Efficiency of retrieval correlates with “logical” reasoning from causal conditional premises

HENRY MARKOVITS and STÉPHANE QUINN
University of Quebec, Montreal, Quebec, Canada

In two experiments, we examined the prediction that there should be a relation between the speed with which subjects can retrieve potential causes for given effects and their reasoning with causal conditional premises (if cause P, then effect Q). It was also predicted that when subjects are given effects for which there exists a single strongly associated cause, speed of retrieval of a second potential cause should be particularly related to reasoning with invalid logical forms—namely, affirmation of the consequent and denial of the antecedent. In the first experiment, 49 university students were given both retrieval tasks and conditional reasoning problems. The results were generally consistent with the predictions. The second experiment, involving 57 university students, replicated the first, with some methodological variations. The results were also consistent with the predictions. An analysis of the combined results of the two experiments indicated that individual differences in efficiency of retrieval of information from long-term memory did predict performance on the invalid logical forms in the predicted ways. These results strongly support a retrieval model for conditional reasoning with causal premises.

Deductive reasoning is an important component of human cognition. In fact, the ability to make inferences on the basis of some given premises underlies much of scientific reasoning and is a key to understanding how human intelligence functions. There is now a great deal of empirical evidence that children and adults have difficulties in reasoning *logically*—that is, that they often give logically incorrect responses to deductive reasoning problems. There is, in addition, a great deal of individual variability in the kinds of responses that are given to deductive problems. This, in turn, raises the question of just what are the characteristics of a *good reasoner*. In the following, we will suggest that one characteristic of good reasoning, in the specific case of deductive reasoning with causal conditional premises, is the speed of retrieval of information from memory.

One notion that is central to many theories of deductive reasoning is that reasoning is inherently constrained by the limited capacity of working memory. Both of the two principal theories of reasoning, the mental model theory (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991) and rules-based theories (Braine & O'Brien, 1991; Rips, 1994), allow the inference that individual differences in

working memory capacity should correlate with the ability to give the logically appropriate responses to more or less complex inferential problems. There is, in fact, some evidence that this is indeed the case (Barrouillet, 1996; Barrouillet & Lécas, 1999; Kyllonen & Christal, 1990; see also Torrens, Thompson, & Cramer, 1999, for an intriguing variant of this idea).

Although the capacity of working memory is clearly a plausible candidate for the explanation both of overall differences in the resolution rate of problems of differing levels of complexity and of individual differences on this dimension, there is an important class of variation in reasoning performance that cannot be thus explained. Specifically, a great deal of empirical research has shown that performance on formally identical problems varies greatly when the specific content of the premises used for reasoning varies (e.g., Cummins, 1995; Cummins, Lubart, Alksnis, & Rist, 1991; Marcus & Rips, 1979; O'Brien, Costa, & Overton, 1986; Thompson, 1994). Such variation is difficult to explain solely by constraints on the capacity of working memory to manipulate more or less complex representations, and it is likely that some other factors are involved. In the following, we will examine a model of conditional (if-then) reasoning that proposes that efficiency of information retrieval from long-term memory is a key component in explaining such variation and that individual differences in this factor should be correlated with content-related differences in reasoning performance (Markovits & Barrouillet, 2002).

Conditional reasoning is a critical component of logical thinking and has been the subject of many empirical and theoretical studies. Conditional reasoning, in its most basic sense, involves making inferences with a given major premise of the form “P implies Q” and one of four possible minor premises. Modus ponens (MP) is the logical

Preparation of this manuscript was supported by scholarships from the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (FCAR) and from the Natural Sciences and Engineering Research Council of Canada (NSERC) to S.Q. and by grants from the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (FCAR) and from the National Sciences and Engineering Research Council of Canada (NSERC) to H.M. The authors thank Michael Wenger, Alessandra Tasso, and an anonymous reviewer of a previous version of this manuscript for their very helpful suggestions. Correspondence concerning this article should be addressed to H. Markovits, Département de Psychologie, Université du Québec à Montréal, C.P. 8888, Succ. “A,” Montréal, PQ, H3C 3P8 Canada (e-mail: markovits.henry@uqam.ca).

principle that involves reasoning with the premises “P implies Q, P is true” and leads to the logically correct conclusion “Q is true.” Modus tollens (MT) involves reasoning with the premises “P implies Q, Q is false” and leads to the logically correct conclusion “P is false.” These two are valid logical forms, since they both lead to a single, logically correct conclusion. Affirmation of the consequent (AC) involves reasoning with the premises “P implies Q, Q is true.” Denial of the antecedent (DA) involves reasoning with the premises “P implies Q, P is false.” In both cases, the implied conclusions—“P is true” for AC and “Q is false” for DA—are not logically correct. Neither of these forms leads to a single, logically correct conclusion, and the correct response would be to deny the implied (biconditional) conclusion in both cases.

