



Who resists belief-biased inferences? The role of individual differences in reasoning strategies, working memory, and attentional focus

Pier-Luc de Chantal¹ · Ian R. Newman² · Valerie Thompson² · Henry Markovits¹

Published online: 2 December 2019
© The Psychonomic Society, Inc. 2019

Abstract

A common explanation for individual differences in the ability to draw rule-based inferences, when a putative conclusion suggests a competing belief-based inference, is that the ability to do so depends on working memory capacity. In the following studies, we examined the hypothesis that the ability to draw rule-based inferences in belief bias tasks can also be explained by individual differences in reasoning strategies and in the related attentional focus. The dual-strategy model differentiates counterexample and statistical strategies that involve different information-processing styles. In the first study ($N = 139$), participants completed a working memory task (operation span), a strategy diagnostic questionnaire, and a belief bias task. The results showed that individual differences in strategy use predicted performance in the belief bias problems over and above any effects of working memory capacity, with counterexample reasoners producing rule-based inferences more often than statistical reasoners. In the second study ($N = 196$), an eye-tracking methodology was used as a process-tracing technique to investigate attentional differences between the two strategies. On problems showing a conflict between rule-based and belief-based information, counterexample reasoners demonstrated longer fixation times on the premises than did statistical reasoners, thus providing direct evidence that individual differences in strategy use reflect different processing styles. These results clearly indicate that individual differences in strategy use are an important determinant of the way that people make inferences when rule-based and belief-based cues are both present.

Keywords Reasoning strategies · Belief bias · Working memory · Attention · Individual differences · Eye tracking

Deductive reasoning consists of making dichotomous judgments of logical validity—that is, determining whether a conclusion necessarily follows from premises that are presumed to be true. Ideally, rule-based reasoning should reflect logical validity and not be sensitive to content-related variations. However, there is convincing evidence that the specific content used for reasoning produces variability in performance (e.g., Cummins, Lubart, Alksnis, & Rist, 1991; Markovits & Vachon, 1990; Thompson, 1994). One of the most well-studied ways that content affects reasoning is the *belief bias* effect, which occurs when belief in the empirical truth of a

putative conclusion conflicts with logical validity (e.g., Evans, Barston, & Pollard, 1983; Evans & Curtis-Holmes, 2005; Klauer, Musch, & Naumer, 2000; Markovits & Nantel, 1989; Oakhill, Johnson-Laird, & Garnham, 1989). Individual differences in the propensity to make rule-based or belief-based inferences are often explained by individual differences in cognitive capacity (De Neys, 2006; Quayle & Ball, 2000; Sá, West, & Stanovich, 1999; Stanovich & West, 2000; Tsujii & Watanabe, 2009). In the following studies, we provide contrary evidence that belief-biased reasoning reflects individual differences in the reasoning strategy used, as is postulated in the dual-strategy model, and that such differences involve different attentional foci.

Markovits and colleagues (Markovits, Brisson, & de Chantal, 2015, 2017a; Markovits, Brisson, de Chantal, & Thompson, 2017b; Markovits, Brunet, Thompson, & Brisson, 2013; Markovits, Lortie Forgues, & Brunet, 2012) have provided evidence that people have access to two types of inferential strategies when making judgments of logical validity, which map onto the basic mechanisms described by

✉ Pier-Luc de Chantal
pldechantal@gmail.com

¹ Department of Psychology, Université du Québec à Montréal, Montréal, QC, Canada

² Department of Psychology, University of Saskatchewan, Saskatoon, SK, Canada

probabilistic (e.g., Evans, Over, & Handley, 2005; Oaksford & Chater, 2007) versus mental model (Barrouillet, Gauffroy, & Lecas, 2008; Johnson-Laird, 2001; Johnson-Laird & Byrne, 1991, 1992; Markovits & Barrouillet, 2002) theories. This integrative model was built following the work initiated by Verschueren, Schaeken, and d'Ydewalle (2005a, 2005b). More specifically, the dual-strategy model proposes that people can either use a statistical or a counterexample strategy, with the former using the probabilistic parameters of the problem and the latter requiring the construction in working memory of an explicit model, which is then examined for the presence of counterexamples. According to this dual-strategy model, any analysis of reasoning outcomes must also consider individual differences in strategy use.

To date, several studies have shown clear correlations between strategy use and inferential performance. These have shown that under time constraint, statistical strategy use greatly increases (Markovits et al., 2013), and that the counterexample strategy is associated with better reasoning on both abstract inferences and explicitly probabilistic inferences (Markovits, Lortie Forgues, & Brunet, 2012). More specifically, Markovits, Brisson, de Chantal, and Thompson (2017b) showed that, when reasoning with inferences for which belief in a putative conclusion conflicted with logical validity, people using a counterexample strategy take longer to reason and are less influenced by conclusion belief than are people using a statistical strategy, although both counterexample and statistical reasoners are equally sensitive to the potential conflict in the problems. Now, one interpretation of these results is that counterexample reasoners are simply more “logical.” However, there is evidence that no direct connection exists between strategy use and “logical” reasoning, with each strategy showing different susceptibilities to information that can lead to both logical and nonlogical responding (Markovits et al., 2015).

Associated with this, some evidence suggests that the two reasoning strategies involve distinct information-processing styles. This is shown most clearly by the fact that effects of strategy use have been found in domains that are quite removed from traditional inferential problems. For example, a recent study showed that women's relative advantage in the processing of negative emotions more strongly occurs among statistical reasoners (Markovits, Trémolière, & Blanchette, 2018b). Similarly, another study showed that stereotypes are more easily activated by statistical than by counterexample reasoners (Gagnon-St-Pierre & Markovits, 2019). Taken together, these results suggest that strategy use represents a broad distinction in information processing (for similar conclusions using a different paradigm, see Newman & Thompson, 2019). These are consistent with the idea that statistical reasoners intuitively process a wider range of associative information, whereas counterexample reasoners are more likely to create a representation focused on a more limited set

of problem parameters. As we shall see, such differences in the way that information is processed provide a novel explanation for strategy differences in reconciling conflict between validity and conclusion belief.

Before we continue, it might be useful to distinguish between the dual-strategy model and dual-process theories that have been proposed as means to explain the joint existence of rule-based and belief-based reasoning. Although there are several versions of dual-process theories (e.g., De Neys, 2018; Evans, 2007; Pennycook, Fugelsang, & Koehler, 2015; Sloman, 1996; Stanovich & West, 2000), these generally distinguish two kinds of processes, referred to as Type 1 and Type 2 processes (it should be noted that these labels may change depending on the author). According to these models, inferences made by one reasoner are the result of the interplay between the two processes (e.g., De Neys, 2006). A key distinction between the two forms of processing is the cognitive cost associated with each. Type 2 processes are assumed to be working memory intensive, whereas Type 1 processes are conceived as autonomous and relatively independent of working memory (Evans & Stanovich, 2013). Research on dual-process models has typically relied on this asymmetry to explain the relative production of belief-based and rule-based responses (De Neys, 2006; Evans & Curtis-Holmes, 2005; for an overview, see Evans & Stanovich, 2013). As a corollary, individual differences in the belief bias effect are often considered to be primarily related to individual differences in measures of cognitive ability, such as working memory (e.g., De Neys, 2006; Quayle & Ball, 2000; Tsujii & Watanabe, 2009) or fluid intelligence (Sá et al., 1999; Stanovich & West, 1997, 2000), which underlie the extent to which more costly Type 2 processes can be deployed. However, there is some indication that cognitive capacity might not be as good a predictor as some dual-process theories might postulate. For example, Robison and Unsworth (2017) found that working memory capacity was related to overall performance on inferential problems but was unrelated to the ability to resolve conflicts between logical validity and conclusion belief (see also Morsanyi & Handley, 2008). Similarly, a recent review by Thompson and Newman (2019) suggests that for relatively simple inferences, the role of working memory capacity is less clear than in more complex forms of reasoning.

The dual-strategy model posits a very different explanation of individual differences in reasoning (whose relation to Type 1 and Type 2 processing remains an open question). In contrast to dual-process models, the dual-strategy model proposes that an inference is under the control of a single strategy at any given moment. In fact, several studies that have examined this model suggest that strategy use is a strong individual difference, with effects appearing on a variety of inferential tasks and noninferential judgments. As we described previously, a key difference between the two strategies has been postulated to be the range of information that is processed, with statistical

reasoners processing a wider range of information at a more intuitive level, whereas counterexample reasoners focus on a smaller range of information relevant to a given judgment.

Both differences in working memory capacity and differences in strategy use have been shown to be related to individual differences in reasoning about inferences for which rule-based and belief-based responses are in conflict. It could be argued that working memory differences underlie both effects, and that strategy use is a simple correlate of working memory. However, the attentional differences postulated by the dual-strategy model suggest that strategy use and working memory capacities modulate reasoning performance in different ways. The goal of the present studies is thus to examine the general hypothesis that strategy use is an individual difference whose effects differ from those related to working memory, with some part of this difference being related to the distribution of attention to the information underlying inferential problems.

Study 1

In this initial study, we examined relations between working memory, strategy use, and performance on a belief bias task. The participants in the present experiment were asked to evaluate conclusions to simple, three-term syllogisms, for half of which logical validity was in conflict with conclusion believability (conflict syllogisms). The other half were nonconflict problems, in which conclusion validity and believability agreed. In addition to accuracy on the reasoning problems, we also recorded response times, since this represents a complementary way of measuring the belief bias effect (e.g., Thompson, Striemer, Reikoff, Gunter, & Campbell, 2003). In addition to the belief bias task, participants were given a frequently used measure of working memory, the operation span task (see Unsworth, Heitz, Schrock, & Engle, 2005), and the strategy diagnostic (Markovits et al., 2012). We predicted that working memory capacity would be positively related to the level of production of logically valid responses in the belief bias task, but that differences in strategy use would explain variance in reasoning over and above that explained by differences in working memory capacity.

