American Psychological Association

Journal of Experimental Psychology: General

0096-3445/18/$12.00

2018, Vol. 147, No. 7, 945–961
http://dx.doi.org/10.1037/xge0000457

Do Smart People Have Better Intuitions?

Valerie A. Thompson
University of Saskatchewan

Gordon Pennycook
Yale University

Dries Trippas
Max Planck Institute for Human Development, Berlin, Germany

Jonathan St. B. T. Evans
University of Plymouth

There is much evidence that high-capacity reasoners perform better on a variety of reasoning tasks (Stanovich, 1999), a phenomenon that is normally attributed to differences in either the efficacy or the probability of deliberate (Type II) engagement (Evans, 2007). In contrast, we hypothesized that intuitive (Type I) processes may differentiate high- and low-capacity reasoners. To test this hypothesis, reasoners were given a reasoning task modeled on the logic of the Stroop Task, in which they had to ignore one dimension of a problem when instructed to give an answer based on the other dimension (Handley, Newstead, & Trippas, 2011). Specifically, in Experiment 1, 112 reasoners were asked to give judgments consistent with beliefs or validity for 2 different types of deductive reasoning problems. In Experiment 2, 224 reasoners gave judgments consistent with beliefs (i.e., stereotypes) or statistics (i.e., base-rates) on a base rate task; half responded under a strict deadline. For all 3 problem types and regardless of the deadline, high-capacity reasoners performed better for logic/statistics than did belief judgments when the 2 conflicted, whereas the reverse was true for low-capacity reasoners. In other words, for high-capacity reasoners, statistical information interfered with their ability to make belief-based judgments, suggesting that, for them, probabilities may be more intuitive than stereotypes. Thus, at least part of the accuracy–capacity relationship observed in reasoning may be because of intuitive (Type I) processes.

Keywords: dual process theories of reasoning, individual differences in reasoning, intuition, analytic thinking

Supplemental materials: http://dx.doi.org/10.1037/xge0000457.supp

Dual process theories of reasoning assume that reasoning and decision making are accomplished by two qualitatively different sets of processes (Evans & Stanovich, 2013). Type I processes are autonomous, require little cognitive capacity, and are generally executed quicker than are Type II processes. Type II processes, by contrast, require working memory and tend to unfold more slowly. These models have traditionally been used to explain biases arising from Type I intuitions, which may or may not be corrected by subsequent Type II analysis. For example, prior beliefs may lead to erroneous judgments of the validity of logical arguments; instructions to reason logically increases accuracy, but only among high-capacity reasoners (Evans, Handley, Neilens, & Over, 2010). Indeed, the relationship between cognitive capacity and reasoning performance is a critical pillar supporting dual process frameworks. The purpose of the current article is to provide a direct test of the basis of this relationship.

Numerous studies have demonstrated that there is a robust relationship between performance on many reasoning tasks and measures of cognitive capacity, such as IQ (Stanovich, 1999, 2009). These tasks are novel from the perspective of most reasoners, such that they cannot rely on past experience to solve them and must, instead, engage Type II thinking. Moreover, in cases where Type I processes produce an initial answer that is based on beliefs or other heuristic processes, Type II processes are needed to overturn the initial answer. As a consequence, the capacity–reasoning relationship is thought to reflect the need for cognitive capacity to overturn Type I outputs in favor of Type II reasoning: namely, to inhibit the Type I answer, to create an alternative representation of the problem, and to engage the requisite logical or numerical operations needed to produce the correct answer. In addition to capacity, reasoners also require the motivation or disposition to implement the override, so that measures of thinking dispositions also correlate with performance on many reasoning tasks (e.g., Stanovich & West, 1998, 2000; Toplak, West, &

Valerie A. Thompson, Department of Psychology, University of Saskatchewan; Gordon Pennycook, Department of Psychology, Yale University; Dries Trippas, Centre for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany; Jonathan St. B. T. Evans, Department of Psychology, University of Plymouth.

This work was funded by the Natural Sciences and Engineering Research Council of Canada. We thank Jamie Campbell and Wim De Neys for comments provided on a draft.

Correspondence concerning this article should be addressed to Valerie A. Thompson, Department of Psychology, University of Saskatchewan, 9 Campus Drive, Saskatoon, SK S7N 5A5, Canada. E-mail: valerie.thompson@usask.ca

Stanovich, 2011). Implicit in this explanation, and indeed, in the dual-process explanation for many phenomena, is the temporal asymmetry of Type I and Type II processes: Autonomous (Type I) processes are normally faster, so that they form a default response, which requires intervention by slower, Type II processes.

Several challenges to this straightforward interpretation have recently arisen. For example, on a variety of tasks, De Neys and colleagues have shown that reasoners are sensitive to the conflict between validity/probability and beliefs, even when they give responses based on belief (for reviews, see De Neys, 2012, 2014). The challenge is that responses based on validity/probability should not be generated quickly enough to interfere with the presumably fast, autonomous, belief-based ones. More recently, Newman, Gibb, and Thompson (2017) have demonstrated that reasoners are able to give responses consistent with logical validity and probability on conditional inference and base rate tasks, even when required to respond under a very challenging deadline (see also Bago & De Neys, 2017). Again, this pattern would not be expected if belief-based judgments were faster and more intuitive than those based on probability.

These findings challenge the pervasive assumption that belief-based responses necessarily reflect Type I processing and logical/probability-based responses reflect Type II responding. Indeed, more recent arguments emphasize that there is no linkage of this kind that necessarily applies regardless of task and instructions (see Evans & Stanovich, 2013). Nonetheless, because many of the dual process-based explanations for phenomena appeal to speed of processing, data showing that both belief-based and validity-based reasoning can happen quickly is a challenge. Regardless, a perhaps more direct challenge is posed by methods that demonstrate that reasoners may generate responses based on validity or probability, even when those responses are goal-inconsistent (Morsanyi & Handley, 2012; though see Klauer & Singmann, 2013 and Trippas, Handley, Verde, & Morsanyi, 2016, for a debate).

To this end, Handley et al. (2011) developed a novel version of a logical reasoning task, who underpinning logic is akin to that of the Stroop task (MacLeod, 1991). Reasoners were asked to solve simple logical reasoning problems with instructions to respond on the basis of believability or validity. If belief-based processes form an initial, default answer, then belief-based judgments should interfere with judgments based on validity more than vice versa. The logic is similar to that underlying a Stroop task; namely, that the process of reading the word gives rise to a prepotent response that interferes with the process of color-naming to a greater extent than the reverse. Instead, Handley et al. observed that in cases where the two responses conflicted (i.e., valid-unbelievable conclusions and invalid-believable conclusions), the validity of the conclusion interfered with the ability to make belief-based judgments at least as much as the reverse. This is an unexpected phenomenon given the assumption that judgments of validity can be made only after a default, belief-based response has been inhibited; by analogy, it would be surprising to find that naming colors interfered with reading as much as reading interfered with color naming.

Handley et al.'s (2011) finding were replicated using a base-rate task, based on Kahneman and Tversky's (1973) lawyer and engineer problem (Appendix B): The value of the base rates interfered with making belief-based judgments, even under time pressure (Pennycook, Trippas, Handley, & Thompson, 2014), which again challenges the assumption that belief-based responses are more likely to form initial, Type I responses to problems. A more nuanced view of these findings is that belief-based and validity-based responses exist on a continuum of complexity (Handley & Trippas, 2015), so that when logical judgments are simple, as in the case of Handley et al. (2011), validity will interfere more with belief than vice versa. However, as the relative difficulty of the logical inferences increases, this relationship reverses, such that on the most difficult problems, beliefs interfere more with judgments of validity than vice versa (Trippas, Thompson, & Handley, 2017). It is this latter pattern (but not the former) that is predicted by past Dual Process Theories (Evans & Stanovich, 2013).

These data have also led to the development of a number of alternative Dual Process formulations, in which reasoning problems may elicit more than one Type I response (De Neys, 2012, 2014; Handley & Trippas, 2015; Pennycook, Fugelsang, & Koehler, 2015; Pennycook & Thompson, 2012; Trippas & Handley, 2018). One of these outputs will form a default response, either because it is more fluent than the other (Pennycook et al., 2015), or because it requires less complex processing than the other (Handley & Trippas, 2015). As in traditional Dual Process Theories, Type II processes may or may not be engaged to overturn this initial response; however, the clear implication is that answers that were previously attributed to Type II processes under the "received view" (Evans & Stanovich, 2013) may have been generated, instead, by Type I processes.

In the current article, we extend this reasoning to the capacity–reasoning relationship. As described previously, the relationship between IQ and reasoning is largely thought to rest on the asymmetry between fast, autonomous Type I processes and slower, WM-dependent Type II processes. On this view, it is assumed that high IQ reasoners have the WM capacity to override the default response and provide an alternative based on logic or probability (Evans & Stanovich, 2013), an assumption justified by the high correlation between measures of IQ and WM (e.g., Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004; Kane, Hambrick, & Conway, 2005). That is, the relationship between IQ and reasoning is usually attributed to Type II processes. High IQ people are assumed to either have greater capacity or greater disposition to overcome a prepotent Type I response (Evans, 2007). The goal of the current article is to provide a direct test of that hypothesis.

There is already preliminary data to challenge the role of Type II thinking in the IQ–reasoning relationship (Thompson & Johnson, 2014). Reasoners were asked to give two responses to each of four types of reasoning problems: a quick, intuitive one and a slower, deliberative one (Thompson, Prowse Turner, & Pennycook, 2011). The correlation between IQ and reasoning performance was observed on *both* the first and second response, which is difficult to explain under the assumption that the accuracy-capacity relationship emerges only when Type II processing is cued. Instead, high-capacity reasoners can quickly give responses based on logic or probability, suggesting that at least part of the accuracy–capacity relationship may be because of Type I processing. This hypothesis is also consistent with suggestions that numerate reasoners may have better numerical intuitions than less numerate ones (Peters, 2012; Peters, Slovic, Västfjäll, & Mertz, 2008). The goal of the current experiments was to test this hypothesis by using a variant of the instructional manipulation developed by Handley et al. (2011).

