & Tversky, 1973)—for example,

> "In a study 1000 people were tested. Among the participants there were 995 nurses and 5 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.
>
> What is the probability that Paul is a nurse?"

In contrast to the traditional paradigm, half of the participants (NoBR condition) made estimates solely on the basis of the personality descriptions (i.e., "995" and "5" were deleted); for the other half, both base rates and personality descriptions were provided (BR condition). For the latter, descriptions were congruent (i.e., stereotypes were

consistent with the large base rate), incongruent (i.e., stereotypes were inconsistent with the large base rate), or neutral (i.e., personality description contained no stereotypes) with respect to the base rates. Comparison of the two base rate conditions allowed us to determine how base rates were used, and comparisons among the congruency conditions allowed us to assess the role of conflict detection in base rate use. Note that although there could be implicit base rate information available in the NoBR condition (e.g., there are generally more nurses than doctors in a normal population), implicit prior probabilities do not normally influence judgments (Tom W. problem, Kahneman & Tversky, 1973).

Models of base rate neglect and predicted response patterns

Base rates and diagnostic information are integrated The first hypothesis was that participants integrate base rates and diagnostic information regardless of their relationship to each other (Koehler, 1996; Novemsky & Kronzon, 1999). For example, changes to base rate quantities affect judgments in a wide range of tasks (Koehler, 1996), and several manipulations increase reliance on base rates, such as presenting them as frequencies, rather than percentages (Nisbett, Krantz, Jepson, & Kunda, 1983). Koehler concluded that base rate "neglect" was really a failure to sufficiently adjust toward the base rate. According to this integration model, probability estimates in the BR condition should be shifted in the direction of the base rate probability relative to the NoBR condition, regardless of congruency.

Base rate and diagnostic information are not integrated An alternative view (Evans & Elqayam, 2007; Evans, Handley, Perham, Over, & Thompson, 2000) is that reasoners tend to give answers consistent with only one piece of information provided in the problem, which may or may not be the base rate. This model therefore predicts that introducing base rates should produce bimodal response distributions for incongruent problems, because participants will sometimes rely on the personality descriptions and will sometimes rely on the base rate information. For congruent problems, which may afford construction of a "set inclusive model" (Evans & Elqayam, 2007, p. 262), it is possible that the two pieces of information may be integrated.

Intuitive and analytic thinking A third approach to understanding base rate neglect comes from dual-process theories (e.g., Evans, 2008), which posit that reasoning and decision making are based on two qualitatively different processes: Heuristic processing is fast, frugal, and intuitive, whereas analytic processing is slow and deliberate. A common explanation for base rate neglect is that the personality description evokes a compelling stereotype, which is made available quickly (Bonner & Newell, 2010; De Neys & Glumicic, 2008) and forms a default basis for judgment unless analytic processes intervene to override the default response (Kahneman, 2003). If this were the case, base rates should influence judgments only under conditions that facilitate analytic processing.

To test this hypothesis, participants were tested using a two-response paradigm (Thompson, Prowse Turner, & Pennycook, 2011). Participants answered each problem twice: with the first answer that came to mind and then with a free time response.[1] The underlying assumption is that the first response that comes to mind is the outcome of largely intuitive processes, whereas rethinking time is a proxy for deliberate, analytic processing (De Neys, 2006). Thus, if processing base rates requires analytic thinking, differences in the probability estimates or response times (RTs) between the BR and NoBR conditions should be observed for final, but not initial, responses.

A recent modification to this theory assumes that analytic thinking is engaged to resolve conflict (De Neys & Glumicic, 2008). This requires a *shallow analytic monitoring* system that detects conflicts, such as those between a base rate and a personality description (De Neys & Glumicic, 2008; De Neys, Vartanian, & Goel, 2008). In the case of base rate reasoning, the stereotype is assumed to form the basis of a default response unless analytic processes are engaged to overcome it. Thus, differences between the BR and NoBR conditions should be observed only for incongruent problems; also, given that incongruent problems trigger analytic processing, they should take longer than congruent problems, especially for the second response.