Method

Participants

A total of 139 students, between 18 and 49 years of age ($M_{\text{age}} = 26.5$, $SD = 6.12$; 68 men and 71 women) were recruited online through the Prolific academic database. The participants were volunteers drawn from a subject pool of students 18–55 years old, born in the United Kingdom or the United States.

Material

Strategy diagnostic We used the questionnaire developed by Markovits, Lortie Forgues, and Brunet (2012), which includes a set of 13 problems embedded in a fantasy context. The problems presented imaginary causal conditional relations “if P then Q” with nonsense terms and frequency information concerning the relative numbers of “P & Q” and “ \neg P & Q” cases out of 1,000 observations. Participants were then given an inference corresponding either to an Affirmation of the Consequent (AC; “if P then Q; Q is true, and therefore P is true”) or Modus Ponens (MP; “if P then Q; P is true, and therefore Q is true”), and they were asked to indicate whether or not the conclusion could be logically drawn from the information provided.

Of the 13 items, ten presented an AC inference, whereas three presented an MP inference. The MP inferences were included in order to provide variability in problem types and were not used to classify the inferential strategies. Of the ten items corresponding to the AC inference, five had a relative frequency of alternative antecedent cases (i.e., \neg P & Q) that was close to 10% of the observations (each individual item varied from 8% to 10%), and five had a relative frequency that was close to 50% (each individual item varied from 48% to 50%). In other words, five items had a relative probability that the conclusion was true of 90%, whereas the other five had a probability of 50%. Task instructions and examples of the problems can be found in the Appendix. Participants who consistently used a statistical strategy would consider that the conclusion was more likely in the 90% problems than in the 50% problems and would judge the 90% conclusions as being valid more often than the 50% conclusions. People who consistently used a counterexample strategy would reject all of these inferences, since they all indicated the presence of counterexamples to the conclusion.

Working memory task We created an online version of the operation span task (see Unsworth et al., 2005). A previous study had suggested that online administration of such a task is reliable (Hicks, Foster, & Engle, 2016). The operation span task requires participants to remember a series of letters while solving very simple math problems. Figure 1 shows an example with the response format. The original task was adapted for online assessment using Qualtrics. It was carefully reproduced according to the documentation provided by the Engle Lab (Attention & Working Memory Lab; englelab.gatech.edu) and the methodology developed by Unsworth et al. (2005). However, some minor modifications were made to the original version of the task, to accommodate online assessment. For instance, pretesting suggested that some participants might be using a facilitating strategy throughout the task. To account for this possibility while keeping the task used in this study as similar as possible to the original, we included an additional control trial on which noncheating participants could not succeed. A description of

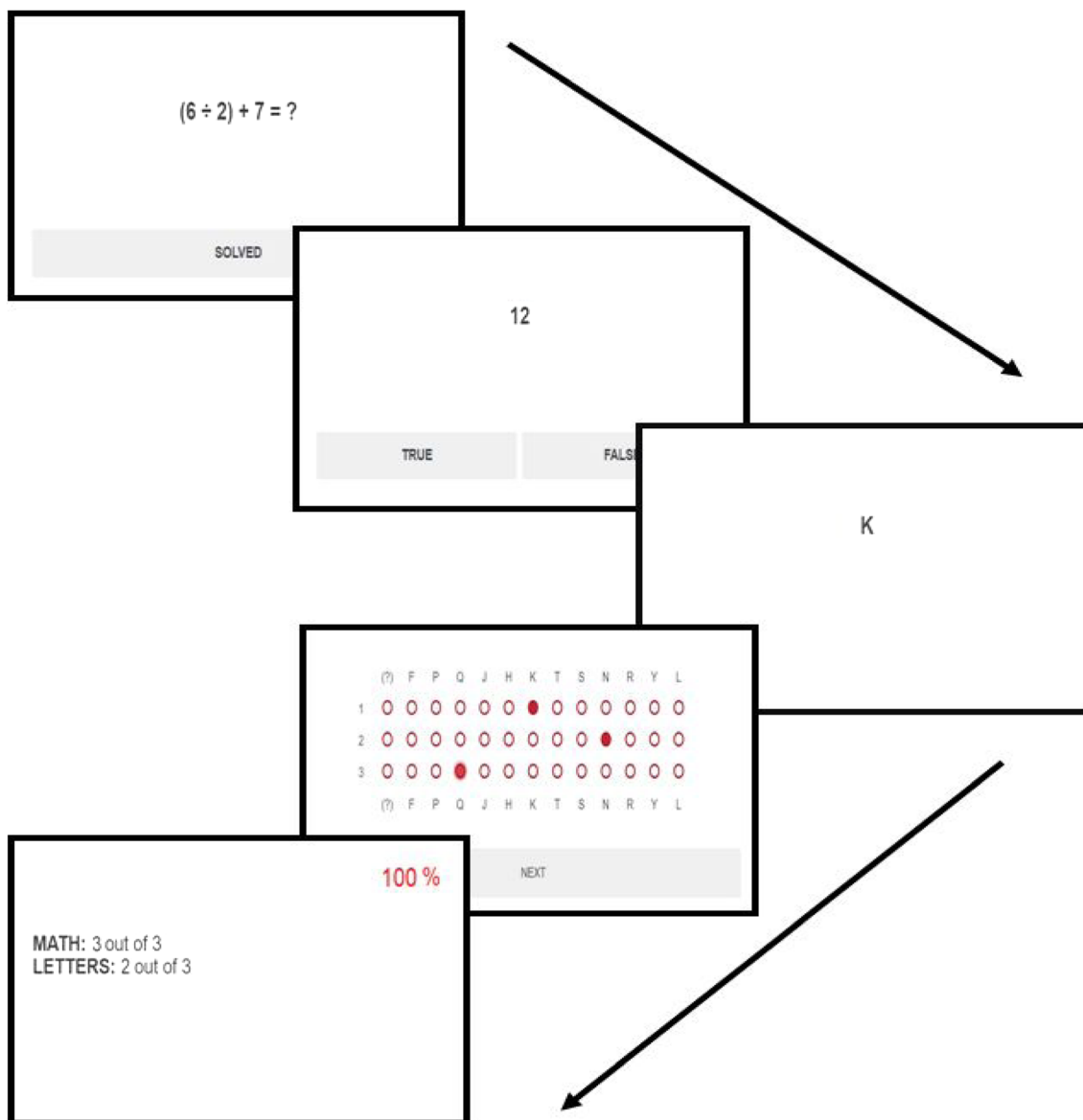


Fig. 1 Details of the operation span task with a response format for a set size of 3. In the task, a math problem is presented first. Participants must solve the problem and click the “solved” button within a prescribed time. A digit is then presented, and participants must indicate “true” or “false” to whether it is the solution of the problem. When a button is clicked, a letter appears for 800 ms. This procedure is repeated according to set size.

Only the first math–letter alternation is presented in this figure. The letters are then recalled using a matrix table, and feedback is given about performance for both the math problems and the recalled letters. The number in the upper right corner of the feedback display corresponds to the cumulative accuracy on the math problems.

the task and the adaptation process for online testing can be found in the [Appendix](#).

Belief bias task Belief bias was assessed with a syllogistic-reasoning task involving 12 problems administered in a random order. The three-term syllogisms were taken from Brisson, de Chantal, Lortie Forgues, and Markovits (2014, Study 3) and varied according to validity of the inference and conclusion belief. The valid syllogisms were equivalent to Modus Tollens inferences (MT; “All A are B, C is not B, therefore C is not A”), and the invalid syllogisms were equivalent to

Affirmation of the Consequent inferences (AC; “All A are B, C is B, therefore C is A”). The major premises were the same for both valid and invalid syllogisms, and the conclusions were designed to be either believable or unbelievable (for a full list of the syllogisms, see the [Appendix](#)). Response times were also recorded for each reasoning problem.

Procedure

The experiment was administered online. All participants were given the strategy diagnostic questionnaire first,

followed by the working memory task and the belief bias task. They were allowed to take a 5-min break between tasks. Finally, they were asked to provide demographic information (i.e., gender, age, and education level) and to indicate whether they had experienced distractions or technical issues during the tasks. It should be noted that no participants reported being distracted in a manner that would justify exclusion from the analysis. The experiment took 26.8 min on average ($SD = 19.1$). Participants were paid between £3 and £4.

Results

In the present study, participants for whom the difference between the numbers of accepted conclusions on 10% problems and 50% problems was greater than or equal to 2 were classified as using a statistical strategy ($N = 62$; 45% of the sample). Participants who rejected all conclusions on both types of problems were classified as using a counterexample strategy ($N = 43$; 31% of the sample). Those who did not adhere to one of these patterns were classified as using another strategy and were excluded from further analysis ($N = 34$; 24% of the sample). This dichotomous criterion for strategy use was chosen in order to ensure unambiguous categorization and has already been discussed previously (see Markovits et al., 2013).

We then analyzed the working memory task. We first excluded 28 participants with an average accuracy of less than 80% in the math problems (similar to Unsworth et al., 2005). The distributions of excluded participants for each strategy were not different, $\chi^2(1, N = 105) = 0.43, p = .51$. A total of 11 participants of the remaining sample perfectly recalled ten letters in the set size of 10, and those were also excluded from the subsequent analysis. Again, the distributions of excluded participants for each strategy were not different, $\chi^2(1, N = 77) = 0.22, p = .64$. We then computed a working memory score, following the absolute scoring method described in Unsworth et al. (2005), which consists of the sum of all perfectly recalled trials. The mean working memory score of the remaining 66 participants was 14.02 ($SD = 7.67$), out of a maximum possible of 25.