Participants solved three-term syllogisms (Experiment 1; Appendix A) and base-rate problems (Experiment 2; Appendix B). For the syllogisms, participants were instructed to respond on the basis of logic or beliefs (Handley et al., 2011) and for the base rate problems, we used Pennycook et al.'s (2014) instructional manipulation to respond either according to "statistics" (prior probabilities or base-rates) "beliefs" (knowledge of real-world stereotypes). On congruent problems, responding on the basis of either beliefs or validity/probability yielded the same response, as illustrated below; in contrast, on incongruent problems, responses based on probability/validity conflicted with those based on belief (see below). Participants also completed a set of measures designed to measure analytic thinking disposition and cognitive capacity. In each experiment we built in a replication. For Experiment 1, we included two different types of problems, and, in Experiment 2, we included a deadline condition to minimize the influence of Type II processing (Evans & Curtis-Holmes, 2005).

Congruent (Believable and valid):
All firters are vegetables.
Some carrots are firters.
Therefore, some vegetables are carrots.
Incongruent (Believable and invalid):
All cannons are pooblings.
No pooblings are weapons.
Therefore, some cannons are weapons.

The critical prediction is as follows: If responses based on logic and probability are relatively more accessible for high-capacity reasoners (Thompson & Johnson, 2014), it follows that we should observe different patterns of interference as a function of capacity: high-capacity reasoners should have relatively more difficulty resolving conflict in favor of beliefs than probabilities/validity, whereas low-capacity reasoners should have relatively more difficulty resolving conflict in favor of probabilities/validity than beliefs. That is, if judgments of probability/validity are relatively more intuitive for high-capacity reasoners, then they should interfere with that group's ability to make judgments based on belief and they should perform more poorly on incongruent problems under belief than logic instructions. In contrast, in cases where there is no conflict, reasoners should perform equally well under the two types of instructions. This would manifest as a three-way interaction between capacity, instructions, and conflict.

Three additional hypotheses were also considered. First, the traditional dual-process interpretation is that high-capacity reasoners will be better than low-capacity reasoners in resolving conflict in favor of probability/validity. This hypothesis is based on the assumption that judgments of belief form a default response, and that capacity is needed to overturn that response to make one based on probability or validity. In contrast, both groups should be able to resolve conflict in favor of beliefs, such that performance should be comparable for high- and low-capacity reasoners, and the groups should do equally well on congruent problems.

The second hypothesis is that reasoners with strong analytic thinking dispositions may not experience a conflict between beliefs and logic at all (as suggested by Svedholm-Häkkinen, 2015). Using a priming paradigm, Svedholm-Häkkinen (2015) found that incongruent problems primed subsequent answers for low, but not high ability reasoners, suggesting that the latter did not experience conflict. However, the evidence was inconsistent, in that many of the critical interaction effects were not significant. Our experiments provided a second opportunity to test this hypothesis, which predicts a two-way interaction between capacity and congruence, such that low, but not high-capacity reasoners will perform more poorly for incongruent than congruent problems.

The third possibility is an instructional set hypothesis (Engle, Carullo, & Collins, 1991), in which cognitive capacity predicts the ability to maintain an instructional set and would therefore correlate with performance regardless of instructions or conflict. That is, high-capacity reasoners would be expected to perform better than low-capacity reasoners regardless of how they were instructed to respond.

Experiment 1: Logical Reasoning

Method

Participants. We recruited 112 University of Saskatchewan students (69% females, mean age = 21.4 years), who received partial course credit or CAN$7. Our goal was to test 100 participants, but to maintain parity with Experiment 2, where multiples of 16 were required to counterbalance the stimuli, we tested to the next highest multiple of 16, giving 112 participants. This number gave us good (.8) statistical power to detect moderate effects (η_p^2 = .07) for all main effects and interactions involving our two-level factors and good (.8) power to detect a slightly larger interaction with ability (η_p^2 = .09). All power calculations were computed using the MorePower calculator (Campbell & Thompson, 2012).[1]

We note that prior approval for the research was obtained from the University of Saskatchewan Research Behavioral Research Ethics Board.

Materials and procedure.

Reasoning problems. Each participant was presented with 64 syllogistic reasoning problems of moderate complexity (for examples, see Appendix A; the complete problem set is available in the online supplemental materials). Half of the problems were three-term categorical syllogisms (e.g., all A are B; all B are C; therefore, all A are C). All of these arguments, could, in principle, be solved by constructing a single mental model (e.g., Johnson-Laird, 2001); the invalid arguments featured determinately invalid conclusions. The remaining problems were disjunctive syllogisms. Within the set of disjunctive syllogisms half featured an affirmation inference (e.g., either A or B but not both; A; therefore not B) and the other half featured a denial inference (e.g., either A or B but not both; not A; therefore B).

We crossed logical validity with the believability of the arguments resulting in two types of congruent problems (valid-believable and invalid-unbelievable) and two types of incongruent problems (valid-unbelievable and invalid-believable). We randomly assigned problem contents to argument structures for each participant independently to ensure that any observed effects of logical validity could not be attributed to incidental confounds between the logical structure and the believability of the argument (cf., Clark, 1973; Trippas, Handley, & Verde, 2013, 2014; Trippas,

[1] Freely available at: https://wiki.usask.ca/pages/viewpageattachments.action?pageId=420413544.

Pennycook, Verde, & Handley, 2015). Examples of the arguments for each cell of the design can be found in Appendix A.

Individual Differences (ID) measures. We used four widely used measures to assess cognitive ability and analytic cognitive style: (a) The Shipley-2, a standardized intelligence test that includes verbal and abstract reasoning components (Shipley, Christian, Martin, & Klein, 2009); (b) A three-item test of numeracy (Lipkus, Samsa, & Rimer, 2001; Schwartz, Woloshin, Black, & Welch, 1997); (c) The Cognitive Reflection Test (CRT; Frederick, 2005), a set of three word problems that cue intuitive but incorrect responses and that measures the willingness to engage analytic reasoning to question a misleading intuition (e.g., Toplak et al., 2011), plus an additional 4 CRT items from Toplak, West, and Stanovich (2014); and (d) a thinking disposition questionnaire that included the 41 items from the actively open-minded thinking questionnaire (Stanovich & West, 2006).

Procedure. The participants were tested on individual computers in small groups. Upon entering the laboratory each participant was briefed about the study and requested to answer a short demographic questionnaire. The Shipley-2 was administered first, followed by the logical reasoning problems, the AOT, Numeracy, and CRT in that order. The instructions for the logical reasoning task were displayed on the computer screen and read:

In this experiment, we are interested in your ability to make two types of judgments: judgments on the basis of LOGIC, and judgments on the basis of BELIEFS

When the word "LOGIC" appears in red at the top of the screen, you should assume all the information ABOVE the line is true (even if it's not, or if it doesn't appear to make much sense).

After a short amount of time, a conclusion sentence BELOW the line will appear which you will be asked about. If you judge that the conclusion necessarily follows from the premises, you should answer "Valid" by pressing the "s"-key, otherwise you should answer "Invalid" by pressing the "k"-key.

For example:

All cars are blurbs

All blurbs are cheap

All cars are cheap

Given the instruction to respond on the basis of LOGIC, you should respond "Valid", because the sentence "All cars are cheap" necessarily follows from the premises above the line (if you assume they are true).

When the word "BELIEF" appears in red at the top of the screen, you should focus on whether the information is in line with your beliefs about what is true in the world. If you think the information BELOW the line is in line with you knowledge of the world, you should respond "Believable" by pressing the "s"-key. Otherwise, please respond "Unbelievable" by pressing the "k"-key

For example:

All cars are blurbs

All blurbs are cheap

All cars are cheap

Given the instruction to respond on the basis of BELIEF, you should respond "Unbelievable", because you presumably know from your experience of the world that the sentence "All cars are cheap" is false (consider, for instance, the cost of a Ferrari or a Porsche).

After reading the instructions, each participant was presented with 64 syllogistic reasoning problems. At the onset of a trial, the premises were presented. After 3 s, the conclusion followed. On half the trials the participant was cued with the word "LOGIC" at the top of the screen to signal that they had to respond according to the logical validity of the syllogism. On the remaining half of the trials the word "BELIEF" appeared at the top of the screen, cueing the participant to produce a response on the basis of their knowledge of what is true in the world. The stimuli were presented in a fully randomized order. It is important that the instructional manipulation was crossed with problem type and conflict to eliminate any potential confounds. On each trial we measured whether the participant responded affirmatively or negatively. After all trials were completed, the participants were thanked and debriefed.

Results

Principal components analysis of individual difference measures and acceptance rates. We adopted two strategies to test our hypotheses. Our first was a multivariate approach in which we subjected our eight dependent variables and four predictor variables (all normalized) to a principal component analysis. This served two functions. The first was as a preliminary test of our hypotheses. As argued previously, there is more than one way that individuals could differ. For example, some may comply better with instructions on all tasks, but others might do better on probabilistic judgments and worse on belief judgments and so on. The principal component analysis (PCA) will enable us to see if there is more than one factor of individual differences, as well as to interpret what each factor represents. Second, we were able to examine the correlations among our predictor variables to determine whether we could simplify the next set of analyses by replacing them with a composite variable.

Three components with eigenvalues greater than one were extracted, explaining 60% of the variance (Table 1). We first note that the categorical syllogisms and disjunctive syllogisms loaded together on the same components, suggesting that the effect of congruency and instructions were similar for the two problem types. In addition, the fact that all of our individual differences measures load on the same factor indicates that producing a composite score based on the four measures is a suitable option.

A more detailed look at the component structure indicated that the component loading for Components A and B are consistent with our hypothesis that high-capacity reasoners would be relatively better at resolving conflict in favor of validity than beliefs: As is clear in Table 1, Component A reflects high capacity and analytic thinking dispositions. It also reflects good performance on three of the four problem types; the exception is for problems that require conclusion validity to be inhibited in favor of belief (Incongruent-Belief), which were strongly associated with Component B. Clearly, these problems capture something unique about reasoning performance that is distinct from both the other types of problems and from cognitive capacity.