However, given that the assumption underlying shallow monitoring is that information about base rates and stereotypes are both made available quickly, one might posit a more active contribution of the base rate information to judgments (De Neys, 2012). Thus, while the conflict *resolution* dimension of the model suggests that analytic processing is required to overcome the default, stereotypical response, the conflict *detection* dimension suggests that base rates are accessible to initial intuitive processing, because such information needs to be available for a conflict to be detected in the first place. If this were the case, differences between the BR and NoBR conditions should emerge, even on the initial response, and for all problem types.

---

[1] See Thompson et al. (2011) for a detailed description of the two-response paradigm and evidence that asking participants to give the initial response does not change response patterns, relative to a more traditional, single-response paradigm.

Method

Participants Sixty-two volunteers from the University of Saskatchewan were paid $5 to participate (46 female, mean age = 22.9 years). Thirty-two were assigned to the BR condition, and 30 were assigned to the NoBR condition.

Materials Eighteen base rate problems (adapted from De Neys and Glumicic, 2008; all similar to the example provided previously) and one practice problem were presented on a computer monitor using E-Prime v1.2. In the BR condition, there were three problem types: (1) Base rates and stereotype pointed to the same response (congruent), (2) base rates and stereotype pointed to different responses (incongruent), and (3) personality description contained no stereotype (neutral). Three base rate ratios were presented equally often: 995/5, 996/4, and 997/3. Extreme ratios were used to maintain consistency with De Neys and Glumicic. To counterbalance content among congruency conditions, two sets of problems were created so that each personality description matched the larger group (congruent) or smaller group (incongruent) an equal number of times. Problem order was randomized for each participant.

Procedure Instructions were adapted from De Neys and Glumicic (2008). Participants were told that they would read a description of studies where participants were drawn randomly from two population groups; these would contain a personality description of the person, as well as information about the composition of the groups. They were asked to provide a probability estimate, out of 100, indicating the likelihood that the person belonged to the specified group.

Participants provided two answers: the first answer that came to mind and a final answer. It was emphasized that the first answer was to be their first inclination or instinct. To reinforce this, the problem changed color and was italicized after 12 s. This deadline was chosen on the basis of pilot studies. Participants were then asked whether they actually had responded with their first answer. Participants responded affirmatively to this question 96.5% of the time. Trials on which participants responded "no" were excluded from further analysis. Participants were then allowed all the time they needed to make their final answer; they were instructed to take their time and think about the problem carefully. Response time was measured for each response, beginning at initial presentation of the problem. Participants were tested individually, and testing took approximately 25 min.

Results

For the BR condition, probability estimates for items that asked about the smaller of the two groups (e.g., there were 995

doctors and 5 nurses; what is the probability that Paul is a nurse?) were subtracted from 100. Thus, high scores always indicated estimates that were close to the base rate, and low scores reflected estimates that deviated from the base rate. To make the data for the NoBR condition comparable for the incongruent problems, the estimates for the NoBR "incongruent" problems were subtracted from 100 so that low numbers for incongruent problems in both conditions reflected estimates based on the stereotypes.

The distribution of probability estimates: Are base rates and diagnostic information integrated? Did participants integrate the base rates and personality descriptions (Koehler, 1996; Novemsky & Kronzon, 1999), or did they base their responses wholly on one or the other (Evans & Elqayam, 2007)? To answer this, we examined the distribution of probability estimates. Figures 1 and 2 plot the initial responses to congruent and incongruent problems in the two conditions.[2]

Consistent with the hypothesis that participants integrated the base rates with the personality descriptions, probability estimates for congruent problems were higher and more uniform in the BR than in the NoBR condition (see Fig. 1), so that judgments clustered around the base rate (an ANOVA supporting this claim is reported below; see Table 1). While it is possible that responses were higher in the BR condition because participants focused on the base rate information and ignored the personality description, this account is inconsistent with the large amount of data showing that participants' judgments are biased in favor of the personality descriptions when they conflict with the base rates (Barbey & Sloman, 2007).

In contrast, for incongruent problems, the distribution of BR responses was bimodal. Figure 2 shows two clusters of responses on opposite sides of the scale in the BR condition, with most higher than 90 or lower than 10 (high estimates reflect consistency with the base rate). Shilling, Watkins, and Watkins (2002) argued that bimodality can be inferred when the means of two distributions differ by more than the sum of their standard deviations. We therefore separated the probability estimates for incongruent problems in the BR condition into two distributions (0–49 and 51–100). For the initial response, the means were 9.6 ($SD = 8.6$) and 92.0 ($SD = 8.6$) for the 0–49 and 51–100 groups, respectively; for the final answer, the means were 8.0 ($SD = 7.4$) and 93.2 ($SD = 6.7$), respectively. The difference between these means (82.4 and 85.2 for initial and final answers, respectively) is much larger than the sums of their standard deviations (17.2 and 14.1 for initial and final answers, respectively), indicating that the distribution is bimodal. Note also that most participants (75%) gave at least one high and low response.