The following analyses were performed with 47.5% of the initial sample—that is, 66 participants ($M_{age} = 26.5, SD = 6.10$; 32 men and 34 women). It should be noted that all of the exclusion criteria were defined a priori. However, to ensure that these criteria did not lead to a selection bias that might influence our conclusions, we conducted complementary analyses with all participants in the sample ($n = 139$). To do so, no exclusion criteria were applied in the working memory task, and the reasoning strategy was measured on a continuous scale, to avoid eliminating patterns that did not fit any strategy. We decided to use the number of accepted conclusions in the 10% problems, given that it is the most discriminating factor on the theoretical level and that it highly correlated with the dichotomous classification ($r = .97, n = 66, p < .001$). A strategy score close to 0 should correspond to a

greater probability of using a counterexample strategy, whereas a score close to 5 should indicate a statistical strategy. The complementary analyzes showed identical results. We therefore decided to present the results using the dichotomous categorization, in order to facilitate interpretation in light of the dual-strategy model and previous studies.

Response accuracy We first analyzed accuracy in the belief bias task. We computed conflict and nonconflict scores, with the conflict score being the proportion of logically correct responses on valid–unbelievable and invalid–believable problems, and the nonconflict score referring to performance on valid–believable and invalid–unbelievable problems. Note that the correct answer to MT inferences is to judge the conclusion to be valid, whereas the correct answer to AC inferences is to judge the conclusion to be invalid. We then performed a 2 (Syllogism Type: conflict, nonconflict) \times 2 (Strategy: statistical, counterexample) repeated measures analysis of variance (ANOVA), with the proportion of correct answers as the dependent variable. Table 1 shows the proportions of correct answers as a function of syllogism type and strategy.

As expected, this produced significant main effects of syllogism type, $F(1, 64) = 61.34, p < .001, \eta_p^2 = .489$, and strategy, $F(1, 64) = 44.09, p < .001, \eta_p^2 = .408$. The proportion of correct answers was higher for nonconflict ($M = .84, SD = .18$) than for conflict ($M = .51, SD = .34$) syllogisms. Replicating past findings (Markovits, Brisson, de Chantal, & Thompson, 2017b), the overall proportion of logically correct answers was higher for counterexample ($M = .82, SD = .19$) than for statistical ($M = .56, SD = .14$) reasoners. These main effects were qualified by a Syllogism Type \times Strategy interaction, $F(1, 64) = 19.90, p < .001, \eta_p^2 = .237$. The interaction was decomposed using the Tukey test, with $\alpha = .05$. This showed that the proportion of correct answers on conflict syllogisms was higher for counterexample ($M = .76, SD = .27$) than for statistical ($M = .32, SD = .26$) reasoners. However, the differences in response accuracy were not statistically significant for nonconflict syllogisms. These results show the expected belief bias effect; syllogisms for which rule-based and belief-based components were in conflict led to lower levels of correct reasoning than did nonconflict syllogisms.

Table 1 Proportions of logically correct answers as a function of syllogism type (conflict, nonconflict) and strategy use (statistical, counterexample) in Study 1

Strategy		Syllogism Type		Total
		Conflict	Nonconflict	
Statistical	($n = 37$)	.32 (.26)	.80 (.19)	.56 (.14)
Counterexample	($n = 29$)	.76 (.27)	.89 (.17)	.82 (.19)
Total	($n = 66$)	.51 (.34)	.84 (.18)	.67 (.21)

Numbers in parentheses are standard deviations.

Critically, they also show that the differences in strategy use are most clearly illustrated by response accuracy on conflict syllogisms, with counterexample reasoners outperforming statistical reasoners.

We then computed a measure of the extent to which participants' responses were influenced by conclusion belief. Belief scores were calculated as the difference between the proportion of logically correct answers on nonconflict problems (i.e., valid–believable and invalid–unbelievable syllogisms) and that on conflict problems (i.e., valid–unbelievable and invalid–believable syllogisms). Higher belief scores indicated higher levels of acceptance or rejection of inferences based on the believability of the putative conclusions. This gave a mean belief score of .33 ($SD = .36$). We then examined the correlations between working memory score, strategy use, and belief score (see Table 2). Strategy was transformed into a dummy variable, with 0 for the statistical strategy and 1 for the counterexample strategy. As expected, the working memory score was negatively related to the belief score; this finding is consistent with past research showing a correlation between the ability to carry out rule-based reasoning and working memory capacity (e.g., Quayle & Ball, 2000). We also observed that the participants using the counterexample strategy had higher working memory capacity than did those using a statistical strategy; they also had lower levels of belief scores.

To examine the separate contributions of strategy use and working memory capacity to response accuracy in the belief bias task, we performed a regression analysis using the belief score as a dependent variable and working memory score and strategy as independent variables. The variance inflation factor ($VIF = 1.12$) indicated an absence of multicollinearity between strategy and working memory. This showed that the model explained a significant part of the variance in belief scores, $F(2, 63) = 12.08$, $p < .001$; working memory and strategy together explained 27.7% of the variance ($R = .53$). However, only strategy was a significant predictor ($\beta = -.42$, $t = -3.67$, $p < .001$), whereas working memory was only marginally significant ($\beta = -.21$, $t = -1.87$, $p = .07$). These findings support our hypothesis that strategy use predicts variability in the belief bias effect that is independent of working memory capacity. We also performed a hierarchical regression in which the working memory score was

entered into the model first, followed by strategy. This showed that 10.9% of the variance in belief scores is explained by the working memory score alone. Strategy explained a significant proportion of the remaining variance (15.5%) in belief scores, above that explained by the working memory score, $F(1, 63) = 13.48$, $p < .001$. We then performed a second hierarchical regression in which strategy was entered into the model first, followed by the working memory score. This showed that strategy alone explained 23.7% of the variation in belief scores, $F(1, 64) = 19.90$, $p < .001$, whereas working memory did not explain a significant proportion of the remaining variance, $F(1, 63) = 3.48$, $p = .07$.

Response time We also analyzed the mean response times in the belief bias task, without regard to accuracy. We performed a 2 (Syllogism Type: conflict, nonconflict) \times 2 (Strategy: statistical, counterexample) repeated measures ANOVA with response time (in seconds) as the dependent variable. Table 3 shows mean response times as a function of syllogism type and strategy. This analysis produced a main effect of syllogism type, $F(1, 64) = 5.52$, $p = .02$, $\eta_p^2 = .079$, as well as a main effect of strategy, $F(1, 64) = 7.29$, $p < .01$, $\eta_p^2 = .10$. Participants responded faster to nonconflict ($M = 6.92$, $SD = 3.08$) than to conflict ($M = 7.98$, $SD = 5.34$) syllogisms, thus showing the expected belief bias effect. Also consistent with prior findings (Markovits, Brisson, de Chantal, & Thompson, 2017b), statistical reasoners responded faster overall ($M = 6.36$, $SD = 2.95$) than counterexample reasoners ($M = 8.84$, $SD = 4.49$). Although the Syllogism Type \times Strategy interaction was not significant, $F(1, 64) = 2.44$, $p = .12$, $\eta_p^2 = .037$, we note that the difference in response times between the two reasoning strategies was greater for conflict ($\Delta = 3.25$ s) than for nonconflict ($\Delta = 1.72$ s) syllogisms.

We then examined the correlations between working memory score, strategy use, and response times for conflict and nonconflict problems (see Table 2). A key observation is that strategy use clearly correlates with response times for both conflict and nonconflict syllogisms, with counterexample reasoners being generally slower than statistical reasoners. On the other hand, individual differences in working memory capacity were not related to response times.

Table 2 Intercorrelations between strategy, working memory score (WM), belief score, and response times (in seconds) on conflict and nonconflict problems in Study 1

	WM	Belief Score	Response Times	
			Conflict	Nonconflict
Strategy	.33**	-.49**	.30*	.28*
WM		-.35**	.05	-.01

Strategy is dummy coded as 0 (statistical strategy) or 1 (counterexample strategy). * $p < .05$; ** $p < .01$.

Table 3 Mean response times (in seconds) as a function of syllogism type (conflict, nonconflict) and strategy (statistical, counterexample) in Study 1

Strategy		Syllogism Type		
		Conflict	Nonconflict	Total
Statistical	($n = 37$)	6.55 (3.94)	6.17 (2.28)	6.36 (2.95)
Counterexample	($n = 29$)	9.80 (6.33)	7.89 (3.70)	8.84 (4.49)
Total	($n = 66$)	7.98 (5.34)	6.92 (3.09)	7.45 (3.88)

Numbers in parentheses are standard deviations.

Discussion

The aim of this first study was to examine the relative contributions of individual differences in strategy use and working memory capacity to the belief bias effect. We found that although individual differences in working memory capacity and strategy use were both individually related to reasoning accuracy on belief-biased problems, as we hypothesized, strategy use predicted individual differences in the ability to make rule-based inferences above and beyond differences related to working memory capacity. In fact, the results suggest that strategy use is a stronger predictor of belief-biased reasoning than working memory capacity. The results also indicate that individual differences in working memory capacity are related to strategy use, with higher-capacity participants tending to use a counterexample strategy more often than a statistical strategy. However, although this difference did not explain the effect of strategy use on reasoning, it is consistent with the idea that the counterexample strategy requires more costly processing.

Our examination of response times further indicated that the effect of strategy use on reasoning reflects different information-processing styles and is not explained by differences in working memory capacity alone. Consistent with Robison and Unsworth (2017), differences in working memory capacity did not correlate with response times. On the other hand, statistical reasoners responded faster overall than counterexample reasoners. Moreover, counterexample and statistical reasoners both showed a relative increase in response times on conflict as compared to nonconflict syllogisms, which suggests that both groups processed the conflict between the rule-based and belief-based components of the problem. In line with past studies (Markovits, Brisson, de Chantal, & Thompson, 2017b), this suggests that both strategies were equally sensitive to the conflict between validity and conclusion belief. On the other hand, differences in working memory capacity did not show any relation to such sensitivity, and strategy use explained variability in response times beyond working memory, especially for conflict syllogisms. These results are consistent with the idea that individual differences in strategy use capture differences in the ways that information is processed that are not captured by measures related to cognitive capacity.