Empirical studies clearly show that inferential behavior is characterized by a great deal of variability (Cummins, 1995; Cummins et al., 1991; Marcus & Rips, 1979; O’Brien et al., 1986; Thompson, 1994). One of the key factors that appears to account for much of this variation is the content of the premises that are used to make deductions. Specifically, reasoners make very different inferences for logically equivalent problems that vary only in the specific content of the major premises (e.g., Cummins, 1995; Cummins et al., 1991; Quinn & Markovits, 1998; Thompson, 1994). Neither of the two major classes of theories that have been proposed to account for inferential behavior provides a clear explanation for such variability. Rules-based theories propose that there exists an initial stage in the process of making a deduction in which natural language premises are interpreted in accordance with pragmatic and semantic principles (Rips, 1994; Rumin, Connell, & Braine, 1983). However, there have been few attempts to specify in enough detail how this stage might work to allow understanding of how variability might affect inferential behavior. Mental model theory, as put forward by Johnson-Laird (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; Johnson-Laird, Byrne, & Schaeken, 1992), proposes that reasoners construct mental models of premises that correspond to possible states of the world that are consistent with these premises. An important postulate in this theory is that reasoners will start with an initial model of the major premise that is as simple as possible. Under some circumstances, this initial model will be “fleshed out”—that is, other models that are also consistent with the major premise will be added to this initial representation. This process will be done in a way that reflects the reasoner’s semantic interpretation of the meaning of the premises. However, the fleshing-out process has never been explained in enough detail to provide for clear explanations of content-related variability.

Markovits (Markovits, 2000; Markovits & Barrouillet, 2002; Markovits, Fleury, Quinn, & Venet, 1998) has recently presented a mental model account of conditional (if-then) reasoning that provides an explicit account of the fleshing-out process. In the following, we will briefly describe this model and then examine some predictions

concerning individual differences in reasoning and access to memory that are derived from it.

This account of conditional reasoning follows that of Johnson-Laird (Johnson-Laird & Byrne, 1991) in many respects. However, in order to account for both developmental and content-related variability, it is assumed that reasoners will often make use of their existing knowledge base about the relations involved in a given major premise in order to activate related information that is stored in long-term memory. Specifically, it is assumed that there are at least two major classes of information concerning conditional relations that are readily available to reasoners. The most easily accessible of these concerns possible cases of [not-p and not-q], which we refer to as the *complementary class*. Reasoners will also attempt to access possible cases of [not-p and q], referred to as the *alternatives class*. For example, consider the major premise “If a rock is thrown at a window, the window will break.” For this premise, “a twig is thrown at the window, the window does not break” is an event in the complementary class, whereas “a bat is thrown at the window, the window breaks” is an event in the alternatives class. It should be noted that there is a third class of information, which concerns potential disabling conditions (see, e.g., Cummins, 1995)—that is, conditions that can allow the antecedent to be true while the consequent is false. An example of this is the possibility that a window is made of Plexiglas, which would allow the conclusion that “if a rock is thrown at a window, which is made of Plexiglas, the window will not break.” Although the number of available disabling conditions has been shown to affect inferences when conditional premises are presented with no specific instructions (Cummins, 1995), most educated adults can use explicit instructions that they should suppose that the premises are true to inhibit retrieval of this class of information, unless these instructions are followed by subsequent demands for information retrieval (Markovits & Potvin, 2001). Since, in the present context, we employed just such explicit instructions, we did not consider this factor.

Specifically, this account of conditional reasoning assumes that the initial representation of the premises is basically the same as that postulated by the model of Johnson-Laird (Johnson-Laird & Byrne, 1991). That is, reasoners start by constructing a model of the major premise that contains only the antecedent and consequent terms of the major premise. For example, suppose that a reasoner is given the major premise: If it rains, then the street will be wet. Their initial model would look like the following:

Rain	street wet
...	

where the second line represents the possibility of other models’ being potentially available.

