

Additional evidence for a dual-strategy model of reasoning: Probabilistic reasoning is more invariant than reasoning about logical validity

Henry Markovits¹ · Janie Brisson¹ · Pier-Luc de Chantal¹

Published online: 7 July 2015
© Psychonomic Society, Inc. 2015

Abstract One of the major debates concerning the nature of inferential reasoning is between counterexample-based strategies such as mental model theory and the statistical strategies underlying probabilistic models. The dual-strategy model proposed by Verschueren, Schaeken, and d’Ydewalle (2005a, 2005b) suggests that people might have access to both kinds of strategies. One of the postulates of this approach is that statistical strategies correspond to low-cost, intuitive modes of evaluation, whereas counterexample strategies are higher-cost and more variable in use. We examined this hypothesis by using a deductive-updating paradigm. The results of Study 1 showed that individual differences in strategy use predict different levels of deductive updating on inferences about logical validity. Study 2 demonstrated no such variation when explicitly probabilistic inferences were examined. Study 3 showed that presenting updating problems with probabilistic inferences modified performance on subsequent problems using logical validity, whereas the opposite was not true. These results provide clear evidence that the processes used to make probabilistic inferences are less subject to variation than those used to make inferences of logical validity.

Keywords Reasoning · Deductive reasoning · Mental models · Probabilistic reasoning

The ability to make deductive inferences is one of the most striking examples of advanced human cognition. Deductive inferences require starting from premises that are assumed to be true and generating conclusions that can be derived from these premises. The standard rules of logic generally assume that the inferences that can be derived depend only on the syntactic structure of the premises. However, many studies have shown that the inferences that people make vary systematically according to the specific content of the premises (e.g., Cummins, Lubart, Alksnis, & Rist, 1991; Markovits & Vachon, 1990; Thompson, 1994). Attempting to understand how people make deductive inferences and why these should show such variability is one of the more urgent tasks of theories of reasoning.

Such variability underlies one of the principal debates about the nature of inferential reasoning. On the one hand, probabilistic theories consider that people’s inferences generate estimations of the likelihood of a given conclusion, with such estimates reflecting stored statistical knowledge about the premises (e.g., Evans, Over, & Handley, 2005; Oaksford & Chater, 2007). Variability related to content can be readily explained by the effect of stored knowledge on likelihood estimations. In addition, such models allow for a process whereby additional information can be used to modify these estimates by Bayesian updating. Such models thus can easily model both the content-related variability of human reasoning and its nonmonotonic character. When asked to make a deductive inference, people will transform their estimation of a conclusion’s likelihood into a dichotomous judgment of validity.

A second category of model focuses particularly on the use of information to generate potential counterexamples. The most influential of these is *mental model theory* (Johnson-Laird, 2001; Johnson-Laird & Byrne, 2002). Although there are variants, the basic underlying principle is that people

✉ Henry Markovits
Henrymarkovits@gmail.com

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

construct internal models (representations) of the premises. If there are counterexamples to a putative conclusion in these models, this conclusion will be considered to be invalid. The nonmonotonic character of reasoning can be explained by the incorporation of additional information into this internal representation, via pragmatic or semantic factors (Johnson-Laird & Byrne, 2002; Markovits & Barrouillet, 2002). If such additional information generates a counterexample, a conclusion that was previously considered to be valid will be considered to be invalid. Content-related variation can be explained by similar processes.

Although both theories have attempted to propose themselves as unitary frameworks for understanding both inferences concerning logical validity and probabilistic inferences, there is increasing evidence that when people are asked to make inferences about validity, they can use a combination of both of these strategies. Such a model was proposed by Verschueren, Schaeken, and d'Ydewalle (2005a, 2005b), with a counterexample strategy being conceived of as a higher-level process requiring more working memory capacity, and a more intuitive statistical strategy as requiring fewer cognitive resources.

Recent results have confirmed and extended this model in several important directions. First, a method to identify which of the two strategies are used when making judgments of logical validity was determined (Markovits, Lortie-Forgues, & Brunet, 2012). Using this method, it has been shown that reasoners who are asked to make an inference concerning logical validity will preferentially use a statistical strategy when they are time-constrained, but will change to a counterexample strategy when allowed more time (Markovits, Brunet, Thompson, & Brisson, 2013). This suggests that the use of the two strategies is not only related to individual differences, but that the strategies correspond to two differing ways of making logical inferences that are accessible to the same individual, depending on such factors as cognitive constraints and the ways that inferential problems are presented (Markovits, Lortie-Forgues, & Brunet, 2010). The fact that a statistical strategy is used more often under time constraint when making logical deductions is consistent with the idea that it is less cognitively complex than a counterexample strategy.