Study 2

The results of Study 1, together with the indirect results of previous studies (e.g., Markovits, Brisson, de Chantal, & Thompson, 2017b), strongly suggest that differences in strategy use not only are a product of the cognitive capacity required for each strategy, but are also related to the way that information is processed. Specifically, we have suggested that statistical reasoners process a wider variety of information associated with a given problem, whereas counterexample reasoners concentrate

on details needed to formulate a mental representation of the problem elements (Markovits, Trémolière, & Blanchette, 2018b). The goal of Study 2 was thus to investigate this hypothesis directly, by examining the information-processing patterns of the statistical and counterexample strategies in belief bias tasks. We decided to use a process-tracing methodology based on eye tracking (for a similar technique, see Ball, Phillips, Wade, & Quayle, 2006).

As we previously indicated, Markovits, Brisson, de Chantal, and Thompson (2017b) showed that counterexample reasoners make higher levels of rule-based inferences on conflict syllogisms. In addition, their results also showed that both counterexample and statistical reasoners are equally sensitive to potential conflicts between logical validity and conclusion belief, as shown by confidence ratings and the fact that both groups took more time to reason with conflict than with nonconflict problems. These results show that both counterexample and statistical reasoners consider conclusion believability in their inferences. Other than the overall decrease in the effect of belief found with counterexample reasoners, the only clear observable difference in the approaches taken by reasoners was the fact that counterexample reasoners increased the time to reason on conflict problems relatively more than did statistical reasoners (Markovits, Brisson, de Chantal, & Thompson, 2017b). This was also the case in the present Study 1, although the difference was only marginally significant.

One explanation for this pattern is suggested by results that have shown that statistical reasoners equally weight all sources of information, whereas counterexample reasoners are more strongly influenced by potential counterexamples to a putative conclusion (Markovits, Brisson, de Chantal, & Singmann, 2018a). In belief bias tasks, the logical structure of the premises and the believability of the conclusion represent two contradictory sources of information. When faced with such a contradiction, all reasoners devote more time to reconciling it. Since potential counterexamples in this case can only be generated by examining the premises, the importance given to these by counterexample reasoners suggests that they will concentrate additional attention on the premises. In contrast, under the assumption that statistical reasoners equally weight both the likelihood of the putative conclusion and that of the conclusion, they should distribute the extra time equally between premises and conclusion. This leads to the major hypothesis that, when faced with a conflict between conclusion belief and validity, counterexample reasoners will attend to the premises of syllogisms for longer than statistical reasoners.

In this study, we measured eye-gaze fixation times, response times, and confidence while reasoners performed a syllogistic reasoning task containing conflict, non-conflict, and neutral items (described below). The neutral problems were included so as to provide a baseline difference of visual attention between the statistical and counterexample strategies and to facilitate interpretation of differences in visual attention on conflict and/or

nonconflict syllogisms. We also elected to simplify our stimuli by only including invalid inferences in this study, since typically the belief bias effect on conflict problems is larger for invalid–believable problems than for valid–unbelievable problems (e.g., Stupple, Ball, Evans, & Kamal-Smith, 2011).

Method

Participants

A total of 196 participants ($M_{\text{age}} = 20.9$ years, $SD = 3.80$; 66 men and 130 women) from the University of Saskatchewan took part in the study for partial course credit or were compensated \$7.50 CDN.

Material

The task was presented on a Microsoft Windows laptop computer using a $1,600 \times 900$ resolution display with the E-Prime program (Psychology Software Tools, 2012). Participants were seated approximately 60 cm (± 5 cm) from the laptop computer screen on a stationary chair; no chinrest was used. An SMI RED-m remote eye tracker (SensoMotoric Instruments, 2017; <http://www.smivision.com>) was used, which allowed for user mobility and minimized intrusiveness (Mele & Federici, 2012). The eye tracker measured the frequency of eye-gaze fixations to each area of interest (AOI). Three rectangular AOIs were drawn around the three sentences in each problem: the first and second premises and the conclusion. Each AOI was equal in size and occupied 13% of the screen (39% in total). The threshold for fixation was set at 80 ms of sustained gaze fixation on a single position with a fixation radius of 1 deg of dispersion (corresponding to approximately 1 cm on the computer screen).

Strategy diagnostic This was the same as in Study 1.

Belief bias task Belief bias was assessed with a syllogistic reasoning task involving 15 problems administered in a random order. All problems had the same logical structure, which led to invalid conclusions that varied in terms of their believability. Five had a believable conclusion (conflict), five had an unbelievable conclusion (nonconflict), and five were neutral problems. The believable and unbelievable syllogisms were taken from Markovits, Brisson, de Chantal, and Thompson (2017b, Study 3). The neutral syllogisms were designed as control problems with content that cued neither strong agreement nor disagreement with the conclusion. For a complete list of the syllogisms, see the [Appendix](#).

The three-term syllogism problems were presented as bitmap images; the background was black with white text (18-point Courier). The two premises and the conclusion were presented concurrently, horizontally centered, and separated by four vertical lines of blank space. Below the problem text, response

options ($V = \textit{conclusion valid}$, $I = \textit{conclusion invalid}$) were displayed in the bottom left and right of the screen, respectively. For each problem, reasoners reported a feeling of rightness (FOR; Thompson, Prowse-Turner, & Pennycook, 2011) for their response to that problem. The responses were entered on a nine-point Likert scale (1 = *not right at all*, 9 = *extremely right*).

Procedure

Participants began the belief bias task with a nine-point calibration of the eye tracker and concluded when the (x, y) coordinate accuracy for both left and right eyes was within 1 deg of the horizontal and vertical deviation from the target. Calibration trials would repeat if deviation was not maintained within the acceptable range, up to ten attempts, at which point the task would proceed without collecting eye-tracking data (but these participants were excluded from the analysis).

Each trial of the belief bias task began with a blank screen (black background) presented for 2 s, followed by the syllogism problem. Participants were instructed to indicate whether the conclusion of each syllogism logically followed from the premises; the conclusions were rated as valid (press “V”) or invalid (press “I”) with a keyboard button press. After each syllogism response, confidence in that response was entered on a separate screen.

Following the belief bias task, reasoners were instructed that the eye-tracking portion of the experiment had ended, and then the strategy diagnostic commenced. The diagnostic problems always followed the syllogism task. Reasoners evaluated conclusion validity by either accepting (press “Y”) or rejecting (press “N”) the conclusion.

Results

For 15 participants, the eye tracker did not record any data; of these, five participants did not successfully complete the calibration, and the remaining ten were able to calibrate but the eye tracker failed to capture their visual gaze during the experimental task. Seven additional participants had insufficient eye-tracking data recorded throughout the task (mean fixation times per trial of less than 50 ms). These 22 participants were excluded from the analyses.

We used the same classification scheme as in Study 1, which led to 74 participants using a statistical strategy (43% of the sample) and 39 using a counterexample strategy (22% of the sample). The remaining participants ($N = 61$, 35% of the sample) were excluded from the analyses, leaving 113 participants ($M_{\text{age}} = 21.3$ years, $SD = 4.20$; 33 men and 80 women). Again, it should be noted that complementary analyses using the continuous scale for strategy assessment led to identical patterns of results.

Response accuracy We predicted a typical belief bias effect, in which more correct responses would be given on nonconflict (i.e., unbelievable) than on conflict (i.e., believable) problems.

Table 4 shows the proportions of correct answers as a function of syllogism type and strategy. Overall, the accuracy for nonconflict problems was nearly perfect ($M = .97, SD = .11$), and considerably better than the accuracy for conflict problems ($M = .42, SD = .34$) [$t(112) = 17.1, p < .001, d = 1.61$]. The counterexample reasoners made no errors on the nonconflict problems (their mean accuracy was 1 and variance was 0), so statistical comparison of accuracies on nonconflict problems between the strategy groups was not possible.

To analyze the conflict and neutral conditions, a 2 (Syllogism Type: conflict, neutral) \times 2 (Strategy: statistical, counterexample) repeated measures ANOVA was computed. The effect of syllogism type was significant, $F(1, 111) = 45.48, p < .001, \eta_p^2 = 0.29$. The proportion of correct responses was higher for neutral problems ($M = .72, SD = .25$) than for conflict problems ($M = .42, SD = .34$). An effect of strategy, $F(1, 111) = 21.83, p < .001, \eta_p^2 = .16$, was qualified by a Syllogism Type \times Strategy interaction, $F(1, 111) = 9.92, p = .002, \eta_p^2 = .08$. Counterexample reasoners had a higher proportion of correct responses on conflict problems than did statistical reasoners [Welch’s $t(61.0) = 4.6, p < .001, d = 1.0$], but the strategy groups did not differ in correct responses for the neutral conclusions [Welch’s $t(102.2) = 1.6, p = .124, d = 0.3$]. This result is consistent with Study 1 and replicated previous research (Markovits, Brisson, de Chantal, & Thompson, 2017b), in which counterexample reasoners were more likely to reason correctly to conflict problems than were statistical reasoners.

Response time Response times (reported in seconds) were analyzed with a 3 (Syllogism Type: conflict, nonconflict, neutral) \times 2 (Strategy: statistical, counterexample) repeated measures ANOVA. Table 5 shows response times as a function of syllogism type and strategy. This showed an effect of syllogism type, $F(1.8, 200.0) = 179.4, p < .001, \eta_p^2 = .62$, which revealed that response times were longer for conflict syllogisms ($M = 12.8, SD = 4.6$) than for nonconflict syllogisms ($M = 7.0, SD = 2.5$) [$t(222) = 18.4, p_{\text{Tukey}} < .001, d = 1.7$]. Overall, these response times are consistent with past research (e.g., Markovits, Brisson, de Chantal, & Thompson, 2017b; Robison & Unsworth, 2017; Stuppel et al., 2011).