We then assume that when reasoners are presented with a given minor premise, they will tend to use both the minor

premise and the major premise as retrieval cues in order to activate one or more of the classes of information mentioned previously—the complementary class and/or the alternatives class (Markovits, 2000; Markovits & Barrouillet, 2002). If a given type of information is sufficiently strongly activated, it will be incorporated into the representation of the premises, and the reasoner will make inferences on this basis. Specifically, mental model theory (Johnson-Laird & Byrne, 1991) supposes that reasoners will generate an inference if there is at least one model that supports the inference and if there is no other model that contradicts it. In any other case, there will be no necessary conclusion drawn.

For example, suppose that after being given the major premise “If it rains, then the street will be wet,” the reasoner is given the minor premise corresponding to the AC logical form (“the street is wet”). The minor premise will be used as a cue to activate potential alternative antecedents [not- p and q —that is, other ways of making a street wet. If one such alternative (e.g., “the street cleaner has passed”) is sufficiently strongly activated, it will be incorporated into the reasoner’s initial representation of the problem:

Rain	street wet
Street cleaner	street wet

With these two models, our reasoner will tend to answer that “if the street is wet, then it may or may not be true that it has rained”—that is, they will give an uncertainty response to the AC form.

This basic framework thus supposes that one of the problems facing reasoners who are given these kinds of premises involves the retrieval of pertinent information from memory. Under less constrained circumstances, retrieval would generally not be a problem. However, information retrieval during reasoning is made more difficult by the requirement to actively maintain premises in working memory, which correspondingly reduces the cognitive capacity available for the retrieval process (Anderson, Reder, & Lebiere, 1996; Rosen & Engle, 1996).

This analysis allows the general conclusion that constraints on information retrieval during reasoning should have an impact on the kinds of inferences that are generated. We can also make some specific predictions about how retrieval constraints should affect reasoning on each of the four logical forms of conditional reasoning. In order to do this, we will present a simplified representation of the retrieval processes involved in reasoning, using a multinomial process tree (MPT) model (e.g., Batchelder & Riefer, 1999). We denote the probability of retrieving a member of the alternatives class on any given logical form by a and the probability of retrieving a member of the complementary class by c . The nature of the representation that will be used by the reasoner will then depend on the outcome of the retrieval process, which can lead to retrieval of one or both of the alternatives and complementary classes. This implies that a reasoner can generate one of four potential model sets. The first one, which we will

refer to as M1, consists of only the elements corresponding to the initial model:

p	q
-----	-----

The second, which we will refer to as M2, consists of the initial model with a member of the complementary class:

p	q
not- p	not- q

The third, which we will refer to as M3, consists of the initial model with a member of the alternatives class:

p	q
not- p	q

The final model set, which we will refer to as M4, consists of the initial model with a member of the alternatives class and a member of the complementary class:

p	q
not- p	q
not- p	not- q

Figure 1 presents the corresponding MPT model that provides a description of the assumed retrieval processes involved in generating these models. This allows calculation of the probabilities of generating the four potential models. Specifically, $\text{Prob}(M1) = (1 - a)(1 - c)$, $\text{Prob}(M2) = (1 - a)c$, $\text{Prob}(M3) = a(1 - c)$, and $\text{Prob}(M4) = ac$. Since we are specifically interested in the extent to which variation in reasoning performance might be related to variation in retrieval processes, we make the assumption that the kind of inference that is made for a given inferential problem is determined by the nature of the model set that is used for reasoning with the standard mental models algorithm. Specifically, reasoners will make an inference if there is at least one model that supports the inference and if there is no other model that con-

Figure 1. Multinomial process tree retrieval model for conditional reasoning.

tradicts it. In any other case, there will be no necessary conclusion drawn. Thus, it is possible to calculate the probability of a reasoner's making a given inference or not by considering the probability of generating each of the four possible model sets and the specific conclusion that is allowed for each model. Table 1 gives a résumé of the conclusions that would be generated for each of the four logical forms as a function of the model set that is used for reasoning. In order to examine how responding is related to retrieval, we calculated the ratio of the probability of generating a response that was equivalent to the logical norm to the probability of generating a response that was not the logical norm, which we will refer to as R , for each of the four logical forms. It must be remarked that the estimates generated by this model remain only approximations, since there are other factors that have not been included in the model.

For the MP form ("p is true"), each of the four model sets leads to the same conclusion, that "q is true." In fact, performance on the MP form with standard instructions to suppose that the premises are true is often close to ceiling (Evans, 1993). This analysis also allows the conclusion that there should be little relation between retrieval of alternatives and reasoning with MP.

