Is inferential reasoning just probabilistic reasoning in disguise?

HENRY MARKOVITS and SIMON HANDLEY
University of Plymouth, Plymouth, England

Oaksford, Chater, and Larkin (2000) have suggested that people actually use everyday probabilistic reasoning when making deductive inferences. In two studies, we explicitly compared probabilistic and deductive reasoning with identical *if–then* conditional premises with concrete content. In the first, adults were given causal premises with one strongly associated antecedent and were asked to make standard deductive inferences or to judge the probabilities of conclusions. In the second, reasoners were given scenarios presenting a causal relation with zero to three potential alternative antecedents. The participants responded to each set of problems under both deductive and probabilistic instructions. The results show that deductive and probabilistic inferences are not isomorphic. Probabilistic inferences can model deductive responses only using a limited, very high threshold model, which is equivalent to a simple retrieval model. These results provide a clearer understanding of the relations between probabilistic and deductive inferences and the limitations of trying to consider these two forms of inference as having a single underlying process.

Conditional (if–then) reasoning is considered by many as the cornerstone of deductive reasoning and has been the subject of an important amount of research. Although deductive reasoning is ideally considered to be abstract in nature, there are many striking forms of variation, due to premise content, in the way that even educated adults make deductive inferences. These suggest that the processes that underlie deduction are sensitive to knowledge about the premises in some way. This form of variation has led Oaksford, Chater, and Larkin (2000) to propose that when people make conditional inferences, they do not attempt to make "logical" inferences but, rather, use everyday probabilistic reasoning, which directly accesses knowledge about premises, to respond to these problems. However, although such a model can account for variation in performance in many instances, there are other ways of doing this that retain the distinction between probabilistic and inferential processes. In the following, we will present two studies that compare deductive and probabilistic reasoning on the same problems.

Conditional reasoning consists in making an inference on the basis of a major premise of the type "if P, then Q" (the first clause, P, is called the *antecedent*; the second, Q, is called the *consequent*). There are four basic inferences that can be made from a major conditional premise. Two of these inferences, the affirmation of the antecedent ("P is true"), called *modus ponens* (MP), and the denial of the consequent ("Q is false"), called *modus tollens* (MT), lead to logically certain conclusions. For example, assuming the truth of the proposition "if it rains, then the street is wet" authorizes the conclusions that "if it rains," then it is certain that "the street is wet" and "if the street is not wet," then "it is not raining." The two other logical forms, consisting of the *denial of the antecedent* (DA; "P is false") and the *affirmation of the consequent* (AC; "Q is true") are uncertain arguments, because they do not allow logically certain conclusions. For example, the DA "it is not raining" does not permit the certain conclusion that "the street is not wet," because it is possible that something other than rain can cause the street to be wet. Likewise, knowing that "the street is wet" (AC) does not permit the certain conclusion that "it has rained," for the same reason.

Research has consistently shown that adult reasoners will often draw what are logically inappropriate conclusions when asked to make deductive inferences. This is particularly the case with the two logical forms for which there is no certain conclusion, AC and DA, where a common response is to accept the invited inference in each case. Several studies have shown that one key factor that appears to influence what kinds of inferences are made on the AC and DA forms concerns the relative number and availability of alternative antecedents—that is, cases of [not-P and Q] (Cummins, 1995; Cummins, Lubart, Alksnis, & Rist, 1991; Janveau-Brennan & Markovits, 1999; Markovits & Vachon, 1990; Quinn & Markovits, 1998; Thompson, 1994). Thus, when the specific content of a conditional premise allows for many potential alternative antecedents (or for one readily accessible one), there will be a tendency to refuse the invited conclusion, and conversely, when there are relatively few available antecedents, there will be a tendency to accept this conclusion.

It is this kind of variability, among others, that has led Oaksford and Chater (Chater & Oaksford, 1999; Oaksford & Chater, 1994; Oaksford et al., 2000) to propose that when people are given "logical" reasoning problems, they will not attempt to make logical inferences; instead, they will use the kind of probabilistic reasoning that is used in many everyday contexts. This hypothesis is particularly appealing in the light of the previously cited studies on conditional reasoning, because it apparently provides a clear explanation for the strong relation between content and inference. For example, suppose that a reasoner is given the following logical problem: "If a rock is thrown through a window, then the window will break. A window is broken." The probabilistic model would claim that the reasoner would first evaluate the probability that, given that the window is broken, a rock was thrown through the window. In this case, the information that there are many ways to break a window other than with a rock would make the conclusion that the window was broken by a rock quite unlikely. This would then lead the reasoner to reject the invited conclusion. On the other hand, when given a problem such as "If a tree is cut down, then the tree will fall," the same process would lead a reasoner to conclude that it was quite likely that the tree was cut down.