Table 4 Proportions of logically correct answers as a function of syllogism type (conflict, nonconflict, neutral) and strategy (statistical, counterexample) in Study 2

Strategy		Syllogism Type			Total
		Conflict	Nonconflict	Neutral	
Statistical	(<i>n</i> = 74)	.31 (.28)	.96 (.13)	.69 (.28)	.66 (.14)
Counterexample	(<i>n</i> = 39)	.63 (.37)	1.0 (.00)	.76 (.19)	.80 (.15)
Total	(<i>n</i> = 113)	.42 (.34)	.97 (.11)	.72 (.25)	.70 (.16)

Numbers in parentheses are standard deviations.

Furthermore, the response times for neutral problems ($M = 11.4, SD = 3.3$) were longer than those for nonconflict syllogisms [$t(222) = 13.1, p_{\text{Tukey}} < .001, d = 1.2$], but shorter than those for conflict syllogisms [$t(222) = 5.4, p_{\text{Tukey}} < .001, d = 0.5$].

We expected that counterexample reasoners would spend more time responding to the conflict syllogisms than statistical reasoners, consistent with the previous results. An effect of strategy, $F(1, 111) = 8.2, p = .005, \eta_p^2 = .07$, was qualified by a Syllogism Type \times Strategy interaction, $F(1.8, 200.0) = 9.2, p < .001, \eta_p^2 = .08$. The counterexample reasoners had longer response times on the conflict problems than did the statistical reasoners [Welch’s $t(55.4) = 3.4, p = .001, d = 0.7$; see Table 5]. Response times did not differ between the strategies for the nonconflict [$t(111) = 1.3, p = .192, d = 0.3$] or the neutral [$t(111) = 1.4, p = .161, d = 0.3$] syllogisms. The counterexample reasoners spent more time reasoning than the statistical reasoners, but only on the conflict syllogisms.

Fixation time Fixation times (in seconds) were recorded as the total fixation time in each AOI per trial; the threshold for a fixation was 80 ms. For fixation times (reported in seconds), we computed a 3 (Syllogism Type: conflict, nonconflict, neutral) \times 3 (AOI: Premise 1, Premise 2, conclusion) \times 2 (Strategy: statistical, counterexample) repeated measures ANOVA (see Table 6). An effect of syllogism type, $F(2, 222) = 112.7, p < .001, \eta_p^2 = .5$, was qualified by a Syllogism Type \times Strategy interaction, $F(2, 222) = 5.3, p = .006, \eta_p^2 = .05$. Post-hoc *t* tests revealed that overall, the counterexample reasoners had longer fixation times on the conflict problems than did the statistical reasoners [$t(111) = 2.65, p = .015, d = 0.5$], but the strategy groups did not differ on fixation times for either the nonconflict [$t(111) = 0.8, p = .402, d = 0.2$] or the neutral [$t(111) = 1.2, p = .225, d = 0.2$] problems. Thus, the difference in visual attention between the strategies manifests on the conflict syllogisms only, which is the same pattern of results as for response times.

We then analyzed fixation times according to the three different AOIs. We found an effect of AOI, $F(1.7, 194) = 19.0, p < .001, \eta_p^2 = .1$, and an AOI \times Syllogism Type interaction, $F(3.4, 377.6) = 47.7, p < .001, \eta_p^2 = .3$, which were qualified by an AOI \times Syllogism Type \times Strategy interaction, $F(3.4, 377.6) = 3.2, p = .018, \eta_p^2 = .3$. We analyzed the latter interaction by computing a separate 3 (Syllogism Type: conflict, nonconflict, neutral) \times 2 (Strategy: statistical, counterexample) repeated measures ANOVA for each AOI (Premise 1, Premise 2, and conclusion; see Fig. 2). We predicted that the counterexample reasoners would fixate on both Premise 1 and Premise 2 for longer than the statistical reasoners, particularly for the conflict syllogisms.

The premises were initially analyzed together, producing an effect of syllogism type, $F(2, 222) = 99.0, p < .001, \eta_p^2 = .47$, and an effect of strategy, $F(1, 111) = 3.94, p = .05, \eta_p^2 = .03$, which were qualified by a Syllogism Type \times Strategy

Table 5 Mean response times (in seconds) as a function of syllogism type (conflict, nonconflict, neutral) and strategy (statistical, counterexample) in Study 2

Strategy		Syllogism Type			Total
		Conflict	Nonconflict	Neutral	
Statistical	(<i>n</i> = 74)	11.70 (3.56)	6.80 (2.13)	11.06 (3.08)	9.85 (2.33)
Counterexample	(<i>n</i> = 39)	14.97 (5.48)	7.45 (3.07)	11.99 (3.77)	11.47 (3.64)
Total	(<i>n</i> = 113)	12.8 (0.46)	7.03 (0.25)	11.4 (0.33)	10.41 (2.94)

Numbers in parentheses are standard deviations.

interaction, $F(2, 222) = 6.49, p = .002, \eta_p^2 = .06$. Post-hoc *t* tests indicated that for conflict syllogisms, counterexample reasoners had longer fixation times on the premises than did statistical reasoners [Welch’s $t(60.3) = 2.64, p = .011, d = 0.57$]. There were no fixation time differences between the strategies for nonconflict syllogisms [$t(111) = 0.83, p = .41, d = 0.2$] or neutral syllogisms [$t(111) = 1.48, p = .142, d = 0.29$]. The predicted pattern of fixation times was observed—that is, counterexample reasoners attended to the premises longer than did statistical reasoners, but this was isolated to the conflict problems. Below we analyze each premise separately, to provide a more nuanced analysis.

For Premise 1, we found an effect of syllogism type, $F(1.6, 180.9) = 71.9, p < .001, \eta_p^2 = .4$; a marginal effect of strategy, $F(1, 111) = 3.4, p = .068, \eta_p^2 = .03$; and a marginal Syllogism Type \times Strategy interaction, $F(1.6, 180.9) = 2.5, p = .099, \eta_p^2 = .02$. Post-hoc *t* tests revealed that counterexample reasoners spent more time fixating on Premise 1 for conflict syllogisms than did statistical reasoners [$t(111) = 2.7, p = .009, d = 0.5$], but that fixation times did not differ between the strategies for

Table 6 Mean fixation times (in seconds) as a function of area of interest (Premise 1, Premise 2, conclusion), syllogism type (conflict, nonconflict, neutral), and strategy (statistical, counterexample) in Study 2

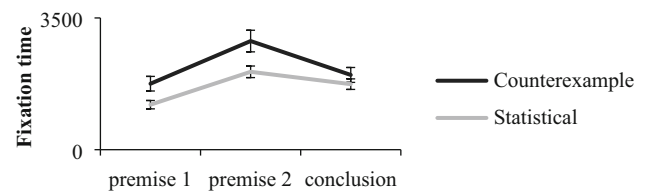
	Strategy	
	Statistical (<i>n</i> = 74)	Counterexample (<i>n</i> = 39)
Premise 1		
Conflict	1.20 (0.96)	1.75 (1.21)
Nonconflict	0.86 (0.78)	1.04 (0.77)
Neutral	1.78 (1.50)	2.18 (1.54)
Premise 2		
Conflict	2.07 (1.32)	2.89 (1.79)
Nonconflict	1.06 (0.77)	1.12 (0.84)
Neutral	1.82 (1.14)	2.19 (1.58)
Conclusion		
Conflict	1.74 (1.20)	1.99 (1.23)
Nonconflict	0.95 (0.73)	1.05 (0.75)
Neutral	1.41 (0.91)	1.45 (0.91)

Numbers in parentheses are standard deviations.

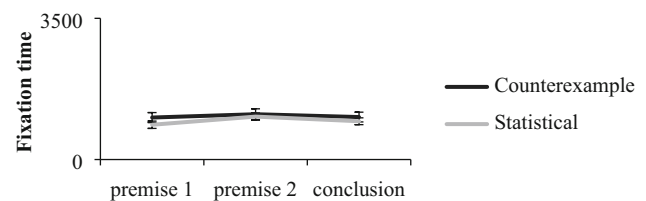
nonconflict syllogisms [$t(111) = 1.2, p = .246, d = 0.2$] or neutral syllogisms [$t(111) = 1.3, p = .184, d = 0.3$]. Thus, fixation times for Premise 1 were consistent with the predicted pattern of visual attention.

The analysis of Premise 2 showed consistent evidence. We found an effect of syllogism type, $F(2, 222) = 112.6, p < .001, \eta_p^2 = .5$, and a marginal effect of strategy, $F(1, 111) = 3.6, p = .061, \eta_p^2 = .03$, which were qualified by a Syllogism Type \times Strategy interaction, $F(2, 222) = 8.3, p < .001, \eta_p^2 = .07$. Post-hoc *t* tests showed that on conflict syllogisms, counterexample reasoners fixated more on Premise 2 than did statistical reasoners [Welch’s $t(60.4) = 2.5, p = .015, d = 0.5$]. Fixation times on Premise 2 for the counterexample and statistical reasoners did not differ for nonconflict syllogisms [$t(111) = 0.38, p = .708, d =$

Conflict syllogisms



Non-conflict syllogisms



Neutral syllogisms

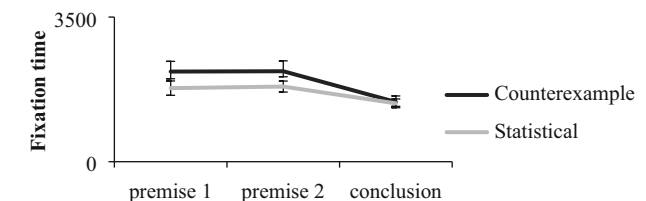


Fig. 2 Mean fixation times as a function of area of interest (Premise 1, Premise 2, conclusion) and strategy (statistical, counterexample) for each syllogism type (conflict, nonconflict, neutral) in Study 2. Error bars represent standard errors.