Now, these studies have examined inferential reasoning requiring deductions concerning logical validity. More convincing evidence that the underlying strategies do indeed correspond to fundamentally different processes would require explicit comparisons between inferences requiring deductions about logical validity and explicitly probabilistic inferences. A recent study has used an inferential-updating paradigm in order to make one such comparison (Markovits, Brisson, & de Chantal, 2015). The basic paradigm of this study required asking people to evaluate an initial abstract inference either as being logically valid or explicitly probabilistically. A

subsequent problem presents the same inference with the additional observational information that out of 1,000 observed cases, 950 of these are consistent with the putative conclusion being true, and only 50 are consistent with the conclusion not being true. The results from this paradigm show that this information results in a significant decrease in the proportion of conclusion acceptances when reasoners evaluate logical validity, but has no effect when reasoners make explicitly probabilistic inferences. In other words, updating information is treated qualitatively differently when making an inference of logical validity than when making a probabilistic inference. Combined with the previous results, this provides a strong basis for the conclusion that the statistical and counterexample strategies are indeed qualitatively different.

In the following studies, we examined a more specific prediction derived from the dual-strategy model. The original formulation of this model specifically suggested that the statistical strategy most often used was a rapid, intuitive form of inference (Verschueren et al., 2005a, 2005b). It should be noted that there is no a priori reason why this should be the case, since it is very possible to envisage a statistical strategy that uses a cognitively complex process to evaluate probabilities (see, e.g., the mental model theory of probabilistic inference: Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). The previously described results show that when people are asked to make an inference about logical validity but are time-constrained, they tend to deploy a statistical strategy. If they have more time, they tend to use a counterexample strategy more often. This is certainly consistent with the idea that the statistical strategy does indeed correspond to an intuitive process. However, stronger evidence is needed before this postulate can be accepted.

Our basic hypothesis is that statistical strategies do indeed correspond to a rapid, intuitive form of likelihood evaluation. Our subsequent reasoning is as follows: First, consistent with previous results (Markovits et al., 2013), we assume that, when making deductions about logical validity with no time constraint, people will tend to use a relatively high proportion of the counterexample strategy more often than an intuitive statistical one. Similarly, we assume that when making explicitly probabilistic inferences, people will strongly tend to use an intuitive statistical strategy. More specifically, we assume that people are capable of intuitively processing the direct kind of explicit statistical information used in the updating problems to revise their judgment of the likelihood of a potential conclusion being true. Such an intuitive strategy will thus be deployed easily, with little variation. By contrast, updating judgments of logical validity will more strongly rely on a counterexample strategy, which requires a more explicit representation of the information presented in problems, and thus is more prone to variability. We can thus make the general prediction that reasoning about

logical validity will exhibit more variability than explicitly probabilistic reasoning.

Study 1

We first examined a fairly direct prediction derived from the previously discussed study on deductive updating (Markovits et al., 2015). Specifically, the results concerning the effects of updated statistical information on inferences of logical validity have shown that updated information indicating a high (but not certain) likelihood of a conclusion being true generates a clear decrease in such inferences. These results nonetheless show a relatively high rate of acceptance of the logical validity of the updated conclusion (varying from 20 % to 35 %). Theoretically, if this kind of updated information were uniformly processed with a counterexample strategy, this rate should be close to zero. One straightforward explanation for this discrepancy is that reasoners who preferentially use a statistical strategy when making simple logical inferences will also tend to process the explicit statistical updating information in the form of likelihood estimates instead of counterexamples. We thus presented participants with two sets of reasoning problems, all of which required judgments about logical validity. The first was the set of AC inferences (“if P then Q, Q is true”) with accompanying statistical information that had been used in previous studies to distinguish between people using a counterexample strategy and those using a probabilistic strategy (Markovits et al., 2012), which we refer to as the *assessment problems*. The second set comprised two high-probability updating problems based on the AC and DA inferences, taken from Markovits et al. (2015). These presented an initial inference, followed by an updated inference with statistical information showing that, of 1,000 observations, 980 showed both the minor premise and the putative conclusion as being true, whereas 20 had the minor premise being true, while the putative conclusion was false. Our basic predictions rely on two sets of results from previous studies. First, previous results had shown that people can use both counterexample and statistical strategies, but can change strategies in different conditions (Markovits et al., 2013). Thus, the assessment method used distinguishes individual tendencies when making inferences of logical validity. Thus, someone who has been classified as a counterexample reasoner would have a strong chance of using this strategy in a subsequent set of reasoning problems, but would have some chance of using a statistical strategy also, with the opposite pattern for someone classified as a statistical reasoner. Second, the use of a consistent counterexample strategy would result in a strong decrease in levels of acceptance of updated conclusions, whereas the use of a statistical strategy would generate relatively equal levels of acceptance of updated conclusions with the parameters used here (Markovits et al., 2015). Thus, we predicted