Although a probabilistic model is an intuitively appealing one, there has been little attempt to specifically compare probabilistic and deductive reasoning, although this question is clearly becoming more salient. One recent study that has compared these two forms of reasoning was done by Rips (2001). This study concluded that, despite a large degree of isomorphism between deductive and inductive inferences, at some key points they diverge. The most critical of these points concerned reasoning of the kind "Car X runs into a brick wall. Therefore car X will stop," for which the rate of acceptance of the inference under deductive instructions was very much lower than the inductive evaluation of the same inference. However, it is possible, in this case, to distinguish two factors. One would be the metacognitive recognition of the form of a deductive argument, whereas the other would be the process by which a deduction is made when a problem is recognized as *deductive*. In fact, one important developmental milestone is the ability to recognize when a deduction can be made and when one cannot (Markovits, Schleifer, & Fortier, 1989). Thus, the key results obtained by Rips (2001) could be attributable to the recognition that it is not appropriate to make a deductive inference with this kind of material, while leaving open the possibility that the underlying processes used in making a deduction, when this is judged appropriate, might still be probabilistic in nature.

The chief aim of this research was, then, to compare patterns of deductive and probabilistic reasoning on the same materials, using the standard forms of conditional reasoning. We also wished to examine the extent to which a purely probabilistic model could account for deductive performance. As we have indicated previously, a probabilistic model appears to be able to account for many of the observed global effects of content on conditional reason-

ing (although see Quinn & Markovits, 2002). However, there has been little attempt to determine the extent to which information about probabilistic inferences can directly predict specific patterns of deductive performance. One problem in doing so is that existing probabilistic models do not specify just how an underlying judgment of the probability that a conclusion is correct is translated into the categorical judgments typically used in deductive reasoning tasks. However, we can distinguish two models for doing this. The first is what Liu (2003) refers to as the *proportionality* hypothesis, which claims that the rate of endorsement of a conditional argument is directly proportional to the estimated probability of the conclusion, given the premises. This is, of course, the most intuitive model. It would claim that if a reasoner estimated that, for a given "If P then Q" premise, the probability of q given p (modus ponens) is 80%, the reasoner should endorse the modus ponens argument about 80% of the time (note, however, that such a model does not specific the mechanism by which such proportionality is achieved). The second model is a *threshold* model, which would use some internal threshold—say, T—which would be used to determine whether a conclusion will be judged to be valid or not. With this model, a conclusion will be judged valid if its estimated probability is greater than or equal to T and invalid otherwise.

There are, however, other models of deductive reasoning that have been used to predict a wide variety of content-related effects in conditional reasoning that do not suppose an underlying probabilistic evaluation. Specifically, Markovits and colleagues (Grosset, Barrouillet, & Markovits, 2005; Markovits & Barrouillet, 2002; Markovits, Fleury, Quinn, & Venet, 1998; Markovits & Quinn, 2002; Quinn & Markovits, 1998, 2002) have developed a retrieval-based model of deductive inference derived from mental model theory (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991). The model claims that when making a logical inference with concrete premises, reasoners will attempt to search for information about the premises that are stored in long-term memory. In the specific case of the AC and DA inferences, this will involve a search for potential alternative antecedents. If at least one such alternative is retrieved, reasoners will tend to reject the invited conclusion; if not, they will tend to accept this conclusion. Because the probability of retrieving an alternative antecedent is clearly related to the number of potential antecedents available in long-term memory, which is, in turn, related to the probability of the conclusion given in both the AC and the DA inferences, the two models often lead to similar predictions.

However, the specific relation between the retrieval probability and the estimated probability of a given conclusion under a probabilistic hypothesis is variable and will depend on how efficient the reasoner's retrieval processes are (Markovits & Quinn, 2002). One clear case in which some divergence might be expected is given by an AC inference with a major premise of the form "If cause P then effect Q," where P is the single most strongly associated cause for the effect Q, and potential alternative

causes, A1, A2, . . . are all considered to be relatively unlikely. In this particular case, the rated probability of "P being true" should be quite high, which should also lead to a correspondingly strong tendency to accept the logical conclusion that "P is true," according to the probabilistic model. However, a retrieval model would claim that the level of acceptance of the conclusion "P is true" will depend only on the probability of retrieval of at least one (even unlikely) alternative. Assuming that most educated adults have reasonably efficient retrieval processes, this would imply that the level of acceptance of the certainty of the conclusion that "P is true" should be much lower than the perceived probability that the same conclusion is true. It should be noted, in this context, that although the retrieval model described here is not the only one that distinguishes deductive and probabilistic reasoning, it is the most completely specified at the moment and allows the clearest comparison with models based on probability estimates.

STUDY 1

The first study presented adult reasoners with the same set of inferences and asked them to make either a "logical" deductive inference or an explicitly probabilistic one. All the reasoners were given the standard logical instructions to suppose that the major premise was true. These instructions limit reasoners' tendencies to retrieve potential disabling conditions (Cummins, 1995) and should generate a situation in which both relative probability ratings and acceptance of the certainty of the conclusion of the MP form would be equally high. For the conditional inferences, causal premises were used, for each of which there was only one very strongly associated cause, which was used in the major premise (these were taken from Quinn & Markovits, 1998). These were explicitly chosen because they generally generate relatively low levels of acceptance of the AC and DA inferences under deductive instructions with educated adults but should be expected to generate relatively high ratings of the probability of these inferences.

