

Interactions between inferential strategies and belief bias

Henry Markovits¹ · Janie Brisson¹ · Pier-Luc de Chantal¹ · Valerie A. Thompson²

Published online: 12 June 2017
© Psychonomic Society, Inc. 2017

Abstract The dual strategy model of reasoning proposed by Verschueren, Schaeken, and d’Ydewalle (*Thinking & Reasoning*, 11(3), 239–278, 2005a; *Memory & Cognition*, 33(1), 107–119, 2005b) suggests that people can use either a statistical or a counterexample-based strategy to make deductive inferences. Subsequent studies have supported this distinction and investigated some properties of the two strategies. In the following, we examine the further hypothesis that reasoners using statistical strategies should be more vulnerable to the effects of conclusion belief. In each of three studies, participants were given abstract problems used to determine strategy use and three different forms of syllogism with believable and unbelievable conclusions. Responses, response times, and feeling of rightness (FOR) measures were taken. The results show that participants using a statistical strategy were more prone to the effects of conclusion belief across all three forms of reasoning. In addition, statistical reasoners took less time to make inferences than did counterexample reasoners. Patterns of variation in response times and FOR ratings between believable and unbelievable conclusions were very similar for both strategies, indicating that both statistical and counterexample reasoners were aware of conflict between conclusion belief and premise-based reasoning.

Keywords Conditional reasoning · Belief bias · Dual process theories · Probabilistic reasoning · Mental models · Dual strategy

The ability to make deductive inferences, that is, to understand that a single conclusion, is a logical consequence of whatever preconditions are assumed is a key component of higher level cognition. This kind of reasoning underlies most scientific and mathematical analyses, and understanding its nature is a critical problem. Unfortunately, many decades of studies examining how people reason, even in limited laboratory conditions, clearly show that the one major characteristic of logical reasoning is its extreme variability.

Understanding the sources of this variability has been one of the major aims of much research. An important current debate about the nature of people’s inferential processes concerns the underlying nature of these processes. Many studies have shown that the inferences people make with familiar premises are at least partly determined by what we can call the statistical structure of implicit information related to premises (i.e., the way that knowledge about premises are structured). There are at least two related but separable dimensions to how this knowledge is structured (Thompson, 2000). One important characteristic when reasoning with a conditional (If P, then Q) relation is the relative number of alternative antecedents that are available in long-term memory, that is, potential cases of A and Q. Many studies have consistently shown that premises for which there are relatively high numbers of alternative antecedents tend to generate high levels of denial of the putative conclusion for both the AC and the DA inferences (Cummins, 1995; Cummins, Lubart, Alksnis, & Rist, 1991; Markovits, 1984; Markovits & Vachon, 1990; Thompson, 1994). A second is a subjective estimate of the conditional probabilities. Probabilistic theories accordingly

✉ Henry Markovits
Henrymarkovits@gmail.com

¹ Department of Psychology, Université du Québec à Montréal, C.P. 8888, Succ A, Montréal, QC, Canada

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

postulate that inference are driven by a Bayesian analysis of the statistical properties of the premises that leads to an estimation of the relative likelihood of a given conclusion being true (Evans, Over, & Handley, 2005; Evans & Over, 2004; Oaksford & Chater, 2007; Oaksford, Chater, & Larkin, 2000). This estimate could be based either on the availability of counterexamples, a subjective estimate of the likelihood that a counterexample occurs, or retrieved knowledge about the conditional probability.

Information about premises can also be processed in a different way; by generating explicit counterexamples to putative conclusions. Such a process is a critical component of mental model theory and variants (Barrouillet, Gauffroy, & Lecas, 2008; Johnson-Laird, 2001; Johnson-Laird, & Byrne, 1991, 2002; Markovits & Barrouillet, 2002). The key component of such an inferential process is that conclusions for which explicit counterexamples can be generated are considered to be invalid. The probability of generating an explicit counterexample, and thus rejecting a putative conclusion, increases as the number of counterexamples accessible to a reasoner increases.

Consistent with the finding that there are multiple ways for people to represent statistical information about conditional relationships (Thompson, 2000), there has recently been increasing evidence that people have access to multiple reasoning strategies. The dual strategy model of reasoning postulates that people have access to both statistical and counterexample-based forms of inference (Markovits, Brisson, & de Chantal, 2015, 2016; Markovits, Brunet, Thompson, & Brisson, 2013; Markovits, Lortie Forgues, & Brunet, 2012; Verschueren, Schaeken, & d'Ydewalle, 2005a, 2005b). Statistical strategies use associative access to knowledge about premises in order to produce likelihood estimations of putative conclusions. Counterexample strategies require generation of internal representations of premises coupled with a search for potential counterexamples to putative conclusions, which also use knowledge about premises. These strategies are generally slower and more cognitively costly and produce dichotomous judgments of validity. Importantly, it has been found that the relation between strategies and norms of logical validity is not constant, with counterexample strategies sometimes generating lower levels of logically valid responses than statistical strategies (Markovits et al., 2016). This latter is consistent with results that suggest that reasoners can reject valid inferences when producing counterexamples (Byrne, Espino, & Santamaria, 1999). In this context, it should be acknowledged that while the original model was developed as a way to integrate probabilistic and mental model approaches to reasoning, the basic distinction that it suggests resides primarily on the nature of the way that statistical information is processed (Markovits et al., 2016), and, as above, does not preclude the possibility that the underlying models might differ. For example, the concept of p validity (Evans, Thompson, &

Over, 2015; Singmann et al., 2014) would generate the same pattern of inferences as is proposed by the counterexample strategy. In fact, the p -validity model is in many respects isomorphic to the mental model description of the counterexample strategy.