that the difference between the initial acceptance and the updated acceptance rates would be greater for people classified as using a counterexample strategy on the assessment problems than for those classified as using a probabilistic strategy.

Method

Participants A total of 107 college (Cégep) students (51 males, 56 females; average age = 21 years, 5 months) took part in this experiment. The students were native French speakers and volunteers.

Material Four paper-and-pencil booklets were prepared. On the first page of each booklet, participants were asked to give basic demographic information. Following this, they were given the following instructions (translated from the original French):

Imagine that a team of scientists are on an expedition on a recently discovered planet called Kronus. On the following pages, we will ask you to answer the question about phenomena that are particular to this planet. For each problem, you will be given a rule of the form if . . . then that is true on Kronus according to the scientists. It is very important that you suppose that each rule that is presented is always true. You will then be given additional information and a conclusion that you must evaluate. In the first booklet, participants were asked to make only deductive inferences. They were given two series of inferences, one of which presented *deductive updating* problems, whereas the other series presented the *strategy assessment* problems.

Deductive updating problems were presented in the following way. On the top of the first page of these problems, the following instructions were given, followed by the *initial* formulation of the high-probability problem set:

For each of the following problems you must consider the statements presented as true and you must indicate whether the proposed conclusion can be drawn logically from the presented information.

The scientists noted that on Kronus:

If it thardannes then the ground will become sticky.

Consider the following statements and respond to the question:

If it thardannes then the ground will become sticky.

Observation: The ground is sticky.

Conclusion: It has thardanned.

Indicate whether this conclusion can be drawn logically or not from the statements.

Participants were given a choice between a NO and a YES response. At the top of the next page, they received the

following information, in which *updated* statistical information for the high-probability problem set was presented:

In one of their monthly communiqués, the scientists sent the following supplementary information. They said that they had made 1,000 observations on Kronus. From this, they found that 980 times it had thardonné and the ground became sticky, while 20 times it had not thardonné and the ground became sticky.

Following this update, participants were given exactly the same inference as had been presented previously.

One further problem set was then presented that followed this same pattern, with the exception that the presented inference was of the (DA) form “if P then Q. P is false.” The updated information showed that of 1,000 observations, 980 had not-P and not-Q, whereas 20 had not-P and Q.

The *strategy assessment* problems presented the set of 13 problems used by Markovits et al. (2012). Each problem described a causal conditional relation involving a nonsense term or relation 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 or not the conclusion could logically be drawn from the premises. The second problem set was identical to the first set, except that the content of the major premise was changed for each problem, while maintaining the same frequency information.

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 individual 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 the 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 was given off, and the TrolYTE was heated.

90 times Philoben gas was 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.

An initial booklet was constructed that presented the *deductive updating* problems first, followed by the *strategy assessment* problems. A second booklet was constructed, which was identical to the first except that the *strategy assessment* problems were presented initially, followed by the *deductive updating* problems. Two further booklets were also constructed, which were identical to the initial two except that the order of the deductive updating problems was inverted, with the DA inference presented first, followed by the AC inference.

Design *Strategy assessment* problems were used as a between-subjects categorization method. The inferences on the updating problems were presented as an *initial* inference followed by an *updated* inference, and this was a within-subjects variable. Order effects were controlled by two between-subjects variables. The first varied the order of the *strategy assessment* problems and the *updating* problems (“updating order”), and the second varied the order of the two forms of inference used in the *updating* problems (“inference order”). Finally, a power analysis showed that with this design, between 90 and 100 participants would give over a 90 % chance of detecting a moderate effect size, and this was the general criterion used to determine the number of participants (within the constraints imposed by using entire classes). This criterion was also employed in the two subsequent studies.