---

[2] Only the response distributions for initial responses are plotted, since the distributions for final answers are similar.
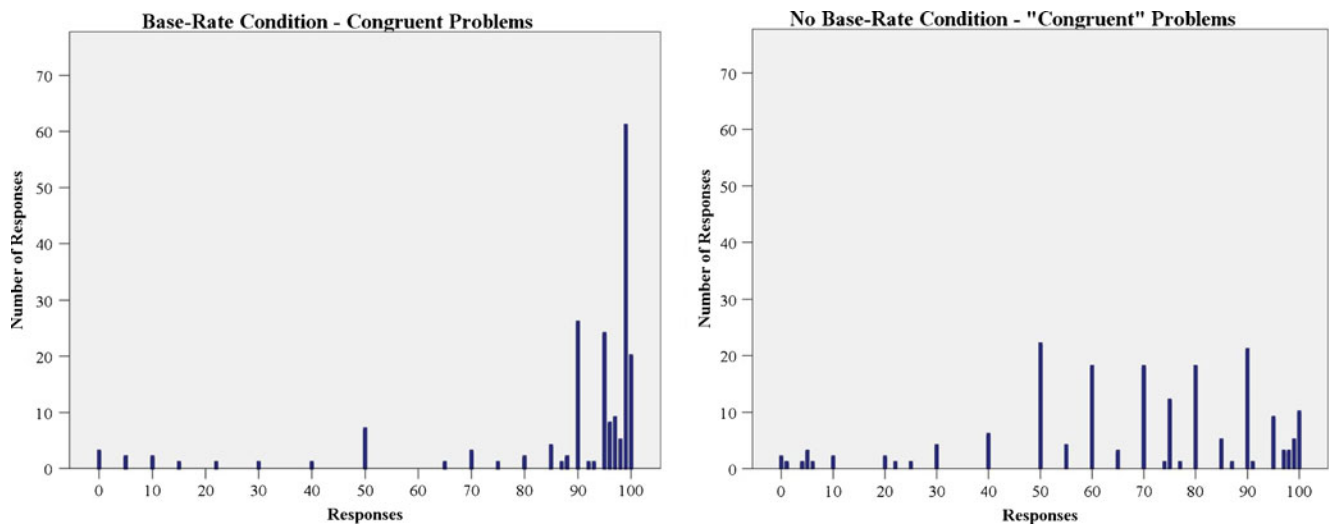
Fig. 1 Distribution of responses for congruent problems. For the base rate condition, high responses are consistent with both stereotypes and base rates. For the no-base-rate condition, high responses are consistent with stereotypes. Note that problems were not "congruent" in the no-base-rate condition, due to the lack of base rate information

In sum, whether or not base rates and stereotypes are integrated depends on whether they converge. When base rates and diagnostic information were consistent with each other, participants successfully integrated them, but when they conflicted, they responded on the basis of one or the other. These data are consistent with the hypothesis that reasoners give answers consistent with only one piece of information, unless the problem affords a set-inclusive model (Evans & Elqayam, 2007; Evans et al., 2000).

The role of conflict detection and analytic thinking in base rate usage We first analyzed probability estimates using separate 2 × 2 (time × base rate condition) mixed ANOVAs for each of the three problem types. The data are presented in Table 1.

For all three problem types, there was a main effect of time, all $Fs(1, 60) \geq 6.79$, all $ps \leq .012$, and condition, all $Fs(1, 60) \geq 12.44$, all $ps \leq .001$; the interaction was not reliable for the congruent or incongruent problems, $F < 1$, but it was for the neutral problems, $F(1, 60) = 4.84$, $MSE = 70.43$, $p = .032$. The difference between the BR and NoBR conditions provided clear evidence that the base rates influenced judgments for all problem types and at both response opportunities.[3] These data are not consistent with the hypothesis that processing base rates requires analytic thinking (Bonner & Newell, 2010; De Neys & Glumicic, 2008), since the effect of the base rate manipulation was observed under conditions designed to minimize analytic thinking. Moreover, conflict was not a

necessary precondition for the use of base rate information, contrary to the assumption that responses to congruent problems are based solely on the personality description (De Neys & Glumicic, 2008).