Method

Participants. A total of 120 university students at the Université du Québec à Montréal (average age, 24.20 years; 36 men and 84 women) took part in this study.

Material. Four booklets of five pages were constructed. On the first page of the probabilistic booklets, the participants read the following instructions (translated from the original French):

In the following pages, you will have to respond to some problems. For each problem, you will be first presented with a rule that you must consider to be true, even if this might not be the case in real life. It is thus very important, in order to respond to the questions that will follow, to suppose that the rule presented at the top of each page IS ALWAYS TRUE. After this rule, you will be given some statements. Your task consists of indicating the probability that a given conclusion is true, given the rule and the statement that is presented. Please read the whole text carefully before putting down your response.

On the top of the second page was presented the conditional premise "If a dog has fleas, then it will scratch constantly," preceded by the instruction "Suppose that it is true that." This premise was fol-

lowed by two multiple-choice questions. The first question corresponded to the MP logical form and was presented as follows:

Michael's dog has fleas.
What is the probability that Michael's dog scratches constantly?

Directly below this question was a 7-point scale that went from 1 (*not at all probable*) to 4 (*moderately probable*) to 7 (*absolutely certain*).

Following this, a second question, corresponding to the DA form, was presented in the same format.

On the third page, the major premise "If a person cuts their finger, then their finger will bleed" was presented, followed by questions corresponding to the AC and MT forms. On the fourth page, the major premise "If one is in a room where there is no light, then the pupils of one's eyes will be dilated" was presented, followed by questions corresponding to the AC and DA forms. On the fifth page, the major premise "If a film is exposed to light before it is developed, then the film will become dark" was presented, followed by questions corresponding to the MP and MT forms. Thus, each participant received two questions for each of the four logical forms.

A second booklet in which the probabilistic format was used was constructed that was identical to the first, except for the order of the questions asked. In this case, each of the major premises was accompanied by the two questions that were not used in the first booklet. For example, for the major premise "If a dog has fleas, then it will scratch constantly," the questions corresponded to the AC and MT forms.

Two booklets presented the same problems in a "logical" format. These were identical to the probabilistic ones, with the following exceptions. The initial instructions were as follows:

In the following pages, you will have to respond to some logical problems. For each problem, you will be first presented with a rule that you must consider to be true, even if this might not be the case in real life. It is thus very important, in order to respond to the questions that will follow, to suppose that the rule presented at the top of each page IS ALWAYS TRUE. After this rule, you will be given some statements. Your task consists of indicating whether a given conclusion is certainly true, given the rule and the statement that is presented. Please read the whole text carefully before putting down your response.

The second difference concerned the specific presentation of the problems. For example, for the conditional premise "If a dog has fleas, then it will scratch constantly," a question corresponding to the MP logical form was presented as follows:

Michael's dog has fleas.
Is it certain that Michael's dog will scratch constantly?
a) Yes
b) No

Procedure. The different booklets were randomly given out to the students during regular class time. No time limits were imposed.

Results and Discussion

Results were compiled separately for each of the problem types. For the logical questions, the number of times the invited conclusion was judged to be certain was calculated for each of the four logical forms. For the probability questions, ratings were added together for each instance of a given logical form. In order to make these measures equivalent, both of them were transformed into a percentage of the maximum value. Table 1 indicates these means for the four logical forms (MP, AC, MT, and DA) for the probabilistic and logical conditions (see Table 1).

Inspection of Table 1 shows a very clear pattern. As was expected, the ratings of the probability of the invited conclusion were very high for the MP form (96%), as were

	Problem Type			
	Deductive		Probabilistic	
	($n = 60$)		($n = 60$)	
Logical Form	M	SE	M	SE
MP	83.3	4.39	96.1	1.07
MT	50.0	5.31	71.7	3.35
AC	30.0	5.34	74.5	2.02
DA	29.2	5.35	70.5	3.04

Note—MP, modus ponens; MT, modus tollens; AC, affirmation of the consequent; DA, denial of the antecedent.

the mean percentages of responses in which this conclusion was accepted as certain on the logical task (83%). However, although the probability ratings for MT, AC, and DA were similar and, also, quite high, the percentages of acceptance of the corresponding deductive conclusions were variable, with the level of acceptance for the MT conclusion (50%) being higher than that for the AC and DA conclusions (about 30%), with all three being lower than comparable ratings of the probability that these conclusions were true (about 70%). We examined the relationships between the four forms, using a *sign* test for the deductive conclusions and paired-subjects t tests for the probabilistic conclusions, with a Bonferroni correction. In the probabilistic condition, this showed that probability ratings for the MP conclusion were significantly higher than those for the AC [$t(59) = 9.37, p < .001$], the DA [$t(59) = 7.96, p < .001$], and the MT [$t(59) = 7.57, p < .001$] conclusions, with no difference among the latter three. In the deductive condition, this showed that the acceptance rate for the MP conclusion was significantly higher than that for the MT conclusion ($z = 4.29, p < .001$). In addition, the rate of acceptance for the MT conclusion was significantly higher than that for the AC conclusion ($z = 2.63, p < .008$) and was higher than that for the DA conclusion ($z = 2.65, p < .008$).