Procedure Booklets were randomly distributed to entire classes. Students who wished to participate were told to take as much time as they needed to answer the questions.

Results and discussion

We first analyzed performance on the *strategy assessment* problem set. Participants who rejected all of the 10 % inferences and all of the 50 % inferences were put into the counterexample category. Participants whose acceptance rates on the 10 % items were greater than their rates on the 50 % items were put into the *statistical* category. All other patterns of responses were put into the *other* category. We then calculated the mean acceptance rates for the initial inference on the AC and the DA inferences, combined, and the mean acceptance rates after the updated information, as a function of reasoning strategies (see Table 1).

We then compared the performance on the *deductive updating* problems for participants showing a counterexample strategy and for those showing a statistical strategy. We performed an analysis of variance (ANOVA) with the number of accepted conclusions as the dependent variable, with

Table 1 Mean numbers of inferences accepted (of a total of two) with standard deviations, as a function of reasoning strategy

Reasoning Strategy	<i>N</i>	Initial Inference	Updated Inference
Counterexample	53	1.51 (0.82)	0.13 (0.44)
Statistical	30	1.50 (0.78)	0.97 (0.89)
Other	24	1.33 (0.70)	0.71 (0.91)

inference (initial, updated) as a repeated measure and strategy (statistical, counterexample), updating order (updating first, updating last), and inference order (AC first, DA first) as between-subjects variables. This analysis showed significant effects of strategy, $F(1, 74) = 10.99, p < .001, \eta_p^2 = .129$, and inference, $F(1, 74) = 66.87, p < .001, \eta_p^2 = .475$, and a significant interaction involving Strategy \times Inference, $F(1, 74) = 15.29, p < .001, \eta_p^2 = .171$. None of the other effects or interactions were significant, with the largest $F(1, 95) = 1.82$.

Post hoc comparisons were done using a Tukey test with $p = .05$. This showed that there was no difference in the mean numbers of conclusions accepted on the initial inferences as a function of strategy. Both statistical and counterexample reasoners showed significant decreases in the extent to which the updated conclusions were accepted. However, reasoners using a counterexample strategy on the initial problems accepted significantly fewer conclusions on the updated inferences than did reasoners using a statistical strategy.

These results are thus consistent with our prediction. Importantly, people's initial rates of conclusion acceptance on the first version of the updating problem sets were similar, irrespective of reasoning strategy. This shows that the differences were uniquely related to ways that the updated statistical information was processed. As hypothesized, reasoners using a counterexample strategy on the diagnostic inferential problems showed a much lower rate of conclusion acceptance following the updated statistical information, which was designed to suggest a high probability of the conclusion being true, relative to reasoners using a statistical strategy. In fact, as can be seen from Table 1, counterexample reasoners almost completely rejected the conclusions following the updated information. By contrast, statistical reasoners, while also showing a significant decrease in conclusion acceptances, had a higher rate of conclusion acceptance than did the counterexample reasoners after updating. This was precisely the pattern predicted.

Study 2

The results of the initial study are consistent with the idea that judgments of logical validity show a high degree of variability. However, the complementary idea that probabilistic reasoning should show relatively little variability has no direct empirical

basis. We examined this hypothesis initially by replicating the first study, but using explicitly probabilistic inferences in the updating procedure.

Method

Participants A total of 86 University students (39 males, 47 females; average age = 24 years, 2 months) took part in this experiment. The students were native French speakers and volunteers.

Material Four paper-and-pencil booklets were prepared. These were identical to the booklets used in Study 1, with one exception. The inferences used on the updating problem sets required participants to estimate the probability that the putative conclusion was true. The initial instructions asked participants to "suppose that each rule that is presented is always true and indicate the probability that the conclusion is true given the presented information." For each problem, after the presentation of the rule and an observation, participants were asked to indicate the probability that the putative conclusion was true on a scale from 0 % to 100 %, in increments of 10 %.

Design *Strategy assessment* problems were used as a between-subjects categorization method. The inferences on the updating problems were presented as an initial inference followed by an updated inference, and this was a within-subjects variable ("inference type"). Order effects were controlled by two between-subjects variables. The first varied the order of the *strategy assessment* problems and the *updating* problems ("updating order"), and the second varied the order of the two forms of inference used in the *updating* problems ("inference order").