0.1] or for neutral syllogisms [$t(111) = 1.44, p = .154, d = 0.3$]. Thus, the fixation times for both premises are in accord with our hypotheses—that is, counterexample reasoners attended to the premises longer overall than did the statistical reasoners.

Finally, the pattern of fixation times for the syllogism conclusions differed from the pattern for the premises. We found an effect of syllogism, $F(1.7, 185) = 68.5, p < .001, \eta_p^2 = .38$, but no effect of strategy was found, $F(1, 111) = 0.55, p = .461, \eta_p^2 = .01$, nor did these factors interact, $F(2, 222) = 1.1, p = .344, \eta_p^2 = .01$. Thus, whereas the fixation times for syllogism conclusions were different between syllogism types (in a pattern consistent with response times and overall fixation times), no differences in fixation times between the strategies emerged.

In sum, the results of the fixation time analyses were consistent with our hypotheses. The pattern of response times indicated that counterexample reasoners spent more time reasoning on the conflict problems. Fixation times indicated that this extra time was spent attending to Premises 1 and 2 for the conflict syllogisms. There were no differences in fixation times for the conclusion on these conflict problems, nor were there any differences between the strategies on consistent or neutral problems. Therefore, the increase in fixation times by counterexample reasoners occurs only when there is a conflict between believability and validity, and is then localized to the problem premises.

Confidence We analyzed confidence with a 3 (Syllogism Type: conflict, nonconflict, neutral) \times 2 (Strategy: statistical, counterexample) repeated measures ANOVA. An effect of syllogism type, $F(2, 222) = 66.9, p < .001, \eta_p^2 = .38$, indicated that overall, confidence was lower for conflict problems ($M = 7.01, SD = 1.04$) than for nonconflict problems ($M = 8.22, SD = 1.01; t = 10.9, p_{\text{Tukey}} < .001, d = 1.0$). This is consistent with previous studies (De Neys, 2014). We found no effect of strategy use, $F(1, 111) = 1.6, p = .208, \eta_p^2 = .01$. However, this null main effect of strategy was qualified by an interaction with syllogism type, $F(2, 222) = 4.7, p = .01, \eta_p^2 = .04$. Whereas the two types of reasoners reported similar levels of confidence for both conflict syllogisms [$t(111) = 0.9, p = .365, d = 0.2$] and nonconflict syllogisms [$t(111) = 1.2, p = .238, d = 0.2$], counterexample reasoners ($M = 7.51, SD = 1.14$) reported higher confidence for neutral syllogisms than did the statistical reasoners ($M = 6.98, SD = 0.95$) [$t(111) = 2.6, p = .01, d = 0.5$]. Despite reporting higher confidence on the neutral items, the counterexample reasoners were not more accurate on these syllogisms, nor were they faster to respond than the statistical reasoners. It should be noted that although counterexample reasoners have been shown to be more accurate with abstract premises (Markovits et al., 2012), the neutral premises used in the present study used familiar content and were amenable to a probabilistic solution, which could explain the lack of a strategy difference on these items.

Discussion

We found clear evidence that counterexample and statistical reasoners differ in their approaches to solving syllogisms for which logical validity and conclusion belief are in conflict. Consistent with previous results (Markovits, Brisson, de Chantal, & Thompson, 2017b), counterexample reasoners produced valid responses on conflict syllogisms more often than did statistical reasoners, indicating that they were less influenced by conclusion believability when making judgments of logical validity. Counterexample reasoners also took longer to respond to these conflict syllogisms. Our hypothesis was that the extra time taken by the counterexample reasoners to solve conflict problems (i.e., invalid–believable) was due to spending a greater amount of time attending to the premises of the argument. The evidence in favor of this hypothesis was clear: Counterexample reasoners demonstrated longer fixation times than statistical reasoners on both Premise 1 and Premise 2, but there was no difference in fixation times on the syllogism conclusions. The fact that this effect was somewhat stronger for the minor premise (Premise 2) provides additional evidence, since the minor premise is the critical element in determining whether or not a given inference allows for a possible counterexample.

Confidence ratings additionally suggest that the two strategies produce equal levels of sensitivity to the conflict, thus indicating that differences in visual behaviors stem from an allocation of attention related to strategy use rather than to differences in the awareness to conflict. Importantly, no such differences were observed with either neutral or nonconflict syllogisms. This is consistent with the idea that the differences in visual fixation are the result of different reactions to conflict between conclusion belief and the premise-based information. These data are consistent with our assumption that counterexample reasoners attend more to premises because those are necessary in order to verify, or confirm, the existence of potential counterexamples that are given priority when they conflict with other sources of information. Since generating counterexamples correlates with logical validity, this explains counterexample reasoners' tendency to give the logically valid response to conflict syllogisms. On the other hand, these results are also consistent with the assumption that statistical reasoners resolve conflict problems by reconciling the conclusion likelihood produced by considering the premises and the certainty of the presented conclusion; given the empirical certainty of the latter, this increases the probability of belief-based judgments. These results provide further evidence consistent with the dual-strategy model, that there are two qualitatively different ways of evaluating argument conclusions. Counterexample and statistical reasoners demonstrate observable behavioral differences in time spent reasoning on conflict problems and in the ability to reject believable but invalid inferences. The eye-tracking measures provided novel evidence that these strategies also differ in the level of attention to the information in the problems, with counterexample reasoners attending relatively more to the premises of the argument.

General discussion

Traditionally, variability in reasoning ability, particularly variability in susceptibility to the belief bias effect (e.g., Evans et al., 1983; Evans & Curtis-Holmes, 2005; Klauer et al., 2000), has been explained by individual differences in working memory capacity (Barrett, Tugade, & Engle, 2004; De Neys, 2006) or other general cognitive abilities (e.g., Sá et al., 1999). Working memory capacity is assumed to be needed to overturn default, belief-based responses; formulate a different representation of the problem; and derive a logically valid response. Indeed, the results of our first study showed that reasoners with higher working memory capacity showed less susceptibility to belief-based responses than did their lower-capacity counterparts, consistent with this explanation.

However, these results also showed that, although working memory was a predictor of individual differences in belief bias, strategy use was actually a better predictor. Contrary to differences in working memory capacity, strategy use also resulted in different patterns of response times on conflict and nonconflict problems (see Study 1). These results are problematic for approaches that consider cognitive capacity to be the most important component in explaining the belief bias effect, but they are consistent with the dual-strategy model of reasoning (Markovits et al., 2015; Markovits, Brisson, de Chantal, & Thompson, 2017b; Markovits et al., 2013; Markovits et al., 2012; Markovits et al., 2017a). This model proposes that two reasoning strategies exist, characterized by differences in the way that information is accessed and processed. More specifically, it is proposed that the statistical strategy involves a relatively intuitive processing of a wide array of associatively available information, whereas the counterexample strategy creates a more structured representation of key problem parameters, which is more working memory intensive. However, the results of the first study show that this difference in working memory was insufficient to explain the effect of strategy use on the extent to which conclusion belief influences participants' inferences, which can be explained by other forms of processing differences, which we hypothesized to be attentional in nature, consistent with some indirect results (Markovits, Trémolière, & Blanchette, 2018b).

The eye-tracking method used in Study 2 gave us a first glimpse into the attentional processes that underlie the functioning of each strategy. In the absence of a conflict between logical validity and conclusion belief, there were no behavioral differences between strategies. Both statistical and counterexample reasoners spent more time on conflict problems, but the relative increase was somewhat greater among the latter group. In addition, the eye-tracking data showed that, on conflict problems, the extra time that counterexample reasoners spent thinking about conflict problems was concentrated on attending to the premises. This is consistent with the idea that conflict detection by counterexample reasoners results in increased examination of information suggesting potential counterexamples (Markovits, Trémolière, & Blanchette,

2018b), which in this case is provided by the premises. Statistical reasoners also spent relatively more time on conflict problems (although their increase was less than with counterexample reasoners). However, in contrast to counterexample reasoners, they continued to give equal attention to premises and conclusion, also consistent with previous results suggesting that statistical reasoners give equal weight to different sources of information. Thus, although both types of reasoners were equally sensitive to the conflict (as indexed by confidence ratings and response times), counterexample reasoners responded by spending relatively more time with the more difficult problems (i.e., conflict syllogisms), time that was concentrated on processing the premises. These data are consistent with Thompson and Johnson (2014), who found that low- and high-capacity reasoners were equally sensitive to conflict, but the high-capacity reasoners spent more time thinking about the conflict problems.

Thus, it is not surprising that Study 1 revealed that counterexample reasoners had higher working memory capacity than statistical reasoners, given that their attentional patterns point toward a more cognitively demanding strategy. Of interest, however, is that differences in strategy use predicted variability in the belief bias effect beyond that predicted by individual differences in working memory capacity. This suggests that the dual-strategy distinction captures differences in information processing that are difficult to explain with differences in working memory capacity alone, but are consistent with the attentional differences captured by the eye-tracking data.

Finally, it is useful to examine how the dual-strategy model explanation of individual differences in belief bias is related to dual-process theories. As we previously stated, the dual-strategy model proposes that reasoning is under the control of a single strategy at any given moment. On the surface, this appears to be difficult to reconcile with dual-process theories, which conceive that both autonomous (Type 1) and working-memory-intensive analytical (Type 2) processes contribute to reasoning outcomes. Dual-process models of reasoning rely on cognitive capacity as the central element explaining individual differences in weighting the contributions of Type 1 and Type 2 processes to reasoning (e.g., Evans & Stanovich, 2013), which is only partially consistent with the present results. We note, however, that dual-process theorists do not claim that cognitive capacity is the only variable that explains individual differences in reasoning. Indeed, Stanovich (2018) has recently proposed a number of individual characteristics that are important determinants of reasoning variability, including understanding of basic concepts, such as logical necessity. One possible variable that may underlie some part of the variance predicted by strategy use is analytic thinking dispositions (Pennycook et al., 2015; Toplak, West, & Stanovich, 2011). That is, counterexample reasoners may be more inclined to engage in analytical thinking than are statistical reasoners, and this motivational variable may explain why they are willing to undertake a cognitively more demanding style of information processing, although this theory remains speculative.