Components B and C offer a view of less successful reasoning. Component B is shows a low correlation with cognitive ability and thinking dispositions. This component reflects poor performance under logic instructions, at least when conflict is involved and good performance under belief instructions, again when conflict was involved. The pattern for Component C is slightly more

Table 1
Component Loadings for the Principal Components Analysis of the Syllogisms and Disjunctive Problems

Variable	Component A	Component B	Component C
Shipley (IQ)	.707	−.092	.138
Numeracy	.472	−.104	.288
Cognitive reflection test	.589	−.021	.442
Actively openminded Thinking	.481	.078	.300
Percentage correct			
Congruent			
Logic			
Syllogisms	.691	.073	−.381
Disjunctions	.648	−.134	−.286
Beliefs			
Syllogisms	.592	−.044	−.416
Disjunctions	.603	.021	−.552
Incongruent			
Logic			
Syllogisms	.712	−.262	.246
Disjunctions	.752	−.347	.209
Beliefs			
Syllogisms	.362	.860	.051
Disjunctions	.321	.860	.122

complex, showing modest, positive correlations with both analytic thinking dispositions and capacity. Performance on the congruent problems, which do not require conflict resolution, load negatively onto this component. In sum, the data suggest that high capacity as well as a disposition to think analytically is conducive to successful reasoning, except when belief-based responses must be inhibited in favor of logical ones. There were two distinct patterns of less successful performance: one that showed opposite patterns for resolving conflict under logic and belief instructions, and another showing poor performance on congruent items.

Factorial analysis of response accuracy. The preceding analysis indicated that computing a composite score for our individual differences measures would be justified. We did this by normalizing each of the four measures and computing their mean. For ease of illustration, the composite ability score was divided into quartiles and the data were then analyzed with a 4 (Ability Group) × 2 (Problem Type) × 2 (Instruction Condition) × 2 (Congruency) mixed analysis of variance (ANOVA). These data are plotted in Figure 1. To ensure that none of our results were artifacts of the way that we chose to divide up the groups, the crucial tests were repeated with the individual differences composite entered as a covariate in an analysis of covariance (ANCOVA).

As is clear from Figure 1, performance for the highest ability group is at ceiling, which means that the data from this group could artifactually create or suppress interaction effects. For this reason, the crucial interaction tests were repeated, using only the data from the first three groups. As would be expected, performance increased with ability, $F(3, 108) = 17.94$, $p < .001$, $\eta_p^2 = .33$. The main effects of instruction, congruency, and problem type were all significant, $F(1, 108) \geq 7.45$, $p \leq .007$, $\eta_p^2 \geq .065$ such that performance was better under logic than belief instructions (.85 vs. .80), for congruent than incongruent problems (.90 vs. .74) and for disjunctive than categorical syllogisms (.84 vs. .80). These main effects were qualified by several higher-order interactions. We will proceed by decomposing the highest-order interaction, which was the predicted three-way interaction between ability, instructions and congruency, $F(3, 108) = 4.71$, $p = .027$, $\eta_p^2 = .12$. This interaction was significant when the disjunctions and syllogisms were analyzed separately, meaning that it replicated across problem types, $F(3, 108) \geq 3.55$, $p \leq .017$, $\eta_p^2 \geq .09$, and when the highest ability group was removed from the analyses, meaning that it was not an artifact of the ceiling effect in that group, $F(2, 81) \geq 4.57$, $p \leq .013$, $\eta_p^2 \geq .10$. We also note that this interaction was reliable when the ability composite was entered into an ANCOVA as a continuous covariate, $F(1, 110) = 4.68$, $p = .033$, $\eta_p^2 = .04$ and when the highest ability group was removed from that analysis, $F(1, 82) = 10.77$, $p = .002$, $\eta_p^2 = .12$.

The form of the interaction is apparent in Figure 1: It is clear that the pattern of interference is quite different for the high and low ability groups. For the incongruent problems, there were main effects of instruction, $F(1, 108) = 6.84$, $p = .010$, $\eta_p^2 = .06$ and ability, $F(3, 108) = 15.46$, $p < .001$, $\eta_p^2 = .30$. More importantly, there was a cross-over interaction between ability and instruction, $F(3, 108) = 5.50$, $p = .001$, $\eta_p^2 = .13$, such that the lowest ability group performed better under belief than logic instructions, whereas the reverse was true for the higher ability groups. As is also clear from the figure, the increase in performance as a function of ability was more marked for logic than belief instructions. For the congruent problems, there was only a main effect of ability, such that performance increased with ability, $F(3, 108) = 8.11$, $p < .001$, $\eta_p^2 = .18$. Finally, we note that performance on incongruent problems was worse than congruent problems in each of the four ability quartiles; smallest $F(1, 27) = 12.89$, $p = .001$, $\eta_p^2 = .32$.

Before we discuss the implications of these findings, there were two significant interactions with problem type to report from the omnibus analysis: Problem type interacted with both instruction and congruency, $F(1, 108) \geq 5.0$, $p \leq .028$, $\eta_p^2 \geq .04$. The effect of congruency was larger for syllogisms than disjunctions (.19 vs. .13), as was the effect of instructions (.07 vs. .03).

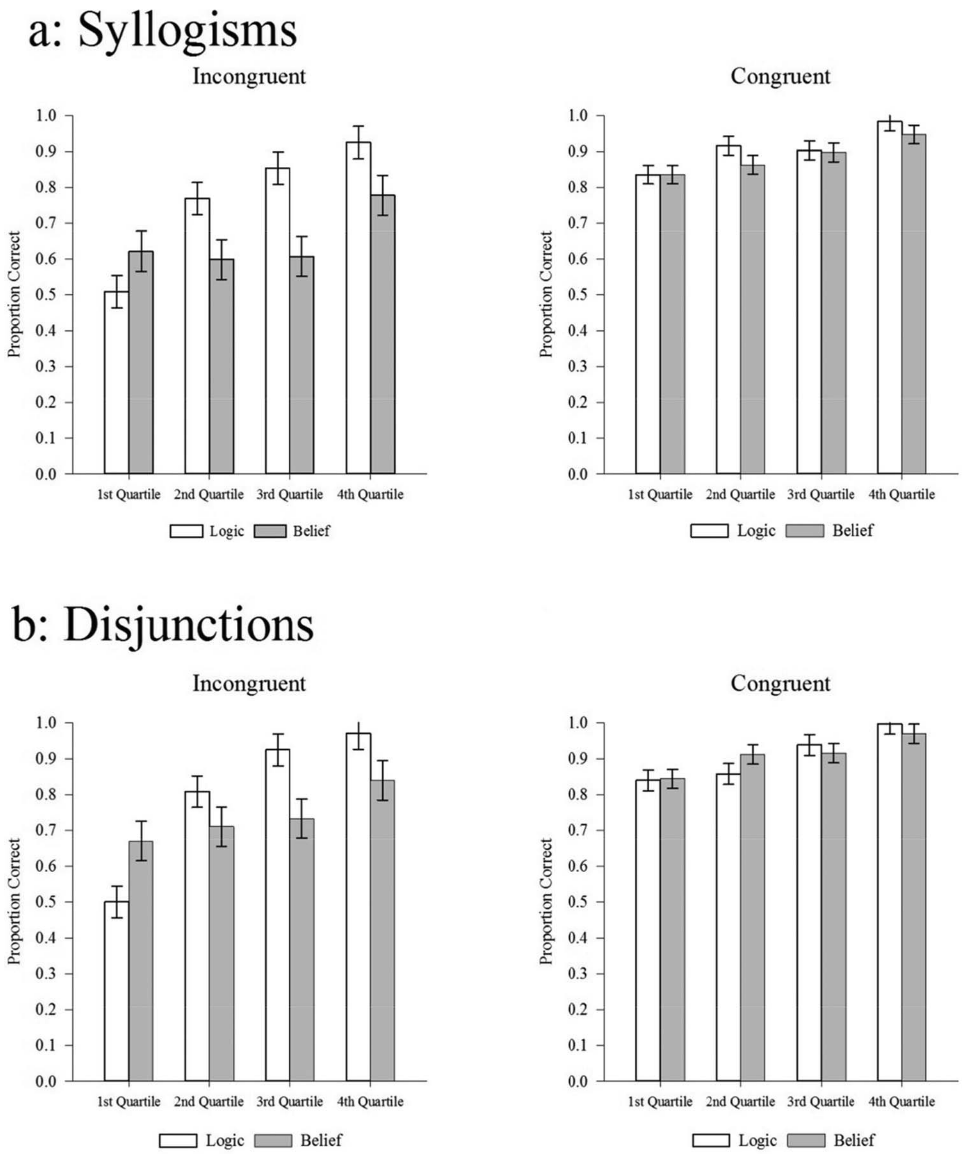

Figure 1. Experiment 1: Proportion correct on the syllogisms and disjunction problems as a function of congruence, instruction type, and cognitive ability. Error bars are *SE*s.

Discussion

These data provide a clear discrimination among the four hypotheses outlined earlier. The data are consistent with the hypothesis that the relative accessibility of logical intuitions varies as a function of ability (Thompson & Johnson, 2014). For higher-ability reasoners, logical validity interfered with their ability to make judgments based on belief, suggesting that the former was the dominant default response for this group. In contrast, for the lower-ability group, we noted that conclusion believability interfered more with judgments of validity than the reverse, suggesting that belief judgments form the default response for this group. By extension, the data support the conclusion that at least some of the relationship between measures of capacity and reasoning performance can be accounted for by differences in Type I processes.

Some of the findings are also consistent with a traditional dual process interpretation. As predicted by that view, higher-capacity reasoners were better at resolving conflict in favor of logical validity than lower-capacity reasoners. More difficult to explain is the fact that, overall, reasoners were better at responding on the basis of validity than belief. Given that belief-based responses should be available via autonomous Type I processes, all groups should quickly have access to belief-based responses, and all groups should therefore have fared well under belief-instructions, regardless of congruency. Instead, our data are consistent with recent evidence showing that logical judgments can happen very quickly (Newman et al., 2017); that even on complex logical reasoning tasks, logical validity interferes with belief-based judgments (Trippas et al., 2017); and with recent proposals that both logical and belief-based judgments can arise from Type I processes (De Neys, 2012; Pennycook, 2018; Pennycook, Fugelsang, and Koehler, 2015; Pennycook & Thompson, 2012; Trippas et al., 2016).