Instead, the critical distinction captured by the dual strategy model is that when making judgments of logical validity, both counterexample and statistical strategies use underlying statistical information derived from knowledge about premises, with the key difference determined by the way that this information is processed. A statistical strategy produces an essentially Bayesian analysis, which evaluates conclusion likelihood in the full context of available knowledge (Fernbach & Erb, 2013). A counterexample strategy, by contrast, must focus more particularly on whether problem premises allow the generation of an explicit counterexample, which implies a more narrow focus (Markovits et al., 2015). This leads to the general hypothesis that reasoners using a statistical strategy will be more open to broader contextual effects than those using a counterexample strategy.

One of the clearest examples of this kind of effect is the well-known effect of conclusion belief on reasoning (Evans, Barston, & Pollard, 1983). There is a clearly documented effect of conclusion believability on people's evaluation of the logical validity of a putative conclusion. Studies have also found that the relative effect of belief is stronger with invalid than with valid logical forms (J. S. B. Evans et al., 1983), that this is stronger when reasoners are asked to evaluate putative conclusions than when they must produce a conclusion (Markovits & Nantel, 1989), and that strong logical instructions can reduce the effects of belief (Evans, Newstead, Allen, & Pollard, 1994). The interaction between validity and empirical belief is in fact one of the cornerstones of dual-process theories that postulate that inferential reasoning is the product of interactions between two systems of reasoning, one of which can generate analytic logical inferences based on premises while the other generates heuristic inferences based on stored knowledge and beliefs. However, different theoretical explanations have been given (e.g., Stuppel & Ball, 2008). Belief-first models assume that the heuristic system cues a rapid and intuitive response that will be accepted unless the analytic system generates an alternative response (e.g., the selective processing model; Evans, 2007). In contrast, reasoning-first models suggest that problems are first processed by the analytic system but that reasoners fall back on the heuristic system if they are lacking the required cognitive resources, such the mental model account of belief bias (Oakhill, Johnson-Laird, & Garnham, 1989) and Quayle and Ball's (2000) mental models-based account. Finally, parallel models suggest that both heuristic and analytic systems are working simultaneously (Handley, Newstead, & Trippas, 2011; Sloman, 1996).

Now, the dual strategy model that we have presented previously does not rely on the same distinctions as dual process models. It is a model of the way that people make inferences based on premises (i.e., a model of analytic thinking) and does not have an explicit heuristic component. However, it does allow a novel prediction about the relative effect of belief on reasoning. This is because while the empirical believability of a putative conclusion is not related to its logical validity, it is a component of a person's real-world knowledge. As such, it would be an intrinsic part of the larger statistical information accessed during reasoning with meaningful premises. Thus, people using a statistical strategy should show a stronger tendency to use conclusion believability as a component of their evaluation of conclusion likelihood (see Oaksford et al., 2000, for just such a prediction), compared to people using a counterexample strategy.

In order to examine this prediction, we adapted the technique used in Markovits et al. (2013) that enables distinguishing the two different forms of strategy. This involves presenting a series of affirmations of the consequent inferences with accompanying statistical information that makes the putative conclusion more or less likely. Reasoners who consistently reject the conclusion to these inferences are classed as using a counterexample strategy. Reasoners who reject the conclusion more often to the low probability inferences are classed as using a statistical strategy. We combined this with a set of syllogisms using the same basic structure but with putative conclusions that were either very believable or very unbelievable.

One further aim of these studies was to examine the information processing patterns associated with these reasoning strategies. In order to do so, we measured both response times and explicit feeling of rightness (FOR) ratings. The basic framework implies that people using a statistical strategy should take less time to make inferences. A previous study has shown that when severely time constrained, people will preferentially use such a strategy (Markovits et al., 2013), which certainly suggests that a statistical strategy is more rapid than a counterexample strategy. However, this is indirect evidence, whereas the current study afforded an opportunity for a direct test. Another open question concerns the way that contradictions between conclusion belief and reasoning based solely on problem premises is processed. Each problem set that is used in the following studies has inferences for which logical validity and conclusion belief are either contradictory or consistent. Differences in response times and FOR ratings between contradictory and consistent conclusions provide some indication of reasoners' ability to detect these differences (e.g., De Neys, 2012; Thompson & Johnson, 2014). The key question is whether statistical and counterexample reasoners have the same ability to do so.

In each of the following studies, we examine a (different) single form of reasoning. It should be noted that many

studies that examine the effects of belief attempt to combine different forms of reasoning, including valid and invalid forms, in order to look at the so-called belief by validity interaction (e.g., Evans et al., 1983). However, a recent study has shown that factors such as the complexity of reasoning influence the extent of belief-bias independently of validity (Brisson, de Chantal, Lortie Forgues & Markovits, 2014). Thus, we decided to conduct separate studies examining interactions between reasoning strategy and individual forms of syllogistic reasoning. These studies examine three different forms of syllogistic reasoning, going from more typical invalid and valid forms using *all* quantifiers, to a less frequently used invalid form based on a *some* quantifier.

Finally, in all of the following studies, we alternate the order of the diagnostic and the belief problems. Although there is a tendency to view the counterexample/statistical distinction as a form of individual difference, which would obviate the necessity for such a control, previous results have clearly shown that people will change strategies under certain conditions (Markovits et al., 2013). In addition, it has been shown that the two strategies do not simply map onto a more or less logical reasoner distinction (Markovits et al., 2016). Such results certainly suggest that strategy use does not correspond to a stable individual difference. Thus, while we make no clear hypothesis in this respect, alternating order allows us to examine the possibility that strategy use might be affected by processing problems with a range of familiar information, such as that used in the belief problems.

Study 1

In this study, the effect of conclusion belief was examined in the context of reasoning with invalid syllogisms, of the general form "All A are B. X is B. X is A." The specific forms used in this study were designed to have many potential counterexamples (i.e., classes of objects that are not A but are B). Half of these syllogisms had putative believable conclusions, while the other half had putative unbelievable conclusions. Given that all of the problems were invalid, those with believable conclusions would be conflict items, and those with unbelievable conclusions represent the consistent items. Reasoning strategies were measured with the strategy assessment problems used by Markovits et al. (2013). Participants are presented with a series of reasoning problems with fictitious content along with a description of the statistical pattern of observations, based on a method used by Geiger and Oberauer (2007). The order of the two problem sets was systematically varied.