Irrespective of how the two models are related, one advantage of the dual-strategy model is that it proposes a relatively detailed understanding of reasoning algorithms and the processing characteristics associated with each strategy. In conclusion, these data provide evidence that individual differences in reasoning strategies and the resulting differences in processing styles modulate the ways that people make inferences. Importantly, the processing differences between the two strategies are not only related to cognitive capacity, but also reflect differences in the ways that people access information. In addition, the fact that reasoning strategies predict differences in reasoning outcomes above and beyond that predicted by cognitive capacity points to the impor-

tance of understanding how differences in strategy use translate to differences in reasoning outcomes. The relation between cognitive capacity, strategy use, and reasoning performance that we have uncovered is complex, and sorting out the details of the relationship will be a rich field for future studies.

Acknowledgements This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada to H.M. [RGPIN-2016-04865] and V.T. [RGPIN 2018-04466].

Open Practices Statement The data and materials for all experiments are available on demand, and none of the experiments was preregistered.

Appendix

Table 7 Description of the strategy diagnostic questionnaire

Instructions		
<p>For this task, imagine that scientists discovered a new inhabited planet, called Planet Kronus. A team of scientists was then sent to this planet. These scientists discovered some things that do not exist on Planet Earth. In the following pages, you will see a description of their discoveries. Read these carefully, because they give important information about the discoveries. For each discovery, you will be given a rule that has been confirmed by scientists, which you must consider to be true. It is very important that you suppose that each rule presented by the scientists is always true when responding to the following question. Following each rule, an observation and a conclusion will be presented. You must indicate whether or not the conclusion can be logically drawn from the presented information. Please take your time to answer the questions.</p> <p>Examples of Problems</p>		
Inference	Prob	Content
MP	100	<p>A team of meteorologists watching the local climate of the planet Kronus noted an interesting phenomenon. They noted that on Kronus: If it thardonnes, then the soil will become sticky. On the last 1,000 times it thardonned, meteorologists made the following observations: 1,000 times it thardonned and the soil became sticky. 0 times it thardonned and the soil did not become sticky. From this information, John reasoned as follows: If it thardonnes, then the soil will become sticky. Observation: It thardonnes. Conclusion: The soil will become sticky. Indicate whether the conclusion of John can be drawn logically or not from the information provided.</p>
AC	90	<p>By exploring a cave of Kronus, geologists have discovered a variety of very special stones, trolytes. Following a series of observations, they argue that on Kronus: If a trolYTE is heated, then it will release philoben gas. On the last 1,000 times they observed trolytes, geologists have made the following observations: 900 trolytes were heated and released philoben gas. 100 trolytes were not heated and released philoben gas. From this information, Mary reasoned as follows: If a trolYTE is heated, then it will release philoben gas. Observation: A trolYTE releases philoben gas. Conclusion: The trolYTE was heated. Indicate whether the conclusion of Mary can be drawn logically or not from the information provided.</p>
AC	50	<p>A team of meteorologists watching the local climate of the planet Kronus noted an interesting phenomenon. They noted that on Kronus: If the sun shines, then the ground will become green. On the last 1,000 times when they observed the ground, meteorologists made the following observations: 505 times the sun was shining and the ground became green. 495 times the sun did not shine and the ground became green. From this information, Jennifer reasoned as follows: If the sun shines, then the ground will become green. Observation: The ground became green. Conclusion: The sun was shining. Indicate whether the conclusion of Jennifer can be drawn logically or not from the information provided.</p>

Prob refers to the relative probability that the conclusion presented is true. MP: modus ponens; AC: affirmation of the consequent.

Table 8 General description of the operation span task and the changes made for online assessment

General Description		
<p>Prior to the official test trials, participants completed three different practice phases. They were first familiarized with the response format for recalling the letters. They then solved 16 math problems in order to establish the time that was to be allotted in the official working memory trials. For this practice phase, participants were instructed to respond to the math problems as fast as possible. Response times were recorded by the program and average time was calculated afterwards. Finally, they were presented a practice phase that combined both the letter and math parts of the task. The official test trials alternatively presented letters and math problems with a random set size varying from 3 to 7. Every set size was presented once and participants were not informed of the number of letters to be remembered until the recall phase. It should be noted that letters must be recalled in the exact same order as they were presented. Similar to the original task, we asked participants to keep their accuracy in the math problems above 80%. The cumulative math accuracy was displayed in a feedback screen at the end of each trial.</p>		
Changes		
	Original Task	Adapted Task
Response format	Consists of clicking the box next to the appropriate letters in the correct order; the boxes were displayed in a 4×3 matrix.	Consists of a $n \times 12$ matrix table, in which n is the number of letters to be recalled in the trial. Rows correspond to the location of the letter in the series and columns present possible letters to be selected. As in the original task, letters in the columns were kept constant for each trial.
Time to solve math problems	The time allotted was the average time required during the practice phase plus 2.5 <i>SDs</i> .	Pretesting suggested that this was too long, with participants being able to rehearse the letters while solving the math problems. Consequently, the time allotted for each participant was simply the average time they needed during the math practice phase.
Number of trial blocks	Three blocks of trials are used.	A single block was used. A study by Foster et al. (2015) has shown that performance in a single block is highly correlated with performance across three blocks.
Detection of facilitating strategies		We pretested our measure, during which time we observed that the number of perfectly recalled trials during pretesting was surprisingly high for some participants. We interpreted this finding as indicating that some participants may be using a facilitating strategy, such as writing down the letters. Indeed, Hicks, Foster, and Engle (2016) documented a tendency for some participants, particularly low ability ones, to write down letters while performing the task, leading to unusually high levels of performance. They addressed this issue by using simple images as the recall material instead of letters. To account for this possibility, while keeping the task used in this study as similar as possible to the original, we included an additional control trial with a set size of 10 at the end of the task. The probability that participants perfectly recall ten letters out of ten without writing them down is very low. In fact, working memory tasks usually involve sets size varying from 3 to 7 because much of the variability in performances occurs within this range. Participants that perfectly recalled all the ten letters were thus excluded from the analysis.

For additional information about the original task, see Unsworth, Heitz, Schrock, and Engle (2005).

Table 9 Content of the belief bias task as a function of syllogism type (valid, invalid) and conclusion belief (believable, unbelievable) in Study 1

Conclusion Belief	Syllogism Type	
	Valid	Invalid
Believable	(Nonconflict)	(Conflict)
	<i>All dogs have legs.</i>	<i>All dogs have legs.</i>
	<i>Fish do not have legs.</i>	<i>German Sheperds have legs.</i>
	<i>Fish are not dogs.</i>	<i>German Sheperds are dogs.</i>
	<i>All trees have roots.</i>	<i>All trees have roots.</i>
	<i>Radios do not have roots.</i>	<i>Maple trees have roots.</i>
	<i>Radios are not trees.</i>	<i>Maple trees are trees.</i>
	<i>All desserts are sweet.</i>	<i>All desserts are sweet.</i>
	<i>Hats are not sweet.</i>	<i>Cakes are sweet.</i>
	<i>Hats are not desserts.</i>	<i>Cakes are desserts.</i>
Unbelievable	(Conflict)	(Nonconflict)
	<i>All things that fly in the sky have lungs.</i>	<i>All things that fly in the sky have lungs.</i>
	<i>Airplanes do not have lungs.</i>	<i>Cadavers have lungs.</i>
	<i>Airplanes do not fly in the sky.</i>	<i>Cadavers fly in the sky.</i>
	<i>All trees have leaves.</i>	<i>All trees have leaves.</i>
	<i>Fir trees do not have leaves.</i>	<i>Tulips have leaves.</i>
	<i>Fir trees are not trees.</i>	<i>Tulips are trees.</i>
	<i>All animals have hairs.</i>	<i>All animals have hairs.</i>
	<i>Birds do not have hairs.</i>	<i>Carpets have hairs.</i>
	<i>Birds are not animals.</i>	<i>Carpets are animals.</i>

Table 10 Content of the belief bias task as a function of syllogism type (conflict, nonconflict, neutral) in Study 2

Conflict	Nonconflict	Neutral
<i>Most cars have valves.</i>	<i>Most cars have valves.</i>	<i>Most servers are knitters.</i>
<i>Most vehicles with four doors have valves.</i>	<i>Most airplanes have valves.</i>	<i>Most campers are knitters.</i>
<i>Most cars are vehicles with four doors.</i>	<i>Most cars are airplanes.</i>	<i>Most servers are campers.</i>
<i>Most cats breathe.</i>	<i>Most soaps have a price.</i>	<i>Most brokers are athletes.</i>
<i>Most animals with fur breathe.</i>	<i>Most dogs have a price.</i>	<i>Most crafters are athletes.</i>
<i>Most cats are animals with fur.</i>	<i>Most soaps are dogs.</i>	<i>Most brokers are crafters.</i>
<i>Most trees have roots.</i>	<i>Most tables are solid.</i>	<i>Most surgeons are bowlers.</i>
<i>Most plants with leaves have roots.</i>	<i>Most chairs are solid.</i>	<i>Most runners are bowlers.</i>
<i>Most trees are plants with leaves.</i>	<i>Most tables are chairs.</i>	<i>Most surgeons are bowlers.</i>
<i>Most poodles have a price.</i>	<i>Most trees have roots.</i>	<i>Most cashiers are weavers.</i>
<i>Most friendly animals have a price.</i>	<i>Most tulips have roots.</i>	<i>Most jugglers are weavers.</i>
<i>Most poodles are friendly animals.</i>	<i>Most trees are tulips.</i>	<i>Most cashiers are jugglers.</i>
<i>Most tables are solid.</i>	<i>Most cats breathe.</i>	<i>Most typists are dancers.</i>
<i>Most things made of wood are solid.</i>	<i>Most birds breathe.</i>	<i>Most curlers are dancers.</i>
<i>Most tables are made of wood.</i>	<i>Most cats are birds.</i>	<i>Most typists are curlers.</i>

All syllogisms are invalid.