In terms of the main effect of time, estimates for congruent problems increased from time 1 to time 2 (Table 1), suggesting that additional thinking led to answers closer to the base rate. However, the congruent problems in the BR condition are ambiguous, since responses could be based on either the description or the base rates. For the nonambiguous incongruent problems, estimates *decreased* over time (Table 1), suggesting that they were pulled toward the stereotypes. Moreover, estimates in both of the NoBR conditions also moved toward the answer suggested by the description (Table 1). In other words, the additional opportunity for analytic processing appeared to increase, rather than decrease, reliance on the stereotype. This, of course, does not mean that base rates are never processed analytically. In fact, estimates for neutral problems in the BR condition shifted toward the base rate, $t(32) = 3.18$, $SE = 2.59$, $p = .003$, with no change in the NoBR condition, $t(30) = 1.13$, $SE = 1.42$, $p = .267$. Thus, base rates had a substantial influence on estimates when they were paired with nondiagnostic personality descriptions. Taken as a whole, these data challenge the dual-process theory assumption (e.g., Bonner & Newell, 2010; De Neys & Glumicic, 2008) that judgments based on the stereotype are default, intuitive responses and that analytic processes change the default in favor of the base rates. Both base rates and stereotypical information are apparently reasoned with via analytic *or* intuitive processing.

Analysis of the RT data also supported this conclusion (see Table 2). RTs that were 3 SDs from the mean were excluded as outliers. All RTs were converted to $\log^{10}$ prior

---

[3] Due to heterogeneity of variances between the BR and NoBR conditions, we also analyzed the data using nonparametric methods and obtained the same pattern of results.
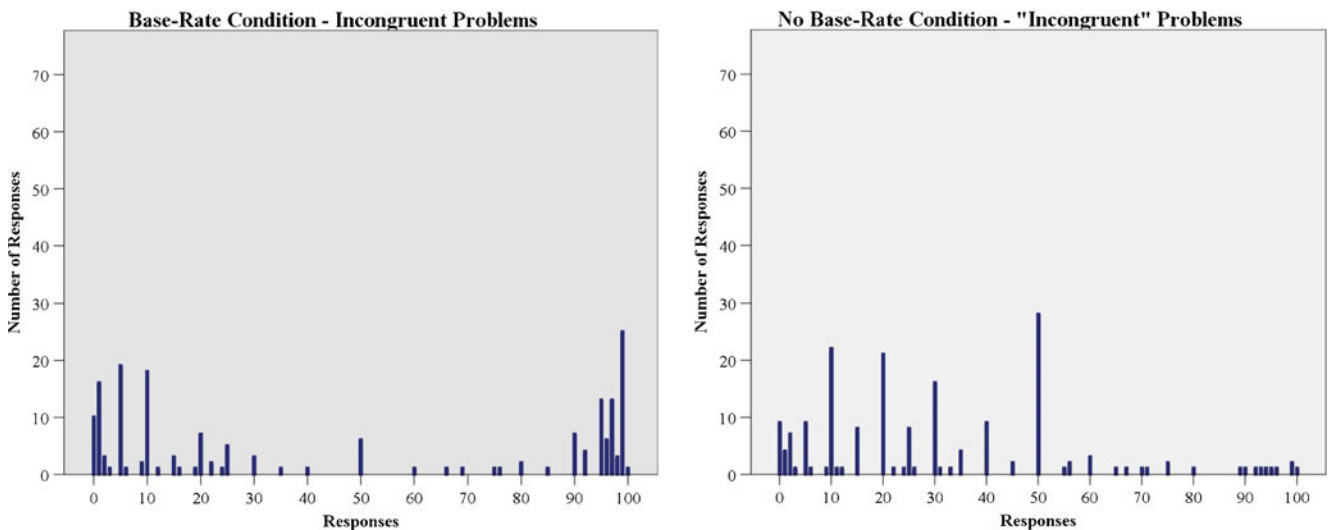
Fig. 2 Distribution of responses for incongruent problems. For the base rate condition, high responses are consistent with base rates, and low responses are consistent with stereotypes. For the no-base-rate condition, low responses are consistent with stereotypes. Note that problems were not "incongruent" in the no-base-rate condition, due to the lack of base rate information