Procedure Booklets were randomly distributed to entire classes. Students who wished to participate were told to take as much time as they needed to answer the questions.

Results and discussion

As before, we categorized participants as using a *counterexample*, a *statistical*, or an *other* strategy. We calculated the mean probability ratings for the initial inferences on the AC and the DA inferences, combined, and the mean ratings after the updated information, as a function of reasoning strategies (see Table 2). We then compared performance on the updating problems for participants showing a counterexample strategy and those showing a statistical strategy. We performed an ANOVA with the number of accepted conclusions as the dependent variable, with inference type (initial, updated) as a repeated measure and strategy (statistical, counterexample), updating order (updating first, updating last), and

Table 2 Mean ratings (with standard deviations in parentheses) of conclusion probability, as a function of reasoning strategy

Reasoning Strategy	<i>N</i>	Initial Inference	Updated Inference
Counterexample	47	.789 (.22)	.775 (.29)
Statistical	28	.763 (.26)	.762 (.24)
Other	11	.668 (.29)	.700 (.32)

inference order (AC first, DA first) as between-subjects variables. This showed two almost significant interactions involving inference order, $F(1, 78) = 3.66, p < .06$, as well as a Strategy \times Updating Order interaction, $F(1, 78) = 3.67, p < .06$. Critically, none of the terms involving inference type were significant, $F_s < 1$ in all cases. In other words, consistent with what can be observed in Table 2, participants using a counterexample strategy responded in the same way as those using a statistical strategy when they were asked to update explicitly probabilistic inferences.

In order to better interpret these results, we performed a post hoc power calculation using G*Power, version 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009). Specifically, we calculated the probability of detecting an interaction effect having the same effect size as was observed in Study 1. The calculated power was 99.9 %.

These results indicate that people using counterexample and statistical strategies process statistical information in the same way when updating explicitly probabilistic inferences. Thus, the variability related to reasoning strategy that is observed when updating inferences asking for judgments of logical validity is not present when updating probabilistic inferences.

Two additional facets of these results, although not directly related to our specific context, are useful to note. First, there was some evidence that reasoning strategy and problem order might impact people's overall probabilistic judgments, without affecting the processing of updating information. Second, it is interesting to note that these results along with those of a previous study (Markovits et al., 2013) show that even when given strong updating evidence that conclusions are true, people's updated probability estimates do not increase much over their initial estimates. This seems somewhat paradoxical. However, both of these results can be understood within a broader Bayesian perspective, which we will come back to in the conclusion.

Study 3

The results of these two studies are consistent with the idea that when people make explicitly probabilistic inferences, they generally use a low-level, intuitive procedure that is both more immediate and less prone to variability than inferences

about logical validity made with identical parameters. By contrast, when they are asked to do the latter, people must process problem parameters in a more complex way, leading to a high degree of individual variability that is related to the specific strategy used. This conclusion is consistent with the observed results but remains indirect. A more direct measure of the difference between the two forms of inference would make a stronger case. We hypothesized that if people are given statistical information and are asked to use this in order to make an explicitly probabilistic inference, the intuitive strategy used on this task would carry over to a subsequent task requiring inferences of logical validity much more easily than to the opposite task. We specifically predicted that simply presenting probabilistic inferences initially should make people more sensitive to the statistical properties of the presented updated information, which would result in a significant increase in conclusion acceptances on the updated inferences. By contrast, making an inference about logical validity would require additional cognitive processing and some transformation of the updated information, which would have little impact on an intuitive statistical strategy.

Method

Participants A total of 92 University students (35 males, 57 females; average age = 22 years, 7 months) took part in this experiment. The students were native French speakers and volunteers.

Material Two paper-and-pencil booklets were prepared. The first booklet (*logical deduction first*) presented the same two updating problems used in Study 1, with the first set using the AC inference and the second set using the DA inference. Each inference required a judgment as to the logical validity of the presented conclusion. Following these problems, participants then received two more updating problem sets with a different content, but in the same order (AC followed by DA), and participants were asked to indicate the probability that the conclusion was true, with the same format as was used in Study 2. The second booklet (*probabilistic inference first*) was identical to the first one, with the exception that the two updating problems with probabilistic inferences were presented first, followed by the updating problems requiring judgments of logical validity.