These results suggest that the pattern of performance on the four conditional inferences differed between the deductive and the probabilistic tasks. This is consistent with what was found by Rips (2001), although in this case, the effect is not attributable to any metacognitive factors, because all four inferences were standard deductive forms, in both conditions.

Another, somewhat separate question is to what extent deductive performance can be predicted from probabilistic inferences. Both a simple inspection of the results and the analysis above clearly show that the proportionality hypothesis is incorrect. The fact that the relationships among the four logical forms differed in the deductive and the probabilistic conditions eliminates any form of linear translation of estimated probability levels as a possible

explanation of rates of acceptance of conclusions in the deductive condition.

We then examined how well a threshold model could explain how probability ratings might translate into deductive judgments. In order to do this, we started by establishing a threshold value of 7, which was the maximal value on the scale used. We then applied the following algorithm to each of the eight probability judgments made in the probabilistic condition; for each conclusion, we converted the rated probability value to a judgment of *certain* if it was equal to or greater than 7, and we converted the rated value to a judgment of *not certain* otherwise (note that in this particular case, 7 is the endpoint, and the *greater than* condition is redundant). Thus, each of the eight probabilistic conclusions was translated into a *certain* or a *not certain* judgment. We then calculated the mean number of certain conclusions for the two MP questions, the mean number for the two MT questions, and so forth. We then repeated this calculation for thresholds of 6 and of 5. The results are summarized in Table 2.

As can be seen from the table, estimating deductive performance from probabilistic ratings with a threshold of 7 gives a remarkably good fit to mean ratings on the deductive task, whereas lower cutoffs give increasingly divergent results. In order to quantify this, we calculated an overall goodness-of-fit measure consisting of the sum of the squares of the difference between the observed values and the estimated values for the four logical forms (thus, a lower value indicates a better fit to the observed data). This gave a value of 27.66 for a threshold of 7, a value of 98.92 for a threshold of 6, and a value of 1,813.45 for a threshold of 5. This analysis confirms what is clearly visible in Table 2—that is, that a cutoff that corresponds to the highest certainty rating gives an excellent fit to the deductive data, whereas lowering this cutoff systematically renders the fit less good. Although we have not included these data in the table, the negative relationship between goodness of fit and lowering the cutoff point is true for values below 5.

These results present a very clear pattern. The premises used in this study were chosen so that empirical knowledge would make the invited conclusions to both the AC and the DA forms relatively probable. In addition, the instructions given were chosen to make the MP conclusion also very probable. When adult reasoners were explicitly

Task	Cutoff	MP	MT	AC	DA
Deductive		83.3	50.0	30.0	29.2
Probability	7	84.2	45.0	29.2	33.3
	6	89.2	52.9	35.0	52.9
	5	96.7	62.5	59.2	54.2

Note—MP, modus ponens; MT, modus tollens; AC, affirmation of the consequent; DA, denial of the antecedent.

asked to make probability ratings of these conclusions, they rated the MP inference as more probable than the MT, AC, and DA inferences, with no difference among the latter three forms, each of which was rated as being about 70% probable. When reasoners were asked to make deductive inferences with the same material, the acceptance rate for the MP inference was also higher than that for the three other forms. However, in this case, the MT inference led to an acceptance rate that was higher than the ones for the AC and DA inferences, which were quite low. This shows that the internal patterns vary between probabilistic and deductive inferences, despite the fact that all the inferences used the same materials and that these were uniformly couched in a standard format. In addition, the relatively low acceptance rates of the AC and DA inferences are consistent with the initial predictions that we made on the basis of a simple retrieval model.

These data are clearly inconsistent with any form of *proportional* model that considers that deductive judgments reflect some direct linear translation of an underlying probabilistic evaluation of conclusions. However, our analyses show that a *threshold* model applied to probabilistic evaluations can provide an excellent fit for observed deductive performance, but only with a threshold at the maximum level of certainty used in this study. In other words, probability evaluations can be used to model deductive judgments if the following rule is used: A deductive inference is accepted if the subjective probability of this inference is close to certainty, and it is rejected otherwise.

STUDY 2

Although the results of the previous study appeared to be quite strong, we decided to look at the relation between deductive and probabilistic inferences in a more direct way. A second study was designed to examine patterns of variation in deductive and probabilistic reasoning, when the real probabilities of the AC and DA forms were systematically varied. In order to directly examine the patterns of variation, we presented participants with an artificially constructed situation (getting free drinks in a bar) for which we presented one, two, three, or four potential antecedents (ways of getting free drinks), with fixed probabilities for each. In addition, we used a within-subjects design, with each participant responding to both deductive and probabilistic problems, which allowed examination of the correlations between the two types of reasoning.

Method

Participants. A total of 152 university students at the University Plymouth (average age, 22.2 years; 24 men and 128 women) took part in this study.