Method

Participants A total of 170 college-level students (86 females, 84 males; average age = 20 years, 5 months) took part in this experiment. All participants were French-speaking students attending one of two colleges (Cégep) in Montréal and were volunteers.

Materials A computer program was created using Microsoft Visual Basic. The opening screen asked for demographic information. In the strategy-first order, participants were given a set of instructions that were identical to those used in Markovits et al. (2013). These explained that they were to be given information taken from scientific studies of a newly discovered planet. They were also told that they were to receive a sequence of arguments followed by a conclusion, and that their task was to indicate whether or not the conclusion was logically valid given the information presented.

The strategy-assessment problems presented the set of 13 problems used by Markovits et al. (2012). Each problem described a causal conditional relation involving nonsense terms or relations that included frequency information concerning the relative numbers of not-P.Q and P.Q cases out of 1,000 observations. Participants were then given an inference corresponding to the affirmation of the consequent inference (P implies Q, Q is true. Conclusion: P is true), and were asked to indicate whether the conclusion could be logically drawn from the premises or not. Of the 13 items, five had a relative frequency of alternative antecedents that was close to 10% (each individual item varied between 8% and 10%), five had a relative frequency that was close to 50% (each item varied between 48% and 50%), and three had a relative frequency of alternative antecedents that was presented as 0% (these last were presented in order to provide greater variability in problem types). The following is an example:

A team of geologists on Kronus have discovered a variety of stone that is very interesting, called a TrolYTE. They affirm that on Kronus, if a TrolYTE is heated, then it will give off Philoben gas.

Of the 1,000 last times that they have observed TrolYtes, the geologists made the following observations:

910 times Philoben gas has been given off, and the TrolYTE was heated.

90 times Philoben gas has been given off, and the TrolYTE was not heated

From this information, Jean reasoned in the following manner:

The geologists have affirmed that: If a TrolYTE is heated, then it will give off Philoben gas.

Observation: A TrolYTE has given off Philoben gas.

Conclusion: The TrolYTE was heated.

Participants were presented with two buttons, one which stated that the conclusion was invalid, while the other stated that the conclusion was valid, with the pointer initially placed between the buttons. They were asked to click on the appropriate button when they were certain of their response. For each inference, both responses and response times were recorded.

Following this initial problem set, participants were given a set of belief-bias problems. For these, they were told that they would be asked to make inferences based on familiar statements. As before, they would be given a rule, an observation, and a conclusion and would be asked to indicate whether the conclusion could be logically drawn from these. They were then presented with a sequence of syllogisms with four invalid (conflict) syllogisms with believable conclusions, four invalid (consistent) syllogisms with unbelievable conclusions, and three filler items. The filler items involved inferences of the form “All A are B. X are A. X are B.” These were presented in a single semirandom order, with no more than two successive items of the same kind.

The believable-conclusion syllogisms were (translated from the original French):

All humans breathe. Italians breathe. Italians are human.

All vehicles have motors. Automobiles have motors.

Automobiles are vehicles.

All soaps have a price. Bath gel has a price. Bath gel is a soap.

All computers use electricity. PCs use electricity. PCs are computers.

The unbelievable-conclusion syllogisms were:

All humans breathe. Birds breathe. Birds are human.

All automobiles have motors. Airplanes have motors.

Airplanes are automobiles.

All soaps have a price. Gasoline has a price. Gasoline is a soap.

All computers use electricity. Lighthouses use electricity. Lighthouses are computers.

For each problem, participants were presented with a single screen presenting both premises and the putative conclusion. At the bottom of the screen, they were presented with two buttons corresponding to the options that the conclusion is valid or not. Response time was measured from the initial presentation of the premises and the pressing of one of the two buttons. For each of these inferences, after the initial response, participants were also asked to rate their feeling of rightness (degree of confidence) on a 1 to 7 scale, with 1 corresponding to *guessing* and 7 corresponding to *certain that I am right* (Thompson, Prowse Turner, & Pennycook, 2011).

Half the participants first received the belief-bias problems followed by the strategy-assessment problems (belief-first

condition), while the other half received these problems in the opposite order (strategy-assessment-first condition).

Design Strategy-assessment problems were used as a between-subjects categorization method. Order of problem sets (belief first, strategy assessment first) was also a between-subjects variable. The dependent variable was conclusion acceptance for belief-bias syllogisms with problem type (believable conclusion, unbelievable conclusion) as a repeated measure. Note that this design was used in all three studies.

Procedure Participants were seen individually in a quiet room and performed the experiment on a portable computer.

Results and discussion

We first analyzed performance on the strategy-assessment problems. For each participant, we categorized reasoning patterns in the following way. Participants who rejected all of the 10% inferences and all of the 50% inferences were put into the *counterexample* category. Participants for whom the mean acceptance rate for the 10% items was greater than that for the 50% items were put into the *statistical* category. All other patterns of responses were put into the *other* category.

Table 1 gives the number of participants in each category as a function of Order. An initial examination showed no significant difference in the distribution of responses in the two conditions, $\chi^2(2) = 3.42, p = ns$, although as can be seen from the table, there is some tendency for the statistical strategy to be used more often when the belief-bias problems were presented first, before the strategy-assessment problems. It should be noted that the *other* category mostly corresponds to reasoners who reject the conclusion to the high and low probability items with the same frequency but are inconsistent in doing so. Such a pattern may be due to inconsistent use of either a statistical or a counterexample strategy. Since we cannot distinguish between these, the *other* patterns will be excluded from further analysis in this and in the following studies.