For the AC form ("q is true"), the logical response is that of uncertainty. An examination of Table 1 shows that this response is generated with model sets M3 and M4, whereas with models sets M1 and M2, the reasoner will generate the (logically false) conclusion that "p is true." The probability of generating a response of uncertainty is then equal to $a(1 - c) + ac$, which equals a . The probability of generating a nonlogical response is equal to $(1 - a)(1 - c) + (1 - a)c$, which equals $1 - a$. Thus, $R(AC) = a/(1 - a)$. Thus, performance on the AC form will vary according to the probability of accessing an alternative antecedent.

For the DA form ("p is false"), the logical response is uncertainty. This response is generated with model sets M1 and M4. The probability of generating a response of uncertainty is, then, equal to $[ac + (1 - a)(1 - c)]$. The probability of generating a nonlogical response is equal to $(1 - a)c + a(1 - c)$. This leads to $R(DA) = [1/(c + a - 2ac) - 1]$. Thus, performance on the DA form will also

vary according to the probability of accessing an alternative antecedent. The inclusion of the c parameter in this equation reflects the particular dynamic of this form. Specifically, the minor premise "p is false" functions as a retrieval cue for both alternatives and complementary cases. This should make the probability of retrieving an alternative generally lower than that for the AC form. In fact, an examination of two studies that have asked for justifications for conditional reasoning problems (Markovits, 2000; Markovits et al., 1998) indicates that this is the case. It will then be generally more difficult to retrieve an alternative with the DA form than with the AC form. This should make the relation between individual efficiency in retrieval of alternatives and reasoning stronger for the DA form than for the AC form. (Note that on the reasonable assumption that a and c vary in a complementary manner, $R[DA]$ will remain similar to $R[AC]$, something that is reflected in the overall similarity in performance with these two logical forms [Evans, 1993].)

For the MT form ("q is false"), the logical response is "p is false." This response is generated with model sets M2 and M4. The probability of generating the logical response is, then, equal to $(1 - a)c + ac$, which equals c . The probability of generating a nonlogical response is $(1 - a)(1 - c) + a(1 - c)$ which equals $1 - c$. Thus, $R(MT) = c/(1 - c)$. Thus, performance with the MT form should not vary with the probability of accessing an alternative antecedent.

Therefore, under this somewhat simplified model, we can predict that performance with both the AC and the DA forms should vary with the probability of retrieving an alternative antecedent during reasoning and that this factor should not significantly affect reasoning with either the MP or the MT form. In the specific context of individual differences, we can predict that reasoners who are more efficient in retrieving alternative antecedents for a given major premise during reasoning should produce relatively more uncertainty responses to the AC and DA forms, with this relation being stronger for the DA form than for the AC form.

Now, the measures that we propose to use for examining individual differences in retrieval efficiency rely on some recent results that show that one important element in overall problem difficulty is the necessity of a reasoner's

Table 1
Conclusion Allowed by Each of the Four Possible Model Sets (M1–M4)
for Each of the Four Logical Forms of Conditional Reasoning

Set	Models	Probability	Inference Made on Each of the Four Logical Forms			
			MP (p is true)	AC (q is true)	DA (p is false)	MT (q is false)
M1	p q	$(1 - a)(1 - c)$	q is true	p is true	uncertain	uncertain
M2	p q	$(1 - a)c$	q is true	p is true	q is false	p is false
M3	not-p not-q	$a(1 - c)$	q is true	uncertain	q is true	uncertain
	p q					
M4	not-p q	ac	q is true	uncertain	uncertain	p is false
	p q					
	not-p not-q					

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

changing semantic fields in order to generate an alternative antecedent (e.g., Quinn & Markovits, 1998). Specifically, an adult reasoner who is given a major premise for which the antecedent term is strongly associated to the consequent, such as "if a dog has fleas, then the dog will scratch," will find it more difficult to produce uncertainty responses to the AC and DA forms than he or she would with the premise "if a dog is unwashed, then the dog will scratch." The reason for this is that in the latter case, generating an alternative antecedent requires overcoming the strong association between "fleas" and "scratching" in order to generate an alternative such as "unwashed," whereas in the former case, this strong association can be used to generate an alternative, "fleas." This is, in fact, quite consistent with Rosen and Engle's (1996) analysis of the effects of working memory load on information retrieval. They concluded that there were four components involved in an information retrieval task. The first was the level of activation spreading from the cue, which functions relatively automatically. The other three involved self-monitoring, suppression of previous responses, and generation of new cues, which require strategic retrieval. According to Rosen and Engle, differences in the strategic components determine the difference between more and less efficient retrievers of information. Our analysis of retrieval processes in reasoning suggests a similar conclusion, since processes such as suppression of activated responses and generation of new cues would be required to generate alternatives with premises such as "if a dog has fleas, then the dog will scratch."