Material. Eight booklets of four pages were constructed. On the half of the booklets for which the probabilistic inferences were asked first, the participants initially read the following instructions:

Read the paragraph on the top of each page. In this paragraph, some facts are presented that must be considered as completely true. Following this, you will be given some arguments based on these facts. For each

argument, there is a suggested conclusion. For each of these, you must indicate the **probability that the conclusion is true**, given the presented facts (on a scale from 0% to 100%).

On the top of the next page, the following description of the situation used to make inferences (in the *one-rule* case) was presented:

You must suppose that everything in the following paragraph is true and then respond to the four questions below.

Imagine that a local club has announced that it is making a special offer. Normally, the club always charges for admission, but from now on, everybody that comes into the club will be given a token. 1 out of every 10 of these tokens has a star on one side. **If a customer has a star on their token**, **they will be charged half price for admission.**

After this, there were four probabilistic inferences corresponding to the MP, AC, DA, and MT inferences. In each case, the questions were phrased as follows:

Given that:	**Tom gets a token with a star at the club.**
What is the probability that:	Tom will be charged half price for admission.

The participants were given a scale going from 0% to 100% in increments of 10% to record their responses.

Subsequently, the participants initially read the following deductive instructions on a separate page:

Read the paragraph on the top of the next page. In this paragraph, some facts are presented that must be considered as completely true. Following this, you will be given some arguments based on these facts. For each argument, there is a suggested conclusion. For each of these, you must indicate whether or not the conclusion is necessarily true, given the presented facts. A conclusion is **necessarily true** when it is **absolutely certain** that the conclusion is true.

On the top of the next page, the same description of the situation used to make the initial set of probabilistic inferences (in the one-rule case) was presented again. After this, there were four probabilistic inferences corresponding to the MP, AC, DA, and MT inferences. In each case, the questions were phrased as follows:

Given that:	**Tom gets a token with a star at the club.**
Is it necessarily true that:	Tom will be charged half price for admission.

The participants could circle either *yes* or *no* in response to these questions.

Another version of this basic booklet was constructed by presenting the deductive inferences first, followed by the probabilistic inferences. Finally, the basic *one-rule* situation was systematically varied by adding one, two or all of the following three alternative rules:

(1) In addition to this, each customer will be given a scratch card, and a number. 1 out of every 20 of the scratch cards allows customers to be charged at half price for admission.
(2) Also, 1 out of 20 of the numbers allows customers to be charged at half price for admission.
(3) Not only that, but 1 out of every 20 customers who do not get charged half price with a token, a card, or a number, will be chosen at random and will be charged at half price for admission.

Specifically, the *two-rules* situation presented the basic situation plus Rule 1. The *three-rules* situation presented rules 1 and 2, and the *four-rules* situation presented all three rules, in addition to the basic situation. For each of these, there were two booklets created by changing the order of the deductive and probabilistic inferences. There were thus a total of eight booklets. Each participant received one question for each of the four logical forms in the deductive condition and one question for the same forms in the probabilistic condition.

Procedure. The different booklets were randomly given out to students during regular class time. No time limits were imposed.

Results and Discussion

Responses to the deductive questions were coded as to the proportion of participants who accepted the invited inference for each of the four forms. An initial analysis looked at potential order effects. These were significant for the MP form, which will be examined separately. For the remaining forms, results were averaged over the two orders. Table 3 indicates the average ratings of the certainty of the conclusions to the probabilistic questions and the mean percentages of acceptances of the invited conclusions to the deductive questions, for the AC, DA, and MT forms. An inspection of this table clearly shows that, overall, the addition of potential antecedents to the original rule had an impact on both deductive and probabilistic reasoning.

There are several different ways to examine these data. The first involves looking at the effect of adding rules to the original premise. The one-rule condition allows for no alternative ways of getting half-price admission. According to a simple retrieval model, this should lead to very high levels of acceptance of the AC and DA deductive inferences, which was indeed the case. At the same time, the lack of any alternative antecedents should lead to an evaluation of the probability of these two inferences that is close to 100%, which also was the case. A retrieval model would claim that the main effect, in the deductive condition, of adding potential alternatives to both the AC and the DA forms should be found between the one-rule and the two-rule conditions, with the addition of subsequent rules having little effect on deductive responding. We used chi-squared analyses to look at differences in AC and DA responding in the deductive condition as a function of number of rules. This showed, as was predicted, that there was a significant drop in acceptance of the AC inference between the one-rule and the two-rule conditions [$\chi^2(1) = 32.03, p < .001$], with no significant differences between the two-rule, the three-rule, and the four-rule conditions. This also showed a significant drop in acceptance of the DA inference between the one-rule and the two-rule condition [$\chi^2(1) = 27.91, p < .001$], with no significant differences between the two-rule, the three-rule, and the four-rule conditions. We analyzed differences in probability

ratings by using t tests. This showed significant drops in the rating of the AC inference between the one-rule and the two-rule conditions [$t(76) = 7.81, p < .0001$] and between the two-rule and the three-rule conditions [$t(76) = 4.36, p < .0001$]. Ratings of the DA inference showed a significant drop between the one-rule and the two-rule conditions only [$t(76) = 5.46, p < .0001$]. The pattern of change in deductive inferences with the AC and DA forms precisely mirror those obtained by Quinn and Markovits (2002), using a different paradigm. These results once again show differing patterns of variation between deductive and probabilistic reasoning.