These data are not consistent with the hypothesis that high-capacity reasoners perform better than low-capacity reasoners because they do not experience conflict (as suggested by Svedholm-Häkkinen, 2015). Instead, performance in all groups was lower for incongruent than congruent problems. Finally, there was evidence to suggest that high-capacity reasoners performed better in all conditions, including the congruent problems, suggesting that part of the capacity-reasoning relationship may arise because high capacity reasoners are better at maintaining an instructional set (Engle et al., 1991).

Experiment 2: Base Rates

The goal of Experiment 2 was to replicate Experiment 1 using a different task. Here, reasoners solved base-rate problems with instructions to respond using statistics or beliefs. As before, half of the problems were incongruent and half were congruent problems. Our built-in replication was to add an additional group of participants who responded under a deadline in order to minimize the role of Type II processes. Strictly speaking, the logic of the Stroop task does not require a deadline to allow one to conclude that reading interferes with color-naming more than the reverse; however, given that our methodology is relatively new to the psychology of reasoning, we thought it prudent to replicate our main finding under conditions that maximize the probability that reasoners would rely on Type I processes. We also gathered confidence measures as an additional means to measure the effect of conflict on reasoning (De Neys, Cromheeke, & Osman, 2011; Shynkaruk & Thompson, 2006; Thompson & Johnson, 2014; Thompson, Prowse-Turner, & Pennycook, 2011; Thompson et al., 2013).

We note that prior approval for the research was obtained from the University of Saskatchewan Behavioral Research Ethics Board.

Method

Participants. We recruited 224 University of Saskatchewan students (69% females), who received partial course credit or CAN$7. Our goal was to test 100 participants per group, but because multiples of 16 were required to counterbalance the stimuli, we tested to the next highest multiple of 16, giving 112 participants per group. This number gave us good (.8) statistical power to detect small ($\eta_p^2 = .035$) for all main effects and interactions involving our two-level factors and good (.8) power to detect a slightly larger three-way interaction with ability ($\eta_p^2 = .05$).

Materials and procedure.

Base-rate problems. The materials and procedure were those used by Pennycook et al. (2014). Participants were given 24 base-rate problems (12 incongruent, 12 congruent; see Appendix B for examples), each providing the prior or base-rate probability of two categories of people as well as a personality sketch that contained stereotypes typical of one of the categories (the complete set of materials is available in the supplementary materials). These items had been pretested to ensure that the description was a good fit to the intended stereotype and was nondiagnostic of the complementary category (Pennycook et al., 2014).

The three elements of the problem were previewed to the participants prior to the presentation of the instruction to respond according to beliefs or statistics (Figure 2). The goal was to familiarize participants with the information but minimize the amount of time they had to integrate it: Participants saw a fixation cross, followed by the presentation of the sample information (e.g., In a study 1,000 people were tested. Jack is a randomly chosen participant of this study) for 4 s. This was followed by the description of the stereotype for 7 s, and the base rate information for 4 s. The order of these latter two pieces of information was randomized across trials. The entire problem then appeared on the screen until the participant entered a response. Participants were cued to respond on the basis of belief half the time and on the basis of statistics half the time via a “BELIEF” or “STATISTICS” prompt that appeared at the bottom of the screen; instructions were counterbalanced across all other factors. In the deadline condition, the screen turned red after five seconds indicating that an immediate response was required. Prior work showed that this deadline was challenging, but not impossible for participants to meet (Pennycook et al., 2014). We also asked for a rating of confidence on a seven-point Likert scale on each trial (after participants gave their response), with 1 marked as “*not at all confident*” and 7 marked as “*extremely confident.*”

Three base-rate probability ratios were presented equally often; 995/5, 996/4, 997/3. Participants were asked to estimate the probability that the person described belonged to one of the categories. Each description was presented to each participant in one of four equally occurring combinations, defined by whether the description and the base-rate were congruent (i.e., indicated the same response), and whether the smaller or larger of the categories was queried (Appendix B). Problem order was randomized for each participant.

Participants were instructed to indicate the likelihood that a randomly chosen participant belonged to one of the two groups. They were instructed that answering according to their beliefs meant answering according to their knowledge of what is true in the world. They were given an example concerning a person who is on the street and dressed in ragged clothing, asking for money. Answering on the basis of beliefs would mean that it is probable that the person is homeless. In contrast, they were told that when answering according to statistics, their prior beliefs about the world weren’t necessarily relevant and to concentrate on the actual probability that something will happen. In the case of the street-person, they were told that because only a small percentage of people in a city are homeless, they would give a low probability to the person being homeless.

Individual difference measures. The numeracy, IQ, and thinking disposition measures were the same as in Experiment 1. For the CRT, the original three items were used, rather than the extended test used in Experiment 1. The tasks were administered in the same order as in Experiment 1.

Results

Scoring. Responses that did not meet the deadline were coded as missing (27.1% of responses in the speeded condition); conse-

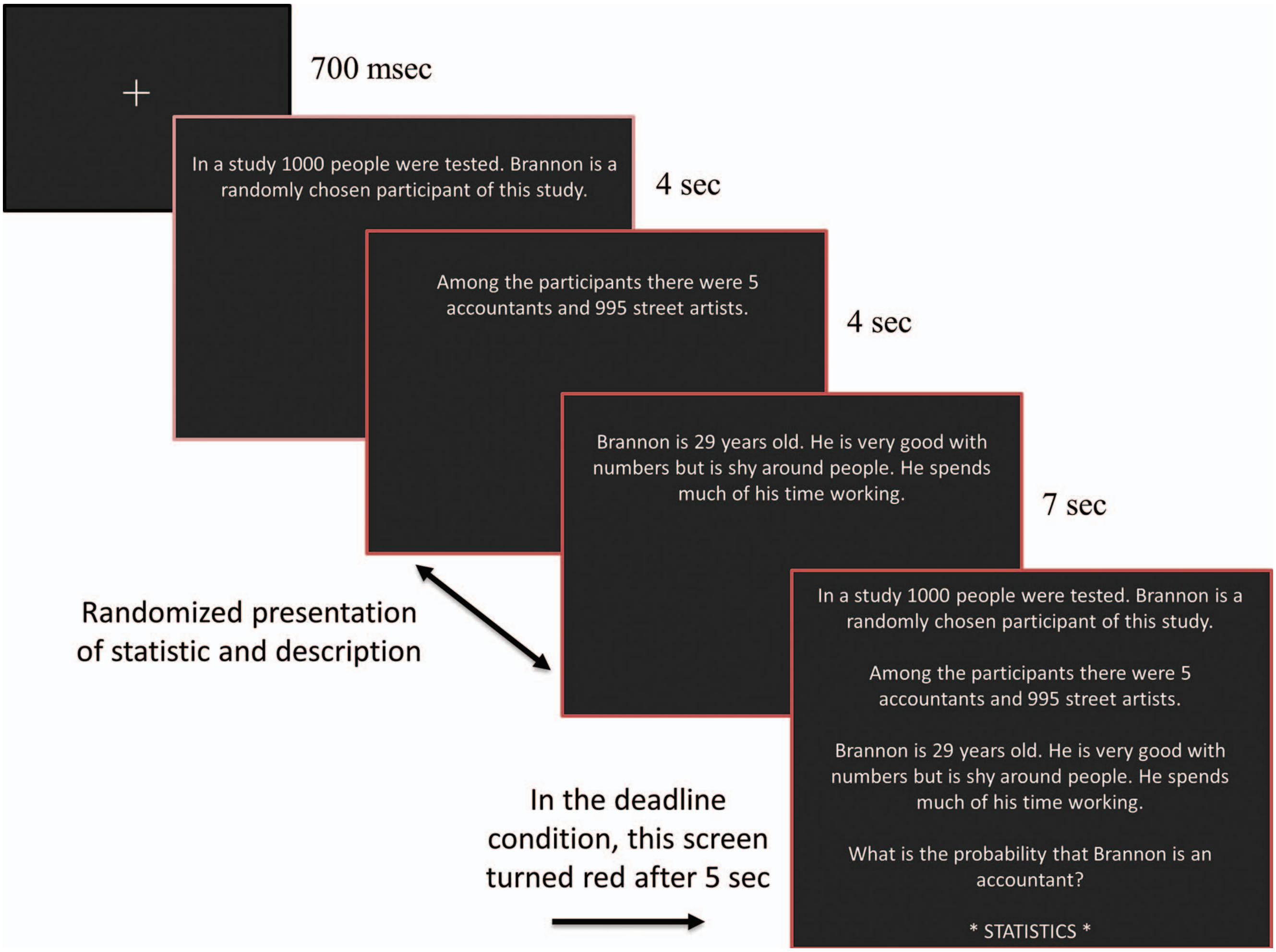

Figure 2. Sequence of events for a single trial in Experiment 2. See the online article for the color version of this figure.

quently, 15 participants were excluded who did not have observations in each of the congruency by instruction cells. We note that the fact that so many responses occurred outside the deadline means that it was challenging to meet and supports our goal to minimize the role of Type II processes in this condition. Three participants did not complete all of the individual differences measures. This still left us with good (.8) power to detect small effects ($\eta_p^2 = .038$) involving our two-level factors.

When the category asked about was the smaller one (Cells 2 and 4 in Appendix B), responses were subtracted from 100 so that high estimates always indicated compliance with the instructions (see Pennycook et al., 2014). For example, in Cell 4 of Appendix B, the correct response under both belief and statistics instructions would be a low probability; to make it comparable to responses in cell 1, the former were subtracted from 100. Overall, probability estimates were high ($M = 73$), indicating that, in general, participants were able to respond according to the instructions.

Principal components analysis of individual difference measures and probability estimates. As in Experiment 1, we adopted two strategies to test our hypotheses. Our first was a multivariate approach in which we subjected our four dependent variables and four predictor variables to a PCA. As was the case in Experiment 1, three components with eigenvalues greater than one were extracted, which were very similar in nature to those we observed in Experiment 1 (Table 2). Once again, the fact that all of our individual differences measures load on the same factor indicates that producing a composite score based on the four measures is a suitable option.

As in Experiment 1, the component loading for Component A reflects high capacity and analytic thinking dispositions. It also reflects good performance on three of the four problem types; the exception is for problems that require statistical information to be inhibited in favor of belief (Incongruent-Belief), which shows null or negative loadings onto this component. Instead, performance under the incongruent-belief instructions was strongly associated with Component B, which reflects relatively low to moderate capacity/thinking dispositions. These data replicate the findings from Experiment 1 and show that problems requiring reasoners to resolve conflict in favor of beliefs appear to tap processes that are distinct from both cognitive ability and the other problem types.