We then computed mean conclusion acceptances (out of four), mean FOR, and mean RT for both believable-conclusion and unbelievable-conclusion syllogisms (see

Table 1 Numbers of participants giving counterexample, statistical, or other patterns as a function of order, and mean acceptance rates for the 0%, 10%, and 50% items in Study 1

Pattern	Strategy first	Belief first
Counterexample	50	37
Statistical	36	47
Other	30	31

Table 2). We first examined conclusion acceptances. We conducted an ANOVA with number of conclusions accepted as dependent variable with problem type (believable conclusion, unbelievable conclusion) as a repeated measure and order and strategy as independent variables. The ANOVA gave main effects of strategy, $F(1, 168) = 30.20, p < .001, \eta_p^2 = .150$, and problem type, $F(1, 168) = 309.34, p < .001, \eta_p^2 = .649$, and a significant Strategy \times Problem Type interaction, $F(1, 168) = 19.92, p < .001, \eta_p^2 = .106$. No effect of order was found. The main effects indicated that believable conclusions were accepted more often than unbelievable ones and that those using a counterexample strategy rejected conclusions more often than those using a statistical strategy (see Table 2). The Strategy \times Problem Type interaction indicated that the difference between acceptance rates for believable-conclusion and unbelievable-conclusion syllogisms was greater for participants using a statistical strategy ($M = 2.47$) than those using a counterexample strategy ($M = 1.47$). In order to examine this in more detail, we calculated a belief score by subtracting the number of conclusion acceptances on the unbelievable-conclusion syllogisms from that on the believable-conclusion syllogisms. The mean effect of belief was greater than zero for participants with a statistical strategy ($M = 2.47$), $t(82) = 15.08, p < .001, d = 1.66$, and for those with a counterexample strategy ($M = 1.47$), $t(86) = 9.66, p < .001, d = 1.04$, with the former being greater than the latter, $F(1, 168) = 19.92, p < .001, \eta_p^2 = .106$.

We then performed an ANOVA with RT as dependent variable, with problem type as a repeated measure and strategy and order as independent variables. This gave a significant effect of problem type, $F(1, 168) = 5.33, p < .05, \eta_p^2 = .032$. Participants took longer to respond to believable-conclusion problems ($M = 18.89$ s) than to unbelievable-conclusion problems ($M = 17.13$ s). In other words, participants took longer to respond when believability conflicted with validity than when it did not. It should also be noted that participants using a counterexample strategy tended to take more time ($M = 18.90$ s) than those using a statistical strategy ($M = 17.10$ s), although this difference was not significant, $F(1, 168) = 2.55, p < .12$.

We performed an ANOVA, with mean FOR as a dependent variable and with problem type as a repeated measure and strategy and order as independent variables. This gave only a significant interaction involving Strategy \times Problem Type, $F(1, 168) = 4.30, p < .05, \eta_p^2 = .027$. Post hoc analysis of the interaction was done using the Tukey test, with $p = .05$. This showed that mean FOR on the unbelievable-conclusion syllogisms (nonconflict) was significantly lower for statistical reasoners than for counterexample reasoners, with no difference on the believable-conclusion syllogisms (conflict).

The results of this initial study indicate that participants who used a counterexample strategy on problems with completely unfamiliar content and explicit statistical

Table 2 Mean number of conclusions accepted (out of four), mean FOR ratings, and mean RT (in seconds) for the unbelievable conclusion and believable conclusion syllogisms for participants using a counterexample or a statistical strategy in Study 1

Strategy	Believable conclusion			Unbelievable conclusion		
	Conclusion acceptance	FOR	RT	Conclusion acceptance	FOR	RT
Counterexample	1.71	5.97	20.11	0.24	6.01	17.63
Statistical	3.01	5.87	17.61	0.54	5.65	16.60

information showed a much smaller effect of conclusion believability on invalid syllogisms than those who used a statistical strategy on the unfamiliar problems. The RT data indicated that both groups were sensitive to conflict, although in the case of the FOR data, the pattern was less clear.

Study 2

The results of the first study are certainly consistent with our general framework. However, the problems used to establish strategy use and those employed to examine the effects of conclusion belief had very similar logical structures. We thus decided to replicate the initial study using syllogisms for which the logically correct response was one of certainty. Accordingly, we used syllogisms with a logical structure corresponding to the Modus Tollens (MT) inference (All X are Y, Z is not Y). These have a single logical conclusion (Z is not Y) but are relatively difficult even for adults. Because all of the inferences are valid, those with unbelievable conclusions were the conflict items and those with believable conclusions were the consistent items.

Method

Participants A total of 189 college-level students (103 females, 86 males; average age = 22 years, 3 months) took part in this experiment. All participants were French-speaking students attending a college (Cégep) in Montréal and were volunteers.

Materials The materials used in this study were identical to those of the initial study, with the exception of the problems used to measure the effects of belief. The syllogisms used had a single valid conclusion. In this case, there were five syllogisms with unbelievable conclusions, five syllogisms with believable conclusions, and two filler syllogisms.

The unbelievable-conclusion syllogisms were:

All vehicles have wheels. Snowmobiles do not have wheels. Snowmobiles are not vehicles.

All birds fly. Ostriches do not fly. Ostriches are not birds.

All fruits are sweet. Lemons are not sweet. Lemons are not fruit.

All trees have leaves. Fir trees do not have leaves. Fir trees are not trees.

All animals have fur. Fish do not have fur. Fish are not animals.

The believable-conclusion syllogisms were:

All vehicles have wheels. Computers do not have wheels. Computers are not vehicles.

All birds fly. Cows do not fly. Cows are not birds.

All fruits are sweet. Broccoli is not sweet. Broccoli is not a fruit.

All trees have leaves. Radios do not have leaves. Radios are not trees.

All animals have fur. Tables do not have fur. Tables are not animals.