We then specifically examined the predictive value of probability estimates for observed deductive performance. The first point here is that, as has been previously stated, an estimation of the probability of a conclusion of 100% should be strongly associated with a perceived lack of potential alternatives to the conclusion, which would, in turn, lead to acceptance of this conclusion in a deductive condition, no matter what theory is used. As a consequence, the key test for the predictive value of probability estimates for deductive judgments must lie in the potential impact of subjective probability estimates at values that differ from 100%. In other words, once it is known that the specific probability attributed to a conclusion is less than 100%, does knowing the exact level of this value provide additional information about deductive performance? We first looked at the AC form. For the 115 participants who rated the AC conclusion at less than 100%, the point biserial correlation between the probability of the AC conclusion and whether this conclusion was rated as certain on the deductive task was .158 (n.s.). For the 107 participants who rated the DA conclusion at less than 100%, the point biserial correlation between the probability of the DA conclusion and whether this conclusion was rated as certain on the deductive task was −.176 (n.s.). For the 61 participants who rated the MT conclusion at less than 100%, the point biserial correlation between the probability of the MT conclusion and whether this conclusion was rated as certain on the deductive task was .219 (n.s.). In all these cases, the relation between the exact level of evaluation of the probability of the conclusion and the rate of

Table 3
Mean Percentages of Responses in Which the Invited Conclusions Were Accepted as Certain on the Deductive Tasks and Mean Certainty Ratings (as Percentages of Maximum Certainty Rating) for the Invited Conclusions on the Probabilistic Tasks for Three Logical Forms as a Function of the Number of Potential Antecedents (With Standard Errors)

	Problem Type											
	Deductive						Probabilistic					
	AC		DA		MT		AC		DA		MT	
Number of Rules	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
1	81.6	0.06	86.8	0.06	100	0.00	92.9	2.29	94.7	2.74	93.4	3.21
2	17.5	0.06	27.5	0.07	80.0	0.06	55.5	4.13	63.3	4.98	83.5	4.03
3	10.3	0.05	20.5	0.07	69.2	0.07	32.8	3.14	57.9	4.18	77.4	4.64
4	11.4	0.05	8.6	0.05	77.1	0.07	33.6	4.00	50.6	4.88	75.4	5.54

Note—AC, affirmation of the consequent; DA, denial of the antecedent; MT, modus tollens.

acceptance of the same conclusion in a deductive reasoning task was quite low, accounting for less than 5% of the variance in all three cases.

One possible way of explaining these results would be to argue that there might, in fact, be relatively little variability in probabilistic responses, despite the intent of the manipulation. We examined this possibility for the AC inference. Table 4 gives the mean percentages of acceptance of the AC conclusion in the deductive task as a function of evaluation of the probability of this conclusion, when the latter was less than 100%. As can be seen from the table, there is quite a large variation in probabilities, indicating that the manipulation of the number of alternatives did have the desired effect.

These results clearly show that, for conclusions with subjective probabilities of less than 100%, knowing the exact probability attributed to the conclusion has very little predictive value in determining people's deductive performance. In other words, we would expect that the best predictor of deductive reasoning would be whether the perceived probability of the conclusion was 100% or less than 100%. In order to directly examine this, we performed a logistic regression on deductive performance for the AC form, for all the participants. We entered the probability of the AC conclusion and a separate dichotomous variable (which we will refer to as *certainty/uncertainty*), which took a value of 1 if the probability of AC was 100% and 0 otherwise. This gave a significant effect of the certainty/uncertainty variable [$\chi^2(1) = 7.58$, $p < .001$]. There was no significant effect of the probability of the AC conclusion. We then performed a similar analysis using deductive performance for the DA form. This also showed a significant effect only of the certainty/uncertainty variable [$\chi^2(1) = 35.20$, $p < .001$].

Another, more synthetic way of looking at this is to examine the partial correlations between the probability of the conclusion with the certainty/uncertainty variable partialled out, and vice versa for the AC and DA forms. For the AC form, the partial correlation for the certainty/uncertainty variable was .38, whereas the partial correlation for the probability of the AC conclusion was .13. For the DA form, the partial correlation for the certainty/uncer-

tainty form was .64, whereas the partial correlation for the probability of the AC conclusion was −.15. Thus, somewhat paradoxically, both forms of analysis show that the perceived probability of the putative conclusion is a worse predictor of inferential performance than the simple dichotomous variable defined by having a probability equal to or less than 100%.