Table 2
Component Loadings for the Principal Components Analysis of the Baserate Responses

Variable	Component A	Component B	Component C
Shipley (IQ)	.707	.329	−.221
Numeracy	.528	−.056	−.116
Cognitive reflection test	.652	.220	−.007
Actively openminded thinking	.629	.074	−.572
Probability estimates			
Congruent			
Statistics	.414	−.253	.621
Beliefs	.392	.347	.585
Incongruent			
Statistics	.658	−.455	.118
Beliefs	−.154	.814	.184

Components B and C offer a view of less successful reasoning that shows some similarities to that observed in Experiment 1. Component B shows low to modest correlations with cognitive ability and thinking dispositions. This component is negatively correlated with reasoning under statistics instructions and positively correlated with reasoning under belief instructions, particularly for the incongruent problems. This component might be labeled "base-rate averse," as it seems to reflect a particular difficulty in reasoning under statistics instructions. Finally, Component C shows a similar pattern to Experiment 1, with opposite loadings for cognitive ability and performance on congruent problems. This factor suggests that motivation and ability are not required for success on problems that do not require conflict resolution, as would be expected if low ability participants do not attempt to reason statistically and rely on belief-based intuitions throughout. In sum, the data suggest that high capacity as well as a disposition to analytic thinking is conducive to successful reasoning, except when inhibition of statistical information is required. There were two distinct patterns of less successful performance: one reflects an inability to follow instructions to use the base rates, and another that reflects an inability to deal with conflict.

Factorial analysis of probability estimates. For our second set of analyses, we computed a composite score from the four individual differences measures (which was justified by the previous analysis, showing that all four loaded onto the same principal component). For this, each of the four measures was normalized separately for each deadline condition. As before, we divided the participants in each condition into quartiles, and as before, we repeated the critical tests with the individual differences composite entered as a continuous covariate in an ANCOVA. Our first analysis in this series was 2 (Instructions) × 2 (Congruency) × 2 (Deadline Condition) × 4 (Ability Group) mixed ANOVA of the probability estimates. The data are plotted in Figure 3. None of the higher-order interactions with the deadline factor were reliable, largest $F = 1.46$, $p = .24$ (recall that there was adequate [.8] power to detect small effects [$\eta_p^2 \geq .038$] involving this variable). The main effects of congruency and deadline condition were significant, $F(1, 201) \geq 8.00$, $p \leq .005$, $\eta_p^2 \geq .04$, with higher accuracy in the congruent ($M = 83$) than incongruent ($M = 64$) conditions and higher accuracy for the standard ($M = 75$) than speeded ($M = 71$) conditions. The main effect of ability group was also reliable, $F(3, 201) = 10.12$, $p < .001$, $\eta_p^2 = .13$, meaning that accuracy was higher overall for those with high ability and analytic thinking dispositions.

The instructions variable interacted with ability group, $F(3, 201) = 5.30$, $p = .002$, $\eta_p^2 = .07$; most critically, however, the predicted three-way interaction between congruency, instructions, and cognitive capacity was significant, $F(3, 201) = 8.18$, $p < .001$, $\eta_p^2 = .11$. This interaction was reliable when the composite cognitive ability measure was entered as a continuous covariate, $F(1, 206) = 7.86$, $p = .006$, $\eta_p^2 = .04$. The fact that the deadline variable did not interact with any of the other variables indicates that the patterns were similar in both the speeded and standard conditions; however, because it was critical to demonstrate that the effect held under time pressure, we computed the interaction for the speeded and unspeeded conditions separately; speeded: $F(3, 93) = 3.23$, $p = .026$, $\eta_p^2 = .09$, unspeeded: $F(3, 108) = 5.33$, $p = .002$, $\eta_p^2 = .13$.[2]

The data plotted in Figure 3 reinforced the conclusions drawn both from the PCA as well as from the data in Experiment 1 in favor of our hypothesis that for high capacity reasoners, responses based on probability are relatively more accessible than those based on beliefs. Statistical analysis verified the visual pattern, and replicated the pattern reported in Experiment 1. An additional 2 (Instructions) × 2 (Deadline Condition) × 4 (Ability Group) ANOVA revealed that for Incongruent problems, there was a cross-over interaction between ability and instructions, $F(3, 205) = 7.41$, $p < .001$, $\eta_p^2 = .10$. For the lowest ability group, performance was better under belief than ability instructions, whereas the reverse was true for the high ability group. The fact that statistical information was available quickly enough to interfere with belief-based judgments for our high-capacity group suggests that for these reasoners, statistical responses may be more intuitive than belief-based ones, whereas the opposite pattern held for the lower-capacity reasoners.

A similar ANOVA for the congruent problems indicated that neither the effect of instructions nor the interaction was reliable, $p \geq .12$. Once again, none of the interactions with the deadline condition were significant in either this or the previous analysis, largest $F(3, 205) = 1.39$, $p \geq .25$, $\eta_p^2 \leq .02$, suggesting that the effects replicated even under conditions meant to minimize Type II processes. As was the case in Experiment 1, the main effect of ability was reliable for both congruent and incongruent problems $F(3, 205) \geq 4.48$, $p \leq .004$, $\eta_p^2 \geq .06$. Finally, when data from the top and bottom quartiles was analyzed separately, we note that the effect of congruency was reliable in both cases, $F(1, 50) \geq 51.24$,

[2] We also ran the preceding analyses including those observations that did not meet the deadline in the Deadline condition. The three-way interaction was observed in first two analyses ($p \leq .001$), and it was marginally significant when the deadline condition was analyzed separately ($p = .059$).

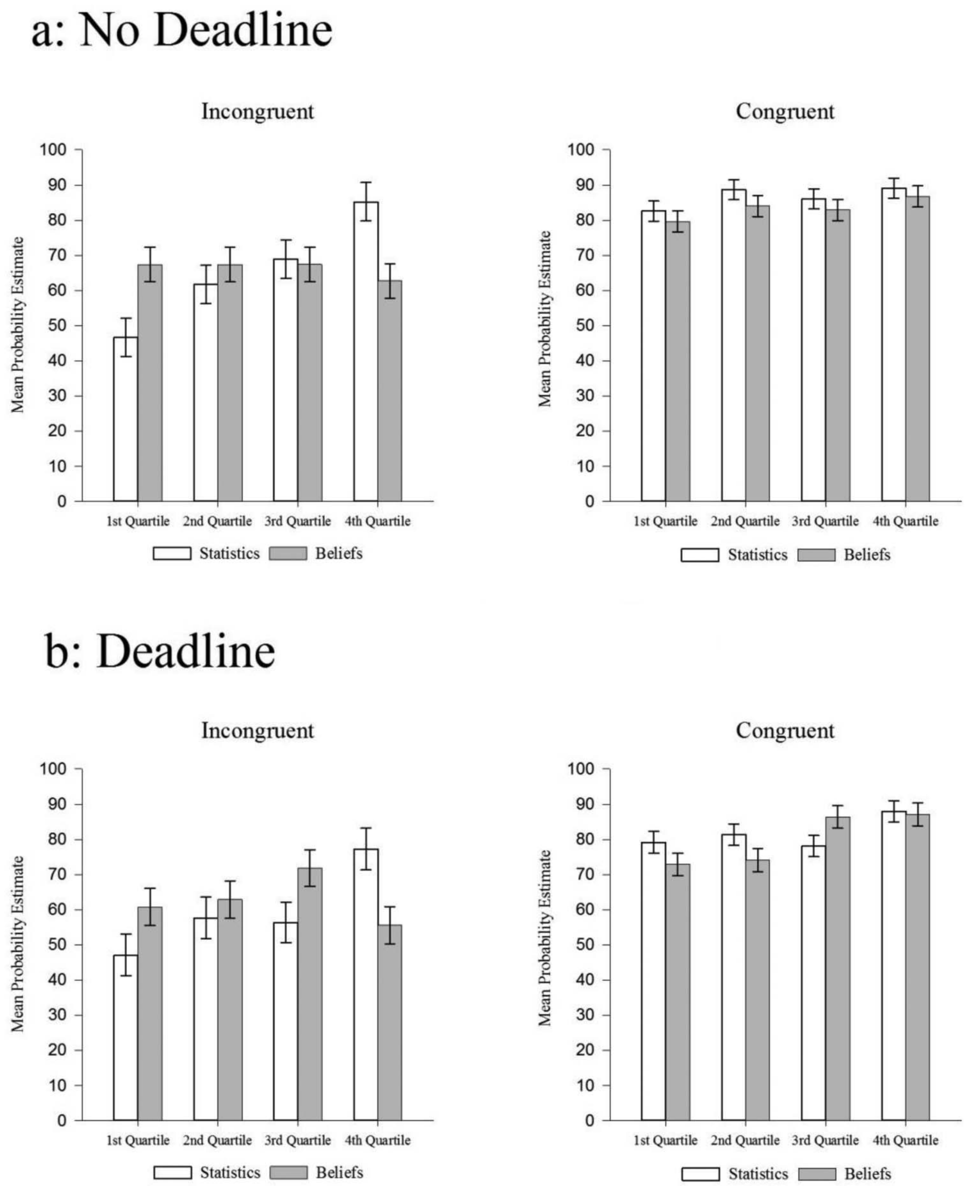

Figure 3. Experiment 2: Mean probability estimates for the base rate task as a function of congruence, instruction type, deadline condition, and cognitive ability. Error bars are *SE*s.

$p < .001$, $\eta_p^2 \geq .51$, meaning that both high- and low-ability reasoners experienced conflict.