Procedure Participants were seen individually in a quiet room and performed the experiment on a portable computer.

Results and discussion

As before, an initial examination showed no significant difference in the distribution of responses in the two conditions, $\chi^2(2) = 2.92, p = ns$, although, as in the initial study, there is some tendency for the statistical strategy to be used more often when the belief-bias problems were presented before the diagnostic problems (this will be examined in the discussion). Also, as previously, for the subsequent analyses, we eliminated participants in the *other* category. We then calculated mean conclusion acceptance, mean RT, and mean FOR for unbelievable-conclusion and believable-conclusion syllogisms as a function of strategy (see Table 3).

We first examined the effect of belief on responses. We conducted an ANOVA with number of conclusions accepted as a dependent variable, with problem type as a repeated measure, and order and strategy as independent variables. This showed a main effect of problem type, $F(1, 153) = 33.92, p < .001, \eta_p^2 = .171$, and an interaction involving Problem Type \times Strategy, $F(1, 153) = 16.16, p < .001, \eta_p^2 = .095$. An analysis of interactions was done using the Tukey test, with $p = .05$. As expected, more believable than unbelievable

Table 3 Mean number of conclusions accepted (out of four), mean FOR ratings, and mean RT (in seconds) for the believable conclusion and unbelievable conclusion syllogisms for participants using a counterexample or a statistical strategy in Study 2

Strategy	Believable conclusion			Unbelievable conclusion		
	Conclusion acceptance	FOR	RT	Conclusion acceptance	FOR	RT
Counterexample	3.05	5.96	25.2	2.79	5.72	24.4
Statistical	3.28	5.86	19.5	1.89	5.36	20.4

conclusions were accepted. Replicating Study 1, the difference between believable and unbelievable conclusions was larger for those using a statistical strategy ($M = 1.38$) than a counterexample strategy ($M = 0.25$; see Table 3). Mean effect of belief was greater than zero for participants with a statistical strategy ($M = 1.38$), $t(75) = 6.94$, $p < .01$, $d = .80$, but not for those with a counterexample strategy ($M = 0.25$), $t(78) = 1.28$, $p = ns$.

We then performed an ANOVA with RT as a dependent variable, with problem type as a repeated measure, and strategy and order as independent variables. This gave a significant effect of strategy, $F(1, 153) = 10.05$, $p < .01$, $\eta_p^2 = .050$, and order, $F(1, 153) = 7.89$, $p < .01$, $\eta_p^2 = .050$. Participants using a counterexample strategy took longer to respond ($M = 24.81$) than those using a statistical strategy ($M = 19.92$). In addition, mean RT was greater when the strategy assessment problems were given initially ($M = 24.84$) than in the inverse order ($M = 19.96$).

We then performed an ANOVA with FOR as a dependent variable, with problem type as a repeated measure, and strategy and order as independent variables. This gave a significant effect of problem type, $F(1, 153) = 15.66$, $p < .001$, $\eta_p^2 = .096$. Mean FOR was greater for believable-conclusion syllogisms ($M = 5.91$) than for unbelievable-conclusion syllogisms ($M = 5.54$).

In this study, we replicated the initial study with syllogisms that had logically valid conclusions. The overall pattern of results was very similar. Participants using a counterexample strategy took significantly more time and were less affected by the believability of conclusions than those using a statistical strategy. The FOR data indicated that both groups were sensitive to conflict; however, the RT showed no difference between the consistent and conflict problems.

Study 3

The results of the two initial studies show quite similar patterns. Overall, the effect of belief is stronger for statistical than for counterexample reasoners, while the latter generally take more time to make inferences. Interestingly, both relative RT and FOR ratings across consistent and conflict syllogisms are fairly similar for statistical and counterexample reasoners. This is particularly interesting, since it suggests that both

strategies, despite global processing differences, are sensitive to a conflict between logical validity and believability. However, both of the syllogisms used in these studies are of the form “All X are Y” and required relatively little time to process. Giving a more complex form of reasoning would induce greater variability and allow a clearer examination of the relationship between metacognitive processing and reasoning strategy. In order to do this, we examined syllogisms of the form “Most X are Y. Most Z are Y. Most X are Z?”, which pretesting suggested would be quite difficult to process. Because these conclusions are invalid, the believable conclusions were the conflict items and the unbelievable conclusions are the nonconflict items.

Method

Participants A total of 121 college-level students (71 females, 50 males; average age = 21 years, 9 months) took part in this experiment. All participants were French-speaking students attending a college (Cégep) in Montréal and were volunteers.

Materials The materials used in this study were identical to those of the initial study, with the exception of the problems used to measure the effects of belief. In this case, there were five (invalid) syllogisms with believable conclusions, five syllogisms with unbelievable conclusions, and two filler syllogisms.

The believable-conclusion syllogisms were:

Most poodles have a price. Most friendly animals have a price. Most poodles are friendly animals.

Most cars have valves. Most vehicles with four doors have valves. Most cars are vehicles with four doors.

Most cats breathe. Most animals with fur breathe. Most cats are animals with fur.

Most tables are solid. Most things made of wood are solid. Most tales are made of wood.

Most trees have roots. Most plants with leaves have roots. Most trees are plants with leaves.

The unbelievable-conclusion syllogisms were:

Most soaps have a price. Most dogs have a price. Most soaps are dogs.

Most cars have valves. Most airplanes have valves.
 Most cars are airplanes.
 Most cats breathe. Most birds breathe. Most cats are birds.
 Most tables are solid. Most chairs are solid. Most tables are chairs.
 Most trees have roots. Most tulips have roots. Most trees are tulips.

Procedure Participants were seen individually in a quiet room and performed the experiment on a portable computer.