These analyses strongly suggest that probability information is not directly informative of deductive reasoning. This reinforces the results of the first study. We then simulated the way that probabilistic inferences might be used to make deductive inferences with a threshold model, in the same way as we did in the first study. We transformed each of the four probability estimates into a deductive judgment of *certain* when the estimate was greater than or equal to a given threshold and into a judgment of *uncertain* otherwise. In order to look at how a threshold model could accurately predict deductive performance, we calculated a simple index of fit by summing the square of the difference between observed deductive performance and modeled performance for each of the four logical forms, for each participant. Table 5 gives the modeled values for the four logical forms and the mean index of fit, using threshold values ranging from 70% to 100%. As can be seen, the lowest index of fit is obtained with a cutoff of 100%, with a steady increase in the index (meaning a worse overall fit) with a decrease in cutoff value. Once again, the best overall fit for deductive performance confirms the rule that a deductive inference is accepted if its probability is close to certain. We also looked at a mixed model, where cutoffs of either 100% or 80% were randomly distributed among the participants (so that half received each of the two values). This gave a goodness-of-fit index of .895, which is intermediate between the values for 100% and 80%, indicating that mixed models would not generally give a better fit than the single cutoff models we have used.

Finally, we examined performance on the MP form, for which there was an effect of order of presentation. Table 6 shows the mean probability ratings and the acceptance rates as a function of number of rules and order. In order to examine the locus of the order effect, we performed separate ANOVAs for the probabilistic ratings and deductive judgments, with number of antecedents and order as independent variables. In the latter case, this indicated a main effect of order [$F(1,144) = 22.88$, $MS_e = 0.137$, $p < .001$] and an order × number of antecedents interaction [$F(1,144) = 3.54$, $MS_e = 0.137$, $p < .02$]. Ratings decreased with number of antecedents and were higher when the deductive problems were done first. A similar analysis using acceptance of the MP conclusion indicated that in this case, there was only an effect of number of antecedents [$F(1,144) = 3.47$, $MS_e = 0.079$, $p < .05$]. Acceptance of the MP conclusion decreased with increasing numbers of antecedents. It must, however, be noted that the MP measure for the deductive condition is dichotomous, and this, coupled with the generally high acceptance rates in both cases, which implies the possibility of ceiling effects, means that these results should be interpreted prudently.

Table 4
Mean Percentages of Acceptances of the Affirmation of the Consequent (AC) Inference as a Function of Evaluation of Its Probability

Probability of the AC Inference (%)	N	Percentage of Acceptance
90	4	50.0
80	7	14.3
70	3	33.3
60	2	50.0
50	35	11.4
40	13	0.00
30	17	11.8
20	12	25.0
10	21	4.8

Table 5
Mean Percentages of Deductive Judgments on the Four
Logical Forms in the Deductive Task and Those That Would
be Obtained Using Probability Ratings With a Cutoff Point
Varying Between 100% and 70% in Study 2

Task	Cutoff	Logical Form				Mean Index of Fit
		MP	MT	AC	DA	
Deductive		90.8	81.6	30.3	36.2	
Probability	100	78.9	59.9	24.3	29.6	.816
	90	80.2	68.4	26.9	38.2	.842
	80	81.6	74.3	31.6	51.3	.947
	70	82.9	78.3	33.6	57.2	.961

Note—MP, modus ponens; MT, modus tollens; AC, affirmation of the consequent; DA, denial of the antecedent.

Overall, the results of this second study are also quite clear. In this study, we looked at the relation between deductive and probabilistic reasoning when each reasoner was asked to make both forms of inference on the same problems. In addition, problems were constructed so that the number of potential alternatives (rules) to the initial antecedent could be varied between one and four. The results show that the effect of the number of potential alternative antecedents on inferential performance with the AC form differs between the deductive and the probabilistic conditions. In addition, these results allow the conclusion that the best predictor of whether a conclusion is accepted under deductive instructions is whether the probabilistic evaluation of this conclusion is less than 100% or not. Information about the exact level of the latter evaluation has no predictive value beyond this. Finally, the threshold modeling exercise leads to the same basic result as that obtained in the first study.

GENERAL DISCUSSION

These two studies allow some clear conclusions about two related questions. First, we have examined the way that probabilistic and deductive reasoning vary, using identical forms of content expressed in standard formats. Both studies show that these two forms of reasoning do not have the same patterns of variation. In Study 1, we looked at reasoning with premises for which there were few potential alternatives to the antecedent (e.g., "If a dog has fleas, then it will scratch constantly"). In this particular case, probability ratings of the AC, DA, and MT conclusions were uniformly high and equal to each other. Proportion of acceptance of the corresponding deductive conclusions was very low for the AC and DA inference, both of which were significantly lower than the rate of acceptance of the MT conclusion. Study 2 looked at variation in AC and DA performance, using a constructed social effect (i.e., getting half-price admission) for which there were one, two, three, or four possible ways for obtaining the effect. Results showed that for both AC and DA deductive inferences, there was a significant and large drop in acceptance of the invited conclusions between the one-rule condition and the two-rule condition, with no difference subse-

quently. The same pattern was found for the DA inference in the probabilistic condition, but the AC inference showed a sustained drop in probability ratings between the two-rule and the three-rule conditions (as would indeed be expected). These results, combined with those of Rips (2001), strongly suggest that probabilistic and deductive reasoning are not isomorphic.