Design Inferences on the updating problems were presented as an initial inference followed by an updated inference, and this was a within-subjects variable ("inference type"). Two forms of inference were used: explicitly probabilistic and judgments of logical validity. Half of the participants received the explicitly probabilistic inferences first, followed by judgments of logical validity, whereas the other half received these in the opposite order ("order").

Procedure Booklets were randomly distributed to entire classes. Students who wished to participate were told to take as much time as they needed to answer the questions.

Results and discussion

For the two problem sets involving logical validity and the two problem sets involving ratings of the probability of conclusions, we calculated the mean numbers of conclusion acceptances and the mean ratings, respectively, for each of the two conditions (see Table 3). We first examined performance on the problems requiring judgments of logical validity. We performed an ANOVA with the number of conclusion acceptances as the dependent variable, with inference type (initial, updated) as a repeated measure and order (logical validity first, probabilistic inference first) as a within-subjects variable. This showed significant effects of inference type, $F(1, 89) = 18.95, p < .001, \eta_p^2 = .225$, and a significant Inference Type \times Order interaction, $F(1, 89) = 4.45, p < .05, \eta_p^2 = .048$. Analysis of the interaction was performed using the Tukey procedure with $p = .05$. This showed that whereas there was no difference in the initial numbers of conclusion acceptances, we observed a significant increase in updated conclusion acceptances when the probabilistic inferences were given first ($M = .96, SD = .88$), as compared to when the judgments of logical validity were given first ($M = .53, SD = .80$).

We then examined performance on the explicitly probabilistic inferences. We performed an ANOVA with mean ratings as the dependent variable, with inference type (initial, updated) as a repeated measure and order (logical validity first, probabilistic inference first) as a within-subjects variable. This showed no significant effects. In order to better interpret these results, we performed a post hoc power calculation using G*Power, version 3.1 (Faul et al., 2009). Specifically, we calculated the probability of detecting an interaction effect having the same effect size that was observed with the deductive inferences. The calculated power was 95.9 %.

These results were again consistent with our predictions. When people are initially asked to make updated inferences using explicitly probabilistic judgments, this has no effect on their initial judgments of logical validity, but produces a clear increase in their tendency to accept the updated inferences. By contrast, probabilistic inferences are relatively indifferent to order. The pattern is exactly the same as that observed when

individual differences in types of reasoning strategy on judgments of logical validity and explicitly probabilistic ratings were examined.

General discussion

A major debate in the psychology of reasoning concerns the underlying nature of inferential processes. The major models that have been proposed, which we refer to as *counterexample-based* or *statistical*, each claim to provide a single, unitary form of inference that can, in principle at least, account for all types of reasoning. The dual-strategy model of reasoning, first proposed by Verschueren et al. (2005a, 2005b), suggested that people might use reasoning strategies that correspond to both of these forms of reasoning. Recent studies have provided increasing evidence for such a model (Markovits et al., 2015; Markovits et al., 2013; Markovits et al., 2012).

One further postulate of this model suggested that the statistical strategy corresponds to an intuitive, low-cost evaluation of conclusion likelihood. Although some indirect evidence had suggested that this is the case, the present studies provide much stronger evidence for this postulate. The results of these three studies show that reasoning requiring a judgment of logical validity is much more variable than reasoning requiring an explicitly probabilistic judgment. In other words, consistent with the dual-strategy model, the processes used for probabilistic reasoning are more invariant than those used when making judgments of logical validity.

Interestingly, the patterns of variation found in these studies are also consistent with a dual-process interpretation of these strategies (Evans, 2007; Sloman, 1996; Stanovich & West, 2000). They show that induction of a lower-level statistical strategy clearly has an interactive effect on judgments of logical validity, making them less consistent with optimal use of a counterexample strategy. By contrast, inducing a counterexample strategy has no impact on judgments of probability. Similarly, the results of Study 1 show that reasoners using a statistical strategy on simple problems show a decrease in judgments of validity, although one that is clearly smaller than the one for reasoners using a counterexample strategy. In other words, this is exactly the pattern of variation that would be predicted if statistical reasoning was a form of heuristic

Table 3 Mean ratings on the probabilistic judgments and mean proportions of conclusion acceptances (with standard deviations in parentheses) on judgments of logical validity (of a total of two), as a function of order

	Logical Validity		Probability Ratings	
	Initial	Updated	Initial	Updated
Order				
Logical validity first	1.40 (0.80)	0.53 (0.80)	.803 (0.23)	.795 (0.23)
Probability ratings first	1.33 (0.85)	0.96 (0.88)	.743 (0.23)	.742 (0.21)

processing, whereas counterexample reasoning was a form of analytic processing.