to analysis (RTs in Table 2 are reported in the original units). Separate 3 × 2 (problem type × base rate condition) mixed ANOVAs were computed for both initial and final responses. For the initial response, there was a main effect of congruency, $F(1.8, 107.8) = 6.04$, $MSE = .002$, $p = .004$ (see note 3), replicating De Neys and Glumicic's (2008) finding that participants take longer to respond to incongruent and neutral problems than to congruent problems (see Table 2). However, there was no main effect of condition, $F(1, 60) < 1$, and no interaction, $F(1.8, 107.8) = 2.06$, $MSE = .002$, $p = .137$ (see note 3). That is, despite the evidence that base rates influenced judgments, initial RTs did not differ for the BR and NoBR conditions for any problem type (see Table 2). Again, this finding is inconsistent with the assumption that base rate use requires slow analytic processing; alternatively, it suggests that reasoning about both base rates and personality descriptions can rely on heuristic processes.

For the final answer RT, there was a main effect of congruency, $F(2, 120) = 3.99$, $MSE = .016$, $p = .022$, a main effect of condition, $F(1, 60) = 4.89$, $MSE = .183$, $p = .031$, and an interaction, $F(2, 120) = 6.57$, $MSE = .016$, $p = .002$. The cause of this interaction was larger RTs for the incongruent problems in the BR condition than in the NoBR condition, $t(60) = 2.27$, $SE = .068$, $p = .027$, but not for the nonconflict congruent problems, $t(60) = 0.794$, $SE = .065$, $p = .430$ (see Table 2). These data are consistent with the conflict detection model proposed by De Neys and Glumicic (2008), in that conflict promoted analytic thinking. However, given the evidence above, it does not necessarily suggest that base rate use requires analytic thinking. Instead, it may be the case that reasoners were spending the additional time attempting to decide which piece of information (i.e., stereotype and/or base rate) they should utilize.

Consistent with this hypothesis, probability estimates for incongruent problems in the BR condition shifted both

Table 1 Probability estimates for initial response and final answer as a function of congruency and condition

	Initial response		Final answer	
	BR condition	NoBR condition	BR condition	NoBR condition
Incongruent				
	49.18 (5.3)	69.63 (1.9)	46.25 (5.9)	75.94 (2.1)
Congruent				
	88.57 (1.9)	68.53 (2.0)	93.88 (1.2)	73.46 (2.5)
Neutral				
	74.01 (3.4)	44.87 (2.2)	82.26 (3.2)	46.48 (2.2)

Note. BR, base rate. Standard errors are in parentheses

Table 2 Response times (RTs, in seconds) for initial response and final answer as a function of congruency and condition

	Initial response		Final answer	
	BR condition	NoBR condition	BR condition	NoBR condition
Incongruent				
	13.39 (.47)	13.66 (.50)	15.39 (1.04)	11.47 (1.09)
Congruent				
	12.81 (.48)	13.13 (.47)	11.86 (.92)	11.06 (1.15)
Neutral				
	14.21 (.62)	13.47 (.51)	14.81 (1.02)	10.69 (1.17)

Note. BR, base rate. Standard errors are in parentheses

toward and away from the base rates (see Table 1). Although the overall trend was toward the stereotype (as evidenced by the main effect reported above), participants who changed their answers to the incongruent problems (53.8%) were just as likely to shift their estimate toward the base rate (29.7%) as away from it (23.1%). These data support the conclusion that participants' analytic thinking was directed toward deciding which piece of information was most reliable.

Discussion

The goal of the present work was to test three models of base rate neglect. In doing so, several novel findings emerged. First, in support of Koehler's (1996) integration model, reasoners appeared to integrate the base rates and stereotypes when they were congruent. However, when they diverged, participants gave answers consistent with only one source of information, suggesting a failure to integrate (Evans & Elqayam, 2007). Thus, while models of base rate neglect may differ on the basis of whether base rate probabilities and diagnostic information are integrated (Koehler, 1996), our data suggest that strategies for utilizing base rates were context dependent, such that integration varied as a function of the relation between prior probability and diagnostic information.