The second question then concerns the potential that the use of probabilistic inferences has as a way of predicting deductive performance. We have suggested two possible models for doing this, the proportional model and a threshold model. The data from both studies clearly allow rejection of any form of proportional model as a possible mechanism for translating probabilistic into deductive evaluations. In contrast, both studies show that a threshold model applied to probabilistic evaluations can accurately predict deductive judgments. However, both the results of the two modeling exercises and more direct comparisons between probabilistic evaluations and deductive performance indicate that the best predictor of deductive performance is simply the presence or absence of a very high degree of certainty in the putative conclusion. In other words, the best way of modeling deductive performance with probabilistic evaluations uses the following rule: If a conclusion is probabilistically estimated as being very close to certain, it will be accepted in a deductive judgment; if it is not estimated as being close to certain, it will be rejected in a deductive judgment.

Certainly, the threshold rule that we have described relies on an evaluation of probabilities. However, this rule is completely isomorphic to the simple retrieval model that has already been described (e.g., Markovits & Barrouillet, 2002). In other words, the rule above is operationally equivalent to a rule that accepts a deductive conclusion only if there is no retrieved counterexample, which, in turn, is the basic "logical" principle that underlies mental model theory. More important for the basic intuition underlying probabilistic models, these results also clearly show that once it is known that a reasoner's estimate of the probability of a putative conclusion is less than certain, knowing the exact value of this estimate has no predictive value for deductive performance.

In sum, although these results cannot entirely eliminate the idea that some sort of probabilistic evaluation may be

Table 6
Mean Ratings and Percentages of Acceptance
of the Modus Ponens Conclusion as a Function
of Number of Antecedents and Task Order

Order	Number of Antecedents	Rating	Acceptance (%)
Probability first	1	91.1	100.0
	2	77.0	100.0
	3	73.0	85.0
	4	55.0	72.2
Deductive first	1	96.3	100.0
	2	98.5	90.0
	3	98.9	89.5
	4	94.7	88.2

used in making deductive judgments, they reduce the potential ways of doing so to a single, very high threshold model, one that is completely isomorphic to the simple retrieval model that has already been described (e.g., Markovits & Barrouillet, 2002). In addition, these results clearly show that most people do make a clear distinction between everyday probabilistic reasoning and deductive reasoning.

Finally, it should be pointed out that we used very strong deductive instructions in these studies. This was done in order to make the distinction between deductive and probabilistic tasks as clear as possible. It is very probably the case that relaxing these instructions would generate a corresponding tendency to use a more relaxed criterion for making deductive judgments and would allow increasing use of the kind of everyday, intuitive evaluation that is used in ordinary contexts.

REFERENCES

CHATER, N., & OAKSFORD, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology*, **38**, 191-258.

CUMMINS, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition*, **23**, 646-658.

CUMMINS, D. D., LUBART, T., ALKSNIS, O., & RIST, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, **19**, 274-282.

GROSSET, N., BARROUILLET, P., & MARKOVITS, H. (2005). Chronometric evidence for memory retrieval in causal conditional reasoning: The case of the association strength effect. *Memory & Cognition*, **33**, 734-741.

JANVEAU-BRENNAN, G., & MARKOVITS, H. (1999). Reasoning with causal conditionals: Developmental and individual differences. *Developmental Psychology*, **35**, 904-911.

JOHNSON-LAIRD, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

JOHNSON-LAIRD, P. N., & BYRNE, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.

LIU, I. (2003). Conditional reasoning and conditionalization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 694-709.

MARKOVITS, H., & BARROUILLET, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review*, **22**, 5-36.

MARKOVITS, H., FLEURY, M.-L., QUINN, S., & VENET, M. (1998). The development of conditional reasoning and the structure of semantic memory. *Child Development*, **64**, 742-755.

MARKOVITS, H., & QUINN, S. (2002). Efficiency of retrieval correlates with "logical" reasoning from causal conditional premises. *Memory & Cognition*, **30**, 696-706.

MARKOVITS, H., SCHLEIFER, M., & FORTIER, L. (1989). The development of elementary deductive reasoning in young children. *Developmental Psychology*, **25**, 787-793.

MARKOVITS, H., & VACHON, R. (1990). Conditional reasoning, representation, and level of abstraction. *Developmental Psychology*, **26**, 942-951.

OAKSFORD, M., & CHATER, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, **101**, 608-631.

OAKSFORD, M., CHATER, N., & LARKIN, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 883-899.

QUINN, S., & MARKOVITS, H. (1998). Conditional reasoning, causality, and the structure of semantic memory: Strength of association as a predictive factor for content effects. *Cognition*, **68**, B93-B101.

QUINN, S., & MARKOVITS, H. (2002). Conditional reasoning with causal premises: Evidence for a retrieval model. *Thinking & Reasoning*, **8**, 179-191.

RIPS, L. J. (2001). Two kinds of reasoning. *Psychological Science*, **12**, 129-134.

THOMPSON, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition*, **22**, 742-758.