Results and discussion

As before, an initial examination showed no significant difference in the distribution of responses in the two conditions, $\chi^2(2) = 4.59, p = ns$, with, again, a small tendency for the statistical strategy to be used more often when the belief-bias problems were presented first. Also, as previously, for the subsequent analyses, we eliminated participants in the *other* category. We then calculated mean conclusion acceptance, mean RT, and mean FOR for the believable-conclusion syllogisms and the unbelievable-conclusion syllogisms as a function of Strategy (see Table 4).

We first examined the effect of belief on responses. We conducted an ANOVA with number of conclusions accepted as a dependent variable, with problem type as a repeated measure, and order and strategy as independent variables. This showed a main effect of strategy, $F(1, 99) = 12.11, p < .001, \eta_p^2 = .109$, and problem type, $F(1, 99) = 131.36, p < .001, \eta_p^2 = .570$, and an interaction involving Problem Type \times Strategy, $F(1, 99) = 7.23, p < .01, \eta_p^2 = .068$. As in Studies 1 and 2, more believable than unbelievable conclusion were accepted, and those using a counterexample strategy accepted fewer conclusions than those using a statistical strategy. An analysis of interactions was done using the Tukey test, with $p = .05$. The interaction indicated that the difference between believable and unbelievable conclusions was larger for those using a statistical ($M = 1.96$) rather than a counterexample ($M = 1.21$) strategy (see Table 4). Mean effect of belief was greater than zero for participants with a statistical strategy, $t(55) = 9.87, p <$

$.001, d = 1.33$, and for those with a counterexample strategy, $t(46) = 6.65, p < .001, d = 0.98$.

We then performed an ANOVA with RT as a dependent variable, with problem type as a repeated measure, and strategy as an independent variable. This gave a significant effect of problem type, $F(1, 99) = 76.45, p < .001, \eta_p^2 = .433$, and interactions involving Problem Type \times Strategy, $F(1, 99) = 5.56, p < .05, \eta_p^2 = .053$, and Strategy \times Order, $F(1, 99) = 7.57, p < .01, \eta_p^2 = .071$. Participants took much longer to respond to believable-conclusion syllogisms ($M = 31.7$) than to unbelievable-conclusion syllogisms ($M = 19.6$), meaning that they took less time to process consistent than conflict items. Analysis of the Strategy \times Order interaction showed that when the belief-bias problems were presented first, counterexample reasoners took significantly more time ($M = 27.7$) than did statistical reasoners ($M = 19.8$), with no significant difference in the opposite order (counterexample = 25.4; statistical = 30.4). An analysis of the Problem Type \times Strategy interactions did not show any individual significant differences; however, the fact that the interaction was significant means that the RT difference for counterexample reasoners ($M = 15.7$) was larger than for statistical reasoners ($M = 9.0$).

We then performed an ANOVA with FOR as a dependent variable, with problem type as a repeated measure, and strategy and order as an independent variable. This gave a significant effect of problem type, $F(1, 101) = 368.36, p < .001, \eta_p^2 = .783$, and an interaction involving Problem Type \times Strategy, $F(1, 101) = 9.02, p < .01, \eta_p^2 = .082$. Mean FOR was greater for unbelievable-conclusion syllogisms ($M = 6.2$) than for believable-conclusion syllogisms ($M = 4.37$), again showing that FORs were higher for the consistent than for the conflict items. Analysis of the interaction did not show any individual significant differences; however, the fact that the interaction was significant means that the FOR differences for the counterexample reasoners ($M = 2.14$) was larger than for the statistical reasoners ($M = 1.57$), and conversely that the difference for the unbelievable conclusions (+.31) was different than the believable ones (-26).

As expected, these problems showed a great deal more variability between believable and unbelievable problems in both response time measures and in FOR ratings. Despite this, the effect of belief remained stronger for statistical reasoners than for counterexample reasoners. There was, however, a clear order effect on response time measures. Statistical

Table 4 Mean number of conclusions accepted (out of four), mean FOR ratings, and mean RT (in seconds) for the believable conclusion and unbelievable conclusion syllogisms for participants using a counterexample or a statistical strategy in Study 3

Strategy	Believable conclusion			Unbelievable conclusion		
	Conclusion acceptance	FOR	RT	Conclusion acceptance	FOR	RT
Counterexample	1.32	4.23	34.2	0.10	6.37	18.5
Statistical	2.36	4.49	29.6	0.39	6.06	20.6

reasoners took less time than counterexample reasoners, but only when receiving the belief bias problems initially. Participants using both strategies took much more time to make inferences when conclusions were believable, that is, in a situation when conclusion belief contradicted logical validity, than when conclusions were unbelievable. Similarly, FOR ratings significantly decreased for the former syllogisms. In other words, both strategies appeared to be sensitive to the presence of conflict between validity and conclusion belief, although the effect was stronger for the counterexample group.

Looking across the three studies, one can see that the effect of conflict on RTs and FORs is larger in this study than in Studies 1 and 2, where the effects were smaller and sometimes nonsignificant. We note that although the RTs were quite a bit longer in the current study, the rate of correct responding was, if anything, higher. Although this must remain speculative, one possible explanation for this pattern is that the internal indices of conflict between reasoning and belief are stronger when reasoning is easier (Handley & Trippas, 2015). In this situation, the ability of people to monitor the conflict increases (as shown by the increased RT and decreased FOR) as does the ability to overcome it, which would then result in a decrease of the effect of conclusion belief. Such an analysis is indeed consistent with recent results relating the extent of belief bias to reasoning difficulty (Brisson et al., 2014).