Confirmatory analysis of confidence measures. As is clear from the Figure 4, the effects of our variables are less dramatic for confidence than for probability estimates. As such, the critical three-way interaction between congruency, instructions, and ability was marginally reliable when we used the grouping variable, $F(3, 201) = 2.12$, $p = .10$, $\eta_p^2 = .03$, but it was reliable when our ability composite was entered as a continuous covariate, with a small effect size, $F(1, 206) = 7.85$, $p = .006$, $\eta_p^2 = .04$. The pattern nonetheless maps onto the pattern observed with the probability estimates. For the incongruent problems, there is a cross-over interaction whereby lower ability participants are more confident under belief than statistics instructions, but this difference is reversed for the higher ability group, $F(3, 205) = 3.95$, $p = .009$, $\eta_p^2 = .06$. Again, this pattern is consistent with the hypothesis that for high ability people, responses based on probability interfere with responses based on belief, lowering confidence in belief-based responses. For the congruent problem, the interaction was not reliable, $F(1, 205) = 1.91$, $p = .13$, $\eta_p^2 = .03$, but the main effect of instructions was reliable, $F(1, 205) = 5.53$, $p = .02$, $\eta_p^2 = .03$, such that reasoners were more confident under statistics ($M = 5.72$) than belief instructions ($M = 5.59$).

Discussion

Reasoners with strong cognitive capacity and who have an analytic thinking disposition were better than their low capacity counterparts at resolving conflict in favor of probabilistic infor-

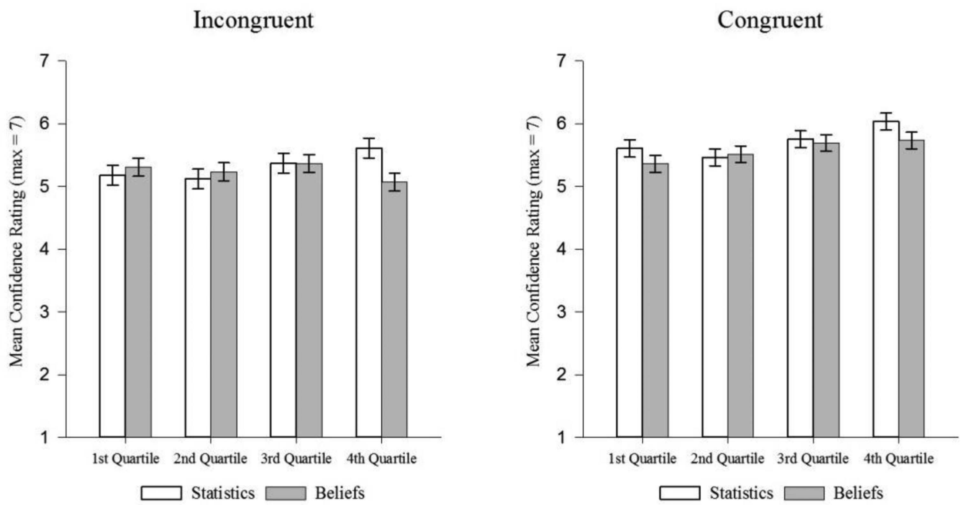

Figure 4. Experiment 2: Confidence judgments for the base rate task as a function of congruence, instruction type, and cognitive ability. Error bars are *SE*s.

mation. These data support the claim that capacity and analytic cognitive style are necessary to inhibit a prepotent belief in favor of one based on probability (Stanovich, 1999, 2009). However, as was the case in Experiment 1, the data do not support a straightforward Dual-Process interpretation: High-capacity reasoners were relatively poorer and less confident at resolving conflict in favor of beliefs than statistics. This finding supports the hypothesis that for high-capacity reasoners, responses based on probability are the default, and cause interference when they are asked to make belief-based judgments. It also converges with other findings showing that capacity differences in reasoning may emerge at an early stage in processing (Peters, 2012; Peters et al., 2008; Thompson & Johnson, 2014). In sum, the data suggest that at least part of the relationship between capacity and reasoning arises early in processing and may be because of difference in the nature of Type I outputs between higher- and lower-capacity reasoners.

Our PCA suggested that failures to succeed on the task are not unidimensional: It appears that one pattern is to perform well only on problems where there is no conflict at all, and the other is to experience difficulty following the instructions to respond on the basis of statistics, even when there is no conflict. These patterns of performance were associated with low to moderate capacity and analytic cognitive style and suggest that a more nuanced approach to understanding unsuccessful, as well as successful performance in reasoning is warranted.

An alternative interpretation of our findings might be that our instructions to reason solely on the basis of the base rates or statistics was confusing, given that the appropriate (Bayesian) response to the task should be to combine these sources of information. We concur that the normative approach to these problems would be to weight the statistical information in light of one's degree of belief in the description. In contrast, however, the evidence is clear that this is not how participants approach the task. For example, Pennycook and Thompson (2012) observed that, in the absence of instructions to reason according to beliefs or statistics, participants produce a markedly bimodal distribution of responses to conflict problems, with a large proportion of responses consistent with *either* the base rate *or* the description, and very few intermediate responses. This indicates that some people can be biased by the base rate—or, in other words, they display what might be called "stereotype neglect" and that the phenomenon known as "base-rate neglect" is produced by averaging over two distinct strategies. That is, rather than attempting to combine information, participants appear to choose which of the two responses is most credible (in cases of conflict; see Pennycook & Thompson, 2012).

Finally, we note that the effect of the deadline was to lower correct responses overall and that the deadline did not interact with any other variable or with cognitive ability. Of course, one must exercise caution when interpreting null findings. Nonetheless, given that we had a reasonably powered study, we can infer that the interaction effects with the deadline are, at best, small in size. There are several conclusions that might be drawn on this basis. First, it tells us that most (if not all) of the interference that we observed occurs because of Type I processing, given that it occurred under conditions designed to minimize Type II processing. This finding is consistent with recent hybrid Dual-Process models (e.g., De Neys, 2012, 2014; Handley & Trippas, 2015; Pennycook et al., 2015) that posit multiple, parallel Type I processes that may conflict prior to intervention by Type II processes. Here, however, the crucial finding is that the 'logical intuition' is the default for high cognitive ability individuals. Thus, not only are logical responses often intuitive, but for some people they are more intuitive (they arise earlier in the reasoning process) than stereotype or belief-based responding (canonical cases of intuitive processing). Note that the fact that high-capacity reasoners tend to have a dominant logical/probabilistic intuition does not imply that, for them, belief-based responding requires Type II processing. Instead, the reason that the deadline does not affect interference is likely because the competition between responses comes before Type II processes are engaged (see Pennycook et al., 2015 for an extended version of this argument).

Finally, we note that our findings may have implications for interpreting reasoning performance under dual task conditions.[3] Specifically, our data suggest that high-capacity reasoners should continue to perform well under load because they can rely on their logical or probabilistic intuitions. Instead, moderate and low-capacity reasoners should be especially disadvantaged by the load, because they are more likely to have to rely on WM capacity to overturn a belief-based response in order to derive a logical or probabilistic response. Although De Neys (2006) did not observe such differential effects, the trend was in the posited direction, suggestion that this might be a worthwhile avenue for further exploration.

General Discussion

In two large experiments, with three different types of problems and two different speed conditions, we have found a consistent pattern of differences between high- and low-capacity reasoners. Our tasks were akin to the Stroop task; answers based on validity/probability conflicted with answers based on belief, accompanied by an instructional set to ignore one dimension and to respond based on the other. High-ability reasoners repeatedly demonstrated more difficulty in resolving such conflicts in favor of beliefs than validity/probability, meaning that, for them, answers based on validity/probability produced more interference than answers based on belief. This interference occurred despite clear instructions to focus only on one dimension, and under conditions designed to minimize Type II engagement. This means that the interference likely occurred between two Type I outputs, and that, for the high-capacity group, answers based on probability/logic are likely the default, Type I response. In contrast, the reverse was true for low-capacity reasoners, who consistently behaved as expected by traditional Dual-Process Theories: belief-based judgments interfered with judgments based on validity/probability, consistent with the assumption that belief-based judgments form a default, Type I response for this group.

Two other findings were consistent across our studies. Both high and low-capacity reasoners did better on congruent than incongruent problems, which means that the observed superiority of high-capacity reasoners is not because of the fact that they do not experience conflict, as suggested by Svedholm-Häkkinen (2015). Second, we observed that high-capacity reasoners performed better than lower-capacity reasoners, even on congruent problems, which would not be expected if the sole role of capacity on these tasks was to resolve conflict. Instead, it suggests that high-capacity reasoners may be better at maintaining an instructional set, meaning that they do well regardless of the task (Engle et al., 1991).

Fundamentally, our data are challenging for the traditional Dual-Process explanation for the relationship between capacity and reasoning, which is a cornerstone of support for that view. A standard default-interventionist view of the relationship between capacity and accuracy is that high-capacity reasoners have the wherewithal to inhibit a prepotent, Type I response, create an alternative model of the problem, and perform the necessary computations to reach the correct answer (e.g., Stanovich, 1999, 2009). In other words, the relationship between capacity and reasoning is explained by Type II processes. Instead, our data supports Thompson and Johnson's (2014) contention that differences in capacity may arise, at least in part, from Type I processes. Thus, one reason that high-capacity reasoners give answers based on probability or logic may be that, for them, those are the first answers that come to mind. As such, at least some of the relationship between cognitive capacity and accuracy arises because high-capacity reasoners intuitively process numerical information and logical relationships, which a) allows them to quickly resolve conflict in favor of probabilities and validity, and b) means that they are relatively less efficient at resolving conflict in favor of beliefs. These findings challenge the common interpretation of the role of cognitive capacity in reasoning, which has focused on the role of Type II processes to explain why high-capacity reasoners are more successful on reasoning tests. Instead, it appears that some of that relationship can be explained by Type I processes.

We say "some" because, of course, our findings do not rule out a role for Type II processes in reasoning (as per Evans & Ball, 2010), nor in explaining the capacity–performance relationship. For example, given that cognitive capacity likely plays a role in the development of logical and probabilistic reasoning during adolescence, reasoning according to logic and statistics may become natural and intuitive for higher-capacity individuals though extended practice (Stanovich, West, & Toplak, 2011). In addition, it is perfectly plausible that cognitive capacity is needed to overturn a belief-based answer in favor of logic/probability, in the manner suggested by the traditional Dual-Process account. That is, those high-capacity reasoners who initially generated a belief-based response, as well as medium-capacity reasoners, may have the wherewithal to overturn the initial response in favor of logic/probability.