References

- Ball, L. J., Phillips, P., Wade, C. N., & Quayle, J. D. (2006). Effects of belief and logic on syllogistic reasoning: Eye-movement evidence for selective processing models. *Experimental Psychology*, *53*, 77–86. <https://doi.org/10.1027/1618-3169.53.1.77>
- Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, *130*, 553–573.
- Barrouillet, P., Gauffroy, C., & Lecas, J.-F. (2008). Mental models and the suppositional account of conditionals. *Psychological Review*, *115*, 760–771. <https://doi.org/10.1037/0033-295X.115.3.760>
- Brisson, J., de Chantal, P.-L., Lortie Forgues, H., & Markovits, H. (2014). Belief bias is stronger when reasoning is more difficult. *Thinking & Reasoning*, *20*, 385–403. <https://doi.org/10.1080/13546783.2013.875942>
- Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, *19*, 274–282. <https://doi.org/10.3758/BF03211151>
- De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, *17*, 428–433. <https://doi.org/10.1111/j.1467-9280.2006.01723.x>
- De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning*, *20*, 169–187. <https://doi.org/10.1080/13546783.2013.854725>
- De Neys, W. (Ed.). (2018). *Dual process theory 2.0*. Oxford, UK: Routledge.
- Evans, J. St. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning*, *13*, 321–339. <https://doi.org/10.1080/13546780601008825>
- Evans, J. St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*, 295–306. <https://doi.org/10.3758/BF03196976>
- 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. <https://doi.org/10.1080/13546780542000005>
- Evans, J. St. B. T., Over, D. E., & Handley, S. J. (2005). Supposition, extensionality and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002). *Psychological Review*, *112*, 1040–1052. <https://doi.org/10.1037/0033-295X.112.4.1040>
- Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*, 223–241. <https://doi.org/10.1177/1745691612460685>
- Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*, 226–236. <https://doi.org/10.3758/s13421-014-0461-7>
- Gagnon-St-Pierre, E., Markovits, H. (2019). *Reasoning strategies explain individual differences in social reasoning*. Manuscript in preparation.
- Hicks, K. L., Foster, J. L., & Engle, R. W. (2016). Measuring working memory capacity on the web with the online working memory lab (the OWL). *Journal of Applied Research in Memory and Cognition*, *5*, 478–489. <https://doi.org/10.1016/j.jarmac.2016.07.010>
- Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences*, *5*, 434–442. [https://doi.org/10.1016/S1364-6613\(00\)01751-4](https://doi.org/10.1016/S1364-6613(00)01751-4)
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.
- Johnson-Laird, P. N., & Byrne, R. M. (1992). Modal reasoning, models, and Manktelow and Over. *Cognition*, *43*, 173–182. [https://doi.org/10.1016/0010-0277\(92\)90061-1](https://doi.org/10.1016/0010-0277(92)90061-1)
- Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, *107*, 852–884. <https://doi.org/10.1037/0033-295X.107.4.852>
- Markovits, H., & Barrouillet, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review*, *22*, 5–36. <https://doi.org/10.1006/drev.2000.0533>
- Markovits, H., Brisson, J., & de Chantal, P. L. (2015). Additional evidence for a dual-strategy model of reasoning: Probabilistic reasoning is more invariant than reasoning about logical validity. *Memory & Cognition*, *43*, 1208–1215. <https://doi.org/10.3758/s13421-015-0535-1>
- Markovits, H., Brisson, J., & de Chantal, P. L. (2017a). Logical reasoning versus information processing in the dual-strategy model of reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 72–80.
- Markovits, H., Brisson, J., de Chantal, P. L., & Singmann, H. (2018a). Multiple layers of information processing in deductive reasoning: combining dual strategy and dual-source approaches to reasoning. *Journal of Cognitive Psychology*, *30*, 1–12. <https://doi.org/10.1080/20445911.2018.1458729>
- Markovits, H., Brisson, J., de Chantal, P. L., & Thompson, V. A. (2017b). Interactions between inferential strategies and belief bias. *Memory & Cognition*, *45*, 1–11. <https://doi.org/10.3758/s13421-017-0723-2>
- Markovits, H., Brunet, M. L., Thompson, V., & Brisson, J. (2013). Direct evidence for a dual process model of deductive inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1213–1222. <https://doi.org/10.1037/a0030906>
- Markovits, H., Lortie Forgues, H., & Brunet, M.-L. (2012). More evidence for a dual-process model of conditional reasoning. *Memory & Cognition*, *40*, 736–747. <https://doi.org/10.3758/s13421-012-0186-4>
- Markovits, H., & Nantel, G. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory & Cognition*, *17*, 11–17. <https://doi.org/10.3758/BF03199552>
- Markovits, H., Trémolière, B., & Blanchette, I. (2018b). Reasoning strategies modulate gender differences in emotion processing. *Cognition*, *170*, 76–82. <https://doi.org/10.1016/j.cognition.2017.09.012>
- Markovits, H., & Vachon, R. (1990). Conditional reasoning, representation, and level of abstraction. *Developmental Psychology*, *26*, 942–951. <https://doi.org/10.1037/0012-1649.26.6.942>
- Mele, M. L., & Federici, S. (2012). Gaze and eye-tracking solutions for psychological research. *Cognitive Processing*, *13*(Suppl 1), 261–265. <https://doi.org/10.1007/s10339-012-0499-z>
- Morsanyi, K., & Handley, S. J. (2008). How smart do you need to be to get it wrong? The role of cognitive capacity in the development of heuristic-based judgment. *Journal of Experimental Child Psychology*, *99*, 18–36. <https://doi.org/10.1016/j.jecp.2007.08.003>
- Newman, I. R., & Thompson, V. A. (2019) Individual differences in base-rate neglect and sensitivity to conflict. Manuscript in preparation.
- Oakhill, J., Johnson-Laird, P. N., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition*, *31*, 117–140.
- Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford, UK: Oxford University Press.
- 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. <https://doi.org/10.1016/j.cogpsych.2015.05.001>
- Psychology Software Tools, Inc. (2012). E-Prime 2.0. Retrieved from <http://www.pstnet.com>
- Quayle, J. D., & Ball, L. J. (2000). Working memory, metacognitive uncertainty, and belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, *53A*, 1202–1223. <https://doi.org/10.1080/713755945>

- Robison, M. K., & Unsworth, N. (2017). Individual differences in working memory capacity and resistance to belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, *70*, 1471–1484. <https://doi.org/10.1080/17470218.2016.1188406>
- Sá, W. C., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, *91*, 497–510.
- SensoMotoric Instruments, Inc. (2017). BeGaze 3.7. <http://www.smivision.com>
- Slooman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22. <https://doi.org/10.1037/0033-2909.119.1.3>
- Stanovich, K. E. (2018). Miserliness in human cognition: The interaction of detection, override and mindware. *Thinking & Reasoning*, *24*, 423–444. <https://doi.org/10.1080/13546783.2018.1459314>
- Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*, 342–357. <https://doi.org/10.1037/0022-0663.89.2.342>
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, *23*, 645–665, disc. 665–726.
- Stuppelle, E. J., Ball, L. J., Evans, J. S. B., & Kamal-Smith, E. (2011). When logic and belief collide: Individual differences in reasoning times support a selective processing model. *Journal of Cognitive Psychology*, *23*, 931–941. <https://doi.org/10.1080/20445911.2011.589381>
- Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition*, *22*, 742–758. <https://doi.org/10.3758/BF03209259>
- Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning*, *20*, 215–244. <https://doi.org/10.1080/13546783.2013.869763>
- Thompson, V. A., & Newman, I. R. (2019). Working memory, autonomy, and dual process theories: A roadmap. In S. Elqayam, I. Douven, J. St. B. T. Evans, & N. Cruz (Eds.), *Logic and uncertainty in the human mind: A tribute to David Over*. London, UK: Routledge.
- Thompson, V. A., Prowse-Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*, 107–140. <https://doi.org/10.1016/j.cogpsych.2011.06.001>
- Thompson, V. A., Striener, C. L., Reikoff, R., Gunter, R. W., & Campbell, J. I. (2003). Syllogistic reasoning time: Disconfirmation disconfirmed. *Psychonomic Bulletin & Review*, *10*, 184–189. <https://doi.org/10.3758/BF03196483>
- 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. <https://doi.org/10.3758/s13421-011-0104-1>
- Tsujii, T., & Watanabe, S. (2009). Neural correlates of dual-task effect on belief-bias syllogistic reasoning: A near-infrared spectroscopy study. *Brain Research*, *1287*, 118–125. <https://doi.org/10.1016/j.brainres.2009.06.080>
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505. <https://doi.org/10.3758/BF03192720>
- Verschueren, N., Schaeken, W., & d’Ydewalle, G. (2005a). A dual-process specification of causal conditional reasoning. *Thinking & Reasoning*, *11*, 239–278. <https://doi.org/10.1080/13546780442000178>
- Verschueren, N., Schaeken, W., & d’Ydewalle, G. (2005b). Everyday conditional reasoning: A working memory-dependent tradeoff between counterexample and likelihood use. *Memory & Cognition*, *33*, 107–119. <https://doi.org/10.3758/BF03195301>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.