General discussion

Understanding the sources of variability in reasoning and their interactions is a critical problem. The results of these studies are novel in several ways. First, they provide clear support for the basic hypothesis that the form of the inferential strategy used by reasoners will generate a differential level of sensitivity to conclusion believability. Statistical strategies involve the integration of stored empirical evidence in order to generate an estimate of the likelihood that a putative conclusion is true. Since such a mode of reasoning requires full consideration of relevant knowledge, of which conclusion believability is one important source, this would be expected to have a relatively strong influence on people's conclusions (see also Oaksford, Chater & Larkin, 2000, for a similar analysis). By contrast, counterexample strategies focus on potential counterexamples to putative conclusions and would place less emphasis on believability. The results of these studies clearly show that the relative impact of conclusion belief is indeed stronger among reasoners using a statistical strategy than among those using a counterexample strategy. This result was found with three very different forms of reasoning, both with invalid syllogisms, based on all and some quantifiers (Study 1 and 3) and valid syllogisms (Study 2). The broad range of reasoning

examined in these studies makes it probable the basic results would generalize to other forms of reasoning.

Second, these results provide direct evidence that statistical reasoners make more rapid inferences than counterexample reasoners. This was true for all three studies, with the sole exception being the order effect in Study 3 (we will examine the implications of order effects later).

Consistent with much previous work (see De Neys, 2012, 2014, for review), reasoners in all three studies were sensitive to the conflict between logical validity and believability as indexed either by FOR or RT (although we note that not all of the comparisons were significant). Particularly relevant was the fact that this was true of both statistical and counterexample reasoners. These data suggest that both sets of reasoners have similar capacity to detect contradictions between logical validity and conclusion belief. An important conclusion to draw from these findings is that even those who rely on a Bayesian or probabilistic approach are nonetheless processing logical relations; if they were not, we would expect them not to show sensitivity to conflict. These findings are not consistent with the assumption that reasoners rely solely on a probabilistic assessment of either the premises or the conclusion, although it is not necessary to assume that the processing of logical relations is done explicitly (De Neys, 2012; Trippas, Handley, Verde, & Morsanyi, 2016).

Actually, it is useful to distinguish between reasoning from premises to a conclusion and logical validity, which is a formal property of a syllogism. When given two premises, reasoners must reason with these in order to generate a conclusion (which may or may not be valid). The relative believability of such a conclusion is, however, the result of a simple empirical judgment based on stored beliefs. Our results are consistent with the idea that the conflict between these two processes is felt by some sort of metacognitive evaluation (Thompson & Johnson, 2014; Thompson, Morley, & Newstead, 2011), which, in this case, is indicated by both comparative RT and FOR ratings and that this was true for both counterexample and statistical strategies. The differential effect of conclusion belief might be at least partially due to the way that the conflict between reasoning from premises and conclusion belief can be resolved. For example, statistical reasoning would produce an estimate of the likelihood of a putative conclusion, which could be more easily modulated by incorporating conclusion belief than the more dichotomous judgment that is produced by counterexample reasoners. We note one other finding that supports differences in how the two groups process these problems. In all three studies, FORs for nonconflict problems were higher in the counterexample group than for the statistical group (although the interaction was significant only in Studies 1 and 3). Thus, when premise-based reasoning and belief-based reasoning converge, it seems to add to a sense of confidence in a way that does not happen for the statistical reasoners. Again, this is speculative, and

invites further experimental work to examine differences in processing strategies between the two groups of reasoners.

One further component of these results concerns the effects of order found in these studies, such that people appeared more inclined to use a statistical strategy if they completed the belief-bias problems first. As we stated in the introduction, the dual strategy model is not a pure individual difference approach. Rather, it suggests that reasoners have access to both statistical and counterexample strategies, with the choice of strategy determined by factors such as cognitive constraints (Markovits et al., 2013). More specifically, our basic hypothesis claims that reasoners using a statistical strategy will be more open to a wider array of information, including conclusion belief. Inversely, it is possible that reasoners who are exposed to more information will tend to use a statistical strategy more often. To examine this, we combined strategy use across all three studies and examined relative use of statistical and counterexample strategies (eliminating the *other* category). Overall, the relative proportion of statistical strategies increased when the belief-bias problems were given initially (55.7% to 43.5%), $\chi^2(1) = 6.33, p < .02$. This indicates that initially reasoning with familiar premises with believable or unbelievable conclusions increased use of a statistical strategy, which is quite consistent with the dual-strategy framework. This, in turn, indicates that reasoning with familiar premises increases the tendency of reasoners to base their inferences on direct consideration of the statistical pattern provided by stored knowledge. The RT data generally also support this interpretation, although they are a bit more complex. The general pattern is consistent with the idea that when people initially reason with the belief-bias problems, they tend to reason faster, which is in turn consistent with the increased tendency to reason statistically, as noted above. In Study 1, with the shortest RT of the three studies, there was no observable order effect. In Study 2, with longer RT than Study 1, there is a clear order effect, with reasoning first with belief-bias problems generating faster RTs. Study 3, with more a complex variation in RT and FOR measures, shows a concomitantly more complex pattern, with statistical reasoners taking less time than counterexample reasoners when the belief-bias problems are presented first, but not in the opposite condition. One way to think about this is that performance, overall, was quite high in Study 3, which, as we have argued, might facilitate conflict detection. Overall, this would produce longer RTs, which might be undone by the combination of a statistical strategy and receiving the belief-bias problems first. In other words, both the forms of order effect observed in these studies are consistent with the idea that reasoning with familiar premises generates related tendencies to both use a statistical strategy and to reason more rapidly.

In sum, our data provide further evidence for the dual strategy model, in that there were clear differences in belief bias and other measures between those using a counterexample

and statistical strategy. They also suggest another factor that can account for some observed differences in susceptibility to belief bias effects. The nature of these differences has been a somewhat contentious subject. For example, response time studies have shown that reasoners who are more susceptible to belief-bias effects take less time to make inferences (J. S. B. Evans & Curtis-Holmes, 2005; Stuppel, Ball, Evans, & Kamal-Smith, 2011; Thompson, Morley, & Newstead, 2011). One interpretation of these results is that faster reasoners are using a simple heuristic form of reasoning. The results of the present studies suggest an alternative explanation of these results. For example, when reasoners are given reduced time to make inferences (J. S. B. Evans & Curtis-Holmes, 2005), they tend to use a statistical strategy (Markovits et al., 2013), which in turn produces stronger effects of belief bias. Similarly, statistical reasoners tend to reason faster than counterexample reasoners and are more susceptible to belief-bias effects, which could account for many of the observed individual differences. The interaction between heuristic responding and the kind of statistical strategy identified in the dual strategy model remains an open, and important, question within the context of dual process theories of belief bias (e.g., J. S. B. Evans, 2007).