Irrespective of just how these two strategies are interpreted, these results, along with those of previous studies, provide strong evidence that people have access to two qualitatively different inferential strategies. Statistical strategies that use the statistical properties of inferential problems to generate an estimate of the probability of a conclusion being true are low-level, intuitive, and generally invariant. Counterexample strategies, which construct explicit representations of premises that may or may not include explicit counterexamples, are more resource-intensive and much more variable. The latter strategies are subject both to individual differences and to the effects of context and time constraint (Markovits et al., 2013).

Finally, it is useful to put the dual-strategy model into a broader perspective. Although the sum of the available empirical evidence clearly supports this model, these results do not provide any indication of the nature of the processes underlying each strategy. For example, as was mentioned in a previous article (Markovits et al., 2013), although the initial formulation of this model identified the counterexample strategy with the mental model theory (Verschuere et al., 2005a, 2005b), it is possible to conceive of this strategy within a generally probabilistic approach (e.g., Oaksford & Chater, 2007). Interestingly, current formulations of probability theory are explicitly Bayesian and consider subjective probabilities to be the results of underlying beliefs derived from a variety of sources. For example, one formulation uses the Ramsey test as a basic criterion for conclusion belief (e.g., Evans, Handley, & Over, 2003). This suggests that when people are evaluating their confidence in a conditional statement, they will hypothetically consider a potential world in which the premises are true and apply their knowledge about the relevant conditions. In fact, such a formulation can explain one of the more interesting results of the present studies. As we previously noted, when people are asked to evaluate the probability of the conclusion based on the initial version of the updating problems, they generally evaluate this as being much higher than 50 %. In addition, when given explicit statistical information suggesting that this conclusion is very probable, they do not modify their initial evaluation a great deal. This, in turn, is consistent with the idea that the initial evaluation is done by some form of Bayesian analysis, which does not consider specific conditions (since these rules concern completely fictitious entities), but does evaluate people's general belief in the conclusion, something that is clearly weighted more strongly than the specific updated information.

Author note This study was financed by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to H.M.

References

- Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, *19*, 274–282. doi:10.3758/BF03211151
- Evans, J. S. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning*, *13*, 321–339.
- Evans, J. S. B. T., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 321–335. doi:10.1037/0278-7393.29.2.321
- Evans, J. S. B. T., 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. doi:10.1037/0033-295X.112.4.1040
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. doi:10.3758/BRM.41.4.1149
- Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences*, *5*, 434–442.
- Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*, 646–678. doi:10.1037/0033-295X.109.4.646
- Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J.-P. (1999). Naïve probability: A mental model theory of extensional reasoning. *Psychological Review*, *106*, 62–88. doi:10.1037/0033-295X.106.1.62
- Markovits, H., & Barrouillet, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review*, *22*, 5–36. doi:10.1006/drev.2000.0533
- Markovits, H., Brisson, J., & de Chantal, P.-L. (2015). Deductive updating is not Bayesian. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/xlm0000092
- 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. doi:10.1037/a0030906
- Markovits, H., Lortie-Forgues, H., & Brunet, M.-L. (2010). Conditional reasoning, frequency of counterexamples, and the effect of response modality. *Memory & Cognition*, *38*, 485–492. doi:10.3758/MC.38.4.485
- Markovits, H., Lortie-Forgues, H., & Brunet, M.-L. (2012). More evidence for a dual-process model of conditional reasoning. *Memory & Cognition*, *40*, 736–747. doi:10.3758/s13421-012-0186-4
- Markovits, H., & Vachon, R. (1990). Conditional reasoning, representation, and level of abstraction. *Developmental Psychology*, *26*, 942–951. doi:10.1037/0012-1649.26.6.942
- Oaksford, M., & Chater, N. (2007). *Bayesian rationality*. Oxford, UK: Oxford University Press.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22. doi:10.1037/0033-2909.119.1.3
- 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.
- Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition*, *22*, 742–758. doi:10.3758/BF03209259
- Verschuere, N., Schaeken, W., & d'Ydewalle, G. (2005a). A dual-process specification of causal conditional reasoning. *Thinking & Reasoning*, *11*, 239–278. doi:10.1080/13546780442000178
- Verschuere, 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. doi:10.3758/BF03195301