In other words, it is possible that cognitive capacity plays a role in the improvement that people show over time. In studies using a two response paradigm, where the first response is given under pressure and the second under free time, accuracy normally increases between the first and second responses (Bago & De Neys, 2017; Newman et al., 2017; Shynkaruk & Thompson, 2006; Thompson et al., 2011). A reasonable hypothesis is that this increase may be a function of cognitive capacity, that is, higher-capacity reasoners may be able to overturn an initial, belief-based response in order to derive a response based on logic or probability.

Our findings demonstrate that the relative impact of rule-based and belief-based information on judgment varies as a function of cognitive capacity: Belief based-information interferes relatively more for high- than low-capacity reasoners, whereas the reverse is true for high-capacity reasoners. The term "relatively" in this context is important, because high-capacity reasoners performed better overall, even on problems that did not require them to resolve conflict. However, one might have expected high ability people to do poorly under belief-instructions, especially under time pressure, given the high degree of interference they experienced from the statistical/logical information contained in the problems. Instead, performance on belief-based conflict problems was relatively constant across the ability spectrum. A parsimonious explanation for this pattern is one that is broadly consistent with dual-process explanations: the ability to access belief-based representations is relatively low-cost, with little variability across individuals. Access to computations based on logical structure or probabilities is more common to individuals with high capacity,

[3] We thank Wim De Neys for this observation.

either because they have previously practiced such computations to the point of automaticity, or because they are able to do the computations quickly and efficiently online.

Two other observations are important in helping us to understand the role that cognitive capacity plays in reasoning. First, we note that the effect of the deadline in Experiment 2 was constant across conditions. That is, performance was about 5% lower in the deadline condition, but not selectively lower for the conflict problems. For this reason, it appears that detecting and resolving conflict on these problems does not necessarily require capacity, in that the conflict problems were not especially affected by the deadline. Instead, the deadline must have affected some other aspect of reasoning performance. One possibility is that it might have affected reasoner's ability to comply with the instructions for all problems. Consistent with this view, we note that in both Experiments, high-capacity reasoners performed better even on congruent problems, where there is no conflict to resolve. Thus, it is possible that one effect of capacity is an enhanced ability (or motivation) to attend to and carry out the task instructions.

One possible concern with our findings is that we used relatively simple reasoning tasks. For example, the statistical information that was available in Experiment 2 was highly salient (e.g., 997 out of 1000), making it possible to be evaluated by Type I processes (Pennycook, Fugelsang, & Koehler, 2015). One might also argue that the reasoning tasks in Experiment 1 were of only moderate complexity, in that the overall rate of performance was high (.82) and the high-capacity group performed at ceiling. We note, however, that in both experiments, the low-capacity group was at floor in one condition (incongruent, statistics/logic-instructions), so that the tasks were not trivially easy. Nonetheless, there are likely to be clear boundary conditions for our findings. Other types of numerical and logical relationships will be more difficult to compute, making them poor candidates for Type I processes. Consequently, on more difficult reasoning tasks (e.g., three series multiple-model syllogisms) or in situations where reasoners need to compare ratios (rather than estimate magnitude), we would not necessarily expect even high-capacity reasoners to be able to produce Type I answers based on logic or probability. In those cases, we would expect much of the relationship between capacity and performance to derive from Type II processes.

Another possible interpretation of our findings is that they dilute the clarity of Dual-Process explanations for many reasoning phenomena, allowing for Type II responses to be delivered by Type I processes. However, this argument would be a typical example of the "normative fallacy" that equates correct answers with Type II reasoning and "biased" answers with Type I reasoning (Elqayam & Evans, 2011). Instead, ours and other's data (e.g., Handley et al., 2011; Pennycook et al., 2014) show that this straightforward mapping is experimentally, as well as theoretically untenable. We do, however, agree that these data demand a task-by-task analysis of how Type I and Type II processes may contribute to performance (see Handley & Trippas, 2015; Thompson & Newman, 2018). In addition, our data speak to the need for a more nuanced accounts of when individual differences are consequential in the reasoning process (De Neys & Bonnefon, 2013), which naturally require more nuanced models of the time course of Type I and Type II processes (Pennycook et al., 2015).

It is, however, clear that our data, along with others (De Neys, 2012, 2014; Handley & Trippas, 2015; Newman et al., 2017) are challenging for a straightforward interpretation of Dual-Process Theory. In fact, one might wonder at what point to draw the line and to reject the basic Dual-Process assumptions that reasoning is best characterized by a distinction between autonomous and WM-intensive processes. As an example of an alternative explanation that does not depend on the autonomy-WM distinction, it is possible that the interactions that we observed in the current experiments were between memory-based and rule-based processes[4]: memory-based processes give rise to belief-based judgments, whereas rule-based processes give rise to judgments of probability and logic. For low-ability participants, memory retrieval may be relatively easier than rule-based processes, and vice versa for high-ability participants. Although this explanation is compelling, we do not see an a priori reason to predict that belief-based judgments should be relatively more difficult than logic judgments for high-ability participants. Moreover, we are uneasy about the implied mapping of the Type I/Type II distinction onto memory versus rules, particularly given the evidence that both rule and memory-based processes can be fast and slow (Newman et al., 2017). Rule-based processes may arise from automated, compacted procedures, or may be implemented sequentially, requiring WM capacity (Kruglanski & Gigerenzer, 2011). Similarly, we can make retrievals from memory autonomously, in response to a stimulus, or make a deliberate search of memory for information relevant to the current context.

Another interpretation might be that performance generally increases with ability, but that performance under logic/probability instructions increases more than performance under belief-instructions. While this explanation is consistent with the pattern observed in the incongruent condition of Experiment 1, it does not hold overall. We note, for example, that performance on incongruent, belief-instructions did not increase as a function of ability in Experiment 2. Moreover, this explanation would predict an overadditive two-way interaction between capacity and instructions, in which the effect of capacity is positive for both belief and logic instructions, but stronger for the latter. Instead, we reliably observed a three-way interaction, in which the effect of capacity on performance for congruent items was similar under belief- and logic-instructions.

In sum, although the data challenge a strict, default interventionist interpretation of Dual Process Theories (Evans & Stanovich, 2013; Stanovich, 2009), our data fit nicely with current, parallel process models in which problems cue multiple Type I outputs (e.g., De Neys, 2012, 2014; Handley & Trippas, 2015; Pennycook et al., 2015). Moreover, the fact that predictions for the current study were grounded in those theories a priori attests to the continuing utility of the Dual Process Framework as a generator of research questions. In addition, as we have argued above, our data do not, in any way, rule out WM capacity as a basis for the reasoning-capacity relationship; instead, we have demonstrated that at least part of that relationship is likely attributable to Type I processes. As we have described above, there is still research to be done to ascertain whether the distinction between autonomous and WM based processes remains a viable basis for making experimental predictions (Thompson & Newman, 2018).

[4] We thank an anonymous reviewer for this suggestion.

Conclusions

In two experiments, in four different conditions, we have demonstrated that when responses based on beliefs and logic/probability conflict, high- and low-capacity reasoners demonstrate different patterns of interference. High-capacity reasoners show more interference when making belief- rather than logic-based responses, suggesting that responses based on logic/probability are the default response for this group. Low-capacity reasoners, instead, show the pattern anticipated by the received view of Dual Process Theories, wherein belief-based responses interfered more with logic-based judgments than vice versa. These data challenge the Dual Process explanation of the reasoning-capacity relationship and suggest that it may be in part because of Type I, rather than Type II processes. We also noted that there are likely to be limits to the ability of even high-capacity reasoners for "intuitive logic" and that there is much work to be done to identify the boundary conditions for this effect and to investigate the relative contributions of Type I and Type II processes in reasoning.

Context

Dual processes theories, which postulate the combination of slow, reflective thought with faster, intuitive thinking have been influential in many areas of psychology. However, dual process theories of reasoning have recently been challenged on a variety of fronts. These challenges appear to undermine the commonly accepted explanation for most reasoning biases, namely that fast, heuristic processes deliver default answers that may not be overturned deliberate, reflective processes (e.g., Kahneman, 2011). Some of these data come from our own laboratories, challenging us to rethink the basic assumptions of the framework. In that context, we decided to reexamine the relationship between cognitive capacity and reasoning biases, which is one of the evidentiary foundations of dual process theories of reasoning. It is widely believed that higher-capacity people resist cognitive biases because of superior reasoning ability, but recent findings of our own suggest that it may instead reflect better logical intuitions (Thompson & Johnson, 2014). On that basis, we made the counterintuitive prediction that high-capacity reasoners would experience relative difficulty overcoming answers based on logic and probabilities to make what should be the easier judgment of belief, and that is exactly what we observed in four separate experimental conditions.

References

Bago, B., & De Neys, W. (2017). Fast logic: Examining the time course assumptions of Dual Process Theory. *Cognition, 158,* 90–109.

Campbell, J. I., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods, 44,* 1255–1265. http://dx.doi.org/10.3758/s13428-012-0186-0

Clark, H. H. (1973). The language-as-fixed effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning & Verbal Behavior, 12,* 355–359. http://dx.doi.org/10.1016/S0022-5371(73)80014-3

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by g. *Intelligence, 32,* 277–296. http://dx.doi.org/10.1016/j.intell.2003.12.002

De Neys, W. (2006). Dual processes in reasoning: Two systems but one reasoner. *Psychological Science, 17,* 428–433.