Acknowledgements This study was supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada to H.M.

References

- Barrouillet, P., Gauffroy, C., & Lecas, J.-F. (2008). Mental models and the suppositional account of conditionals. *Psychological Review*, *115*(3), 760–771.
- Brisson, J., de Chantal, P.-L., Lortie Forgues, H., & Markovits, H. (2014). Belief bias is stronger when reasoning is more difficult. *Thinking and Reasoning*, *20*(3), 385–403.
- Byrne, R. M. J., Espino, O., & Santamaria, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory and Language*, *40*, 347–373.
- Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition*, *23*(5), 646–658.
- Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, *19*(3), 274–282.
- De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science*, *7*(1), 28–38.
- De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning*, *20*(2), 169–187.
- Evans, J. S. B. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning*, *13*, 321–339.
- Evans, J. S. B., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*(3), 295–306.
- Evans, J. S. B., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, *11*(4), 382–389.

- Evans, J. S. B., Newstead, S. E., Allen, J. L., & Pollard, P. (1994). Debiasing by instruction: The case of belief bias. *European Journal of Cognitive Psychology*, 6(3), 263–285.
- Evans, J. S. B. T., & Over, D. E. (2004). *If*. New York: Oxford University.
- Evans, J. S. B., Over, D. E., & Handley, S. J. (2005). Suppositionals, extensionality, and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002). *Psychological Review*, 112, 1040–1052.
- Evans, J. S. B., Thompson, V. A., & Over, D. E. (2015). Uncertain deduction and conditional reasoning. *Frontiers in psychology: Cognition*. doi:10.3389/fpsyg.2015.00398
- Fernbach, P. M. & Erb, C. D. (2013). A quantitative causal model theory of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(5), 1327.
- Geiger, S. M., & Oberauer, K. (2007). Reasoning with conditionals: Does every counterexample count? It's frequency that counts. *Memory & Cognition*, 35(8), 2060–2074.
- 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(1), 28–43. doi:10.1037/a0021098
- Handley, S. J., & Trippas, D. (2015). Chapter two—dual processes and the interplay between knowledge and structure: a new parallel processing model. *Psychology of Learning and Motivation*, 62, 33–58.
- Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences*, 5(10), 434–442.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove: Erlbaum.
- Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics and inference. *Psychological Review*, 109, 646–678.
- Markovits, H. (1984). Awareness of the “possible” as a mediator of formal thinking in conditional reasoning problems. *British Journal of Psychology*, 75(3), 367–376.
- Markovits, H., & Barrouillet, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review*, 22(1), 5–36.
- 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(8), 1208–1215. doi:10.3758/s13421-015-0535-1
- Markovits, H., Brisson, J., & de Chantal, P.-L. (2016). Logical reasoning versus information processing in the dual strategy model of reasoning. *Journal of Experimental Psychology: Learning, Memory and Cognition*. Advance online publication.
- 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*, 49(6), 1213–1222.
- Markovits, H., Lortie Forgues, H., & Brunet, M.-L. (2012). More evidence for a dual-process model of conditional reasoning. *Memory & Cognition*, 40(5), 736–747.
- Markovits, H., & Nantel, G. (1989). The belief-bias effect in the production of 'logical' conclusions. *Memory and Cognition*, 17(1), 11–17.
- Markovits, H., & Vachon, R. (1990). Conditional reasoning, representation, and level of abstraction. *Developmental Psychology*, 26(6), 942–951.
- Oakhill, J., Johnson-Laird, P. N., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition*, 31(2), 117–140. doi:10.1016/0010-0277(89)90020-6
- Oaksford, M., & Chater, N. (2007). *Bayesian rationality*. Oxford: Oxford University Press.
- Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(4), 883–899.
- Quayle, J. D., & Ball, L. J. (2000). Working memory, metacognitive uncertainty, and belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology: Section A*, 53(4), 1202–1223. doi:10.1080/713755945
- Singmann, H., Klauer, K. C., & Over, D. (2014). New normative standards of conditional reasoning and the dual-source model. *Frontiers in Psychology*, 5, 316. doi:10.3389/fpsyg.2014.00316
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119(1), 3–22. doi:10.1037/0033-2909.119.1.3
- Stuppelle, E. J. N., & Ball, L. J. (2008). Belief–logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. *Thinking & Reasoning*, 14(2), 168–181. doi:10.1080/13546780701739782
- Stuppelle, E. J. N., 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(8), 931–941.
- Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition*, 22(6), 742–758.
- Thompson, V. A. (2000). The task-specific nature of domain-general reasoning. *Cognition*, 76(3), 209–268.
- Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking and Reasoning*, 20(2), 215–244.
- Thompson, V. A., Morley, N. J., & Newstead, S. E. (2011). Methodological and theoretical issues in belief-bias: Implications for dual process theories. In K. I. Manktelow, D. E. Over, & S. Elqayam (Eds.), *The science of reason: A festschrift for Jonathan St. B. T. Evans* (pp. 309–338). Hove: Psychology Press.
- Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, 63(3), 107–140.
- 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(9), 1448.
- Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005a). A dual-process specification of causal conditional reasoning. *Thinking & Reasoning*, 11(3), 239–278.
- Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005b). Everyday conditional reasoning: A working memory-dependent tradeoff between counterexample and likelihood use. *Memory & Cognition*, 33(1), 107–119.