While the failure to integrate for conflict problems may have been due to a lack of capacity or motivation to do the necessary calculations, we suggest that participants view the two information sources as incompatible and focus their efforts, instead, on attempting to determine which is most reliable. The latter explanation is consistent with Evans and Elqayam's (2007) hypothesis that integration can occur only when the problem cues construction of "set-inclusive mental models" (p. 262). Our data suggest that conflict cues separate mental models: one based on the statistical information and the other based on the personality description.

A separate but related question concerns the cognitive mechanisms that underlie reasoning with base rates. Many researchers have categorized responses based on personality descriptions as *heuristic* and those based on the base rate as *analytic* (e.g., De Neys & Glumicic, 2008). Our data suggest that this categorization is too simplistic. When offered the opportunity to rethink their initial answer, participants were just as likely to shift toward the stereotype as toward the base rate, suggesting that both types of answers can be the outcome of analytic processing. Similarly, many participants gave answers consistent with the base rates when making their initial, presumably intuitive, response, and doing so did not require additional time relative to the NoBR condition. Thus, it appears that fast "intuitive" decisions could be based on the statistical information—a situation that was perhaps facilitated by the very large proportions provided.

While this conclusion is counter to much theorizing in the field (e.g., Barbey & Sloman, 2007), others have made similar claims. For example, Koehler (1996) surmised that participants may be intuitively aware of large base rates (see also De Neys, 2012).

Thus, consistent with De Neys and Glumicic's (2008) shallow monitor, it is clear that information about both base rates and stereotypes are available quickly, leading to highly efficient conflict *detection* (De Neys, 2012). Inconsistent with their view on conflict *resolution*, however, the base rate information appeared to have been processed regardless of whether a conflict had been detected. How then do we explain the evidence that seems to suggest a link between base rate usage and analytic processing? Specifically, participants are more likely to reread and later recall the base rates when they are incongruent with the description (De Neys & Glumicic, 2008), and base rate usage decreases under cognitive load (Franssens & De Neys, 2009). We concur with De Neys and colleagues that an analytic mode of thinking was triggered in response to the conflict, and indeed, we found evidence that participants took longer rethinking incongruent than congruent problems. However, while reasoners may be thinking analytically, we propose that they are attempting to determine which piece of information to base their decisions on—a choice that is not necessary for the congruent problems. Under this explanation, participants tend to base their decision on the stereotype when put under cognitive load, because it is more salient intuitively than the base rate information (Barbey & Sloman, 2007; De Neys, 2012), and not because base rates require analytic processing (Franssens & De Neys, 2009). Thus, the phenomenon known as "base rate neglect" arises from averaging across two routine and relatively effortless strategies: one that relies on the base rate information and the other that relies on the personality description.

References

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioural and Brain Sciences, 30,* 241–256.

Bonner, C., & Newell, B. R. (2010). In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. *Memory & Cognition, 38,* 186–196.

De Neys, W. (2006). Automatic-heuristic and executive-analytic processing in reasoning: Chronometric and dual task considerations. *Quarterly Journal of Experimental Psychology, 59,* 1070–1100.

De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science, 7,* 28–38.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition, 106,* 1284–1299.

De Neys, W., Vartanian, O., & Goel, V. (2008). When our brains detect that we are biased. *Psychological Science, 19,* 483–489.

Evans, J. St B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59,* 255–278.

Evans, J. St B. T., & Elqayam, S. (2007). Dual-processing explains base-rate neglect, but which dual-process theory and how? *Behavioural and Brain Sciences, 30,* 261–262.

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

Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking and Reasoning, 15,* 105–128.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58,* 697–720.

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

Koehler, J. J. (1996). The base-rate fallacy reconsidered: Descriptive, normative and methodological challenges. *Behavioural and Brain Sciences, 19,* 1–53.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90,* 339–363.

Novemsky, N., & Kronzon, S. (1999). How are base-rates used, when they are used: A comparison of additive and Bayesian models of base-rate use. *Journal of Behavioral Decision Making, 12,* 55–67.

Shilling, M. F., Watkins, A. E., & Watkins, W. (2002). Is human height bimodal? *The American Statistician, 56,* 223–229.

Thompson, V. A., Prowse Turner, J., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology, 63,* 107–140.