De Neys, W. (2012). Bias and conflict a case for logical intuitions. *Perspectives on Psychological Science, 7,* 28–38. http://dx.doi.org/10.1177/1745691611429354

De Neys, W. (2014). 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., & Bonnefon, J.-F. (2013). The "whys" and "whens" of individual differences in thinking biases. *Trends in Cognitive Sciences, 17,* 172–178. http://dx.doi.org/10.1016/j.tics.2013.02.001

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

Elqayam, S., & Evans, J. St. B. T. (2011). Subtracting "ought" from "is": Descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences, 34,* 233–248. http://dx.doi.org/10.1017/S0140525X1100001X

Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *The Journal of Educational Research, 84,* 253–262. http://dx.doi.org/10.1080/00220671.1991.10886025

Evans, J. S. B., Handley, S. J., Neilens, H., & Over, D. (2010). The influence of cognitive ability and instructional set on causal conditional inference. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 63,* 892–909. http://dx.doi.org/10.1080/17470210903111821

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8,* 223–241. http://dx.doi.org/10.1177/1745691612460685

Evans, J. St. B. T. (2007). On the resolution of conflict in dual process theories. *Thinking & Reasoning, 13,* 321–339. http://dx.doi.org/10.1080/13546780601008825

Evans, J. St. B. T., & Ball, L. J. (2010). Do people reason on the Wason selection task? A new look at the data of Ball et al. (2003). *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 63,* 434–441. http://dx.doi.org/10.1080/17470210903398147

Evans, J. St. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning, 11,* 382–389. http://dx.doi.org/10.1080/13546780542000005

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19,* 25–42. http://dx.doi.org/10.1257/089533005775196732

Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 28–43. http://dx.doi.org/10.1037/a0021098

Handley, S. J., & Trippas, D. (2015). Dual processes and the interplay between knowledge and structure: A new parallel processing model. *Psychology of Learning and Motivation, 62,* 33–58. http://dx.doi.org/10.1016/bs.plm.2014.09.002

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences, 5,* 434–442. http://dx.doi.org/10.1016/S1364-6613(00)01751-4

Kahneman, D. (2011). *Thinking fast and slow*. Toronto, Ontario, Canada: Double Day Canada.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251. http://dx.doi.org/10.1037/h0034747

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131,* 66–71. http://dx.doi.org/10.1037/0033-2909.131.1.66

Klauer, K. C., & Singmann, H. (2013). Does logic feel good? Testing for intuitive detection of logicality in syllogistic reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1265–1273. http://dx.doi.org/10.1037/a0030530

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review, 118,* 97–109. http://dx.doi.org/10.1037/a0020762

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. http://dx.doi.org/10.1177/0272989X0102100105

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203. http://dx.doi.org/10.1037/0033-2909.109.2.163

Morsanyi, K., & Handley, S. J. (2012). Logic feels so good-I like it! Evidence for intuitive detection of logicality in syllogistic reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 596–616. http://dx.doi.org/10.1037/a0026099

Newman, I. R., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 1154–1170. http://dx.doi.org/10.1037/xlm0000372

Pennycook, G. (2018). A perspective on the theoretical foundation of dual-process models. In W. De Neys (Ed.), *Dual process theory 2.0* (pp. 3–27). Oxon, United Kingdom: Routledge.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology, 80,* 34–72. http://dx.doi.org/10.1016/j.cogpsych.2015.05.001

Pennycook, G., & Thompson, V. A. (2012). Reasoning with base rates is routine, relatively effortless, and context dependent. *Psychonomic Bulletin & Review, 19,* 528–534. http://dx.doi.org/10.3758/s13423-012-0249-3

Pennycook, G., Trippas, D., Handley, S. J., & Thompson, V. A. (2014). 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. (2012). Beyond comprehension the role of numeracy in judgments and decisions. *Current Directions in Psychological Science, 21,* 31–35. http://dx.doi.org/10.1177/0963721411429960

Peters, E., Slovic, P., Västfjäll, D., & Mertz, C. K. (2008). Intuitive numbers guide decisions. *Judgment and Decision Making, 3,* 619–635.

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine, 127,* 966–972. http://dx.doi.org/10.7326/0003-4819-127-11-199712010-00003

Shipley, W. C., Christian, P. G., Martin, T. A., & Klein, A. M. (2009). *Shipley-2*. Los Angeles, CA: Western Psychological Services.

Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory & Cognition, 34,* 619–632. http://dx.doi.org/10.3758/BF03193584

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

Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 55–88). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199230167.003.0003

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127,* 161–188.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23,* 645–726.

Stanovich, K. E., & West, R. F. (2006). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning, 13,* 225–247. http://dx.doi.org/10.1080/13546780600780796

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). The complexity of developmental predictions from dual process models. *Developmental Review, 31,* 103–118.

Svedholm-Häkkinen, A. M. (2015). Highly reflective reasoners show no signs of belief inhibition. *Acta Psychologica, 154,* 69–76. http://dx.doi.org/10.1016/j.actpsy.2014.11.008

Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning, 20,* 215–244. http://dx.doi.org/10.1080/13546783.2013.869763

Thompson, V. A., & Newman, I. R. (2018). Logical intuitions and other connundra for Dual Process Theories. In W. De Neys (Ed.), *Dual Process Theory 2.0* (pp. 121–136). Oxon, United Kingdom: Routledge.

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology, 63,* 107–140. http://dx.doi.org/10.1016/j.cogpsych.2011.06.001

Thompson, V. A., Turner, J. A., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition, 128,* 237–251. http://dx.doi.org/10.1016/j.cognition.2012.09.012

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition, 39,* 1275–1289. http://dx.doi.org/10.3758/s13421-011-0104-1

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Rational thinking and cognitive sophistication: Development, cognitive abilities, and thinking dispositions. *Developmental Psychology, 50,* 1037–1048. http://dx.doi.org/10.1037/a0034910

Trippas, D., & Handley, S. J. (2018). The Parallel Processing Model of Belief Bias: Review and extension. In W. De Neys (Ed.), *Dual process theory 2.0* (pp. 28–46). Oxon, United Kingdom: Routledge.

Trippas, D., Handley, S. J., & Verde, M. F. (2013). The SDT model of belief bias: Complexity, time, and cognitive ability mediate the effects of believability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1393–1402. http://dx.doi.org/10.1037/a0032398

Trippas, D., Handley, S. J., & Verde, M. F. (2014). Fluency and belief bias in deductive reasoning: New indices for old effects. *Frontiers in Psychology, 5,* 631. http://dx.doi.org/10.3389/fpsyg.2014.00631

Trippas, D., Handley, S. J., Verde, M. F., & Morsanyi, K. (2016). Logic brightens my day: Evidence for implicit sensitivity to logical validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 1448–1457. http://dx.doi.org/10.1037/xlm0000248

Trippas, D., Pennycook, G., Verde, M. F., & Handley, S. J. (2015). Better but still biased: Analytic cognitive style and belief bias. *Thinking & Reasoning, 21,* 431–445. http://dx.doi.org/10.1080/13546783.2015.1016450

Trippas, D., Thompson, V. A., & Handley, S. J. (2017). When fast logic meets slow belief: Evidence for a parallel-processing model of belief bias. *Memory & Cognition, 45,* 539–552. http://dx.doi.org/10.3758/s13421-016-0680-1

(*Appendices follow*)

Appendix A

Examples of Problems in Each Congruency by Logic Condition for Experiment 1

Logically valid	Logically invalid
[CONGRUENT: valid-believable]	[INCONGRUENT: invalid-believable]
Categorical syllogism:	*Categorical syllogism:*
All salmons are bunges	All salmons are bunges
No bunges are fruits	No bunges are fish
Therefore, no salmons are fruits	Therefore, some salmons are fish
Disjunctive syllogism (affirmation):	*Disjunctive syllogism (affirmation):*
Either beers are boats or they are drinks	Either beers are boats or they are drinks
Beers are drinks	Beers are boats
Therefore, beers are not boats	Therefore, beers are drinks
Disjunctive syllogism (denial):	*Disjunctive syllogism (denial):*
Either beers are boats or they are drinks	Either beers are boats or they are drinks
Beers are not boats	Beers are not drinks
Therefore, beers are drinks	Therefore, beers are not boats
Correct according to logic: accept	Correct according to logic: reject
Correct according to belief: accept	Correct according to belief: accept
[INCONGRUENT: valid-unbelievable]	[CONGRUENT: invalid-unbelievable]
Categorical syllogism:	*Categorical syllogism:*
All salmons are bunges	All salmons are bunges
No bunges are fish	No bunges are fruits
Therefore, no salmons are fish	Therefore, some salmons are fruits
Disjunctive syllogism (affirmation):	*Disjunctive syllogism (affirmation):*
Either beers are boats or they are drinks	Either beers are boats or they are drinks
Beers are boats	Beers are not boats
Therefore, beers are not drinks	Therefore, beers are not drinks
Disjunctive syllogism (denial):	*Disjunctive syllogism (denial):*
Either beers are boats or they are drinks	Either beers are boats or they are drinks
Beers are not drinks	Beers are drinks
Therefore, beers are boats	Therefore, beers are boats
Correct according to logic: accept	Correct according to logic: reject
Correct according to belief: reject	Correct according to belief: reject

Note. In the actual study, several other syllogistic structures were used. The examples serve only to demonstrate the manipulations.

(*Appendices continue*)

Appendix B

Examples of Problems in Each Congruency by Base-Rate Condition for Experiment 2

High base-rate	Low base-rate
[CONGRUENT, CELL #1] In a study, 1,000 people were tested. Brannon is a randomly chosen participant of this study. Among the participants there were 995 accountants and 5 street artists. Brannon is 29 years old. He is very good with numbers but is shy around people. He spends much of his time working. What is the probability that Brannon is an accountant? Correct according to statistics ~ 100 Correct according to beliefs ~ 100 Correct according to base-rate ~ 100	[INCONGRUENT, CELL #2] In a study, 1,000 people were tested. Brannon is a randomly chosen participant of this study. Among the participants there were 5 accountants and 995 street artists. Brannon is 29 years old. He is very good with numbers but is shy around people. He spends much of his time working. What is the probability that Brannon is an accountant? Correct according to statistics ~ 0 Correct according to beliefs ~ 100 Correct according to base-rate ~ 0
[INCONGRUENT, CELL #3] In a study, 1,000 people were tested. Brannon is a randomly chosen participant of this study. Among the participants there were 5 accountants and 995 street artists. Brannon is 29 years old. He is very good with numbers but is shy around people. He spends much of his time working. What is the probability that Brannon is a street artist? Correct according to statistics ~ 100 Correct according to beliefs ~ 0 Correct according to base-rate ~ 100	[CONGRUENT, CELL #4] In a study, 1,000 people were tested. Brannon is a randomly chosen participant of this study. Among the participants there were 995 accountants and 5 street artists. Brannon is 29 years old. He is very good with numbers but is shy around people. He spends much of his time working. What is the probability that Brannon is a street artist? Correct according to statistics ~ 0 Correct according to beliefs ~ 0 Correct according to base-rate ~ 0

Received February 8, 2017
Revision received March 18, 2018
Accepted April 10, 2018 ■