In order to examine the relation between retrieval and reasoning, our basic measure was the following. We presented subjects with effects for which there was one cause that was strongly associated to the given effect, whereas others were less easily activated or belonged to differing semantic fields (see Quinn & Markovits, 1998)—for example, the effect "a dog scratches a lot," for which people easily retrieve the cause "the dog has fleas," whereas other potential causes are generally unrelated to this (e.g., "skin dirty") and are less easily retrieved. We asked the subjects to generate (at least) two potential causes as quickly as possible. In this case, they would tend to activate the most highly associated cause first and with relatively little effort, given the strong association between the cause and the given effect. Then they would have to activate another cause, which involves retrieving a relatively weakly associated cause, while the first one remained in working memory. In this second instance, we could expect that the speed with which a subject could retrieve a second cause would reflect their use of the kinds of strategic components required to overcome the effects of having generated the initial cause. Thus, the relative speed with which a reasoner can generate a second cause in this situation should be related to their overall retrieval efficiency. We would then predict that individual differences in responding to the AC and DA forms would be particularly reflected in the relative time required to produce a second cause. Time

required to produce an initial cause should reflect more associative processes and should be less indicative of retrieval efficiency and, thus, less highly related to individual differences in reasoning.

Finally, the items used to examine reasoning involved causal premises ("If cause P, then effect Q). These kinds of items were chosen for their relative difficulty. When one reasons with causal premises, generating alternatives to a given cause P requires looking for "things other than P that lead to the same effect Q." This requires the reasoner to use a naive causal theory (Cummins, 1995) in order to construct an appropriate definition that can be used as a retrieval cue (e.g., things that can make a street wet). Except in particular cases, such retrieval cues would lead to an ad hoc class (Barsalou, 1983), which must be reconstructed by the subject on line and whose elements are more difficult to retrieve than corresponding class-based information. Note that the specific items used to measure conditional reasoning differed from those used to measure retrieval speed.

In the following, we will present two experiments in which we examined the relation of individual differences in speed of retrieval of one or two alternative causes as a predictor of reasoning performance with causal conditional premises.

EXPERIMENT 1

In this first experiment, we looked at the predicted relationship between conditional reasoning and retrieval times for first and second causes. In order to do this, subjects were given a series of conditional reasoning problems in a paper-and-pencil format. They were also given a series of verbally presented items and were asked to generate three possible antecedents as rapidly as possible. Half of these items used the kinds of causally related effects discussed above (referred to as *causal* items), whereas half used class-based items (requiring retrieval of members of a given class), which were used as a control.

Method

Subjects

A total of 49 university students, with French as their first language (40 females, 9 males; average age, 24 years 5 months) took part in this experiment. Each subject was a volunteer.

Materials

Reasoning task. A paper-and-pencil test of conditional reasoning was constructed. On the front page, the subjects were asked for their age and sex. Below this, were the following instructions (translated from the original French):

In the following pages, you will have to answer a certain number of questions. On each page you will first see a statement. *You must suppose that this statement is always true* and then choose, for each of the four following questions, the response that you consider to be the best. Please read each statement and the questions that follow attentively.

On the top of each of the following nine pages, was the statement "Suppose that it is always true that:" followed by one of nine if-then major premises. After each major premise, there were four questions

corresponding to the four logical forms, MP, AC, DA, and MT. Each question was composed of a statement involving either the antecedent or the consequent of the major premise, followed by three possible choices. For example, the question corresponding to AC with the major premise "If a person has a cavity, then they will have a toothache" was

- A person has a toothache,
 (a) It is certain that the person has a cavity.
 (b) It is certain that the person does not have a cavity.
 (c) One cannot be certain that the person has a cavity or not.

Although each major premise was followed by four questions, the order of these questions was different each time. The specific premises used are shown in the Appendix. Half of the premises referred to concrete content and were used for evaluating reasoning performance in this context. The other premises had content that contained abstract referents for the consequent terms and sometimes for the antecedent (this was done in order to preclude retrieval of a specific alternative antecedent in these cases) and were used primarily to offset any possibility of subjects' carrying over responses from one concrete problem to another. A pretest had previously shown that these items were at a level of difficulty appropriate to the kinds of subjects used in the experiment.

Retrieval task. A recording on audio tape was used to present the items in the retrieval task. At the beginning of the tape, the subject heard an introduction in which the basic task was presented. The subjects were told that they would be given a statement asking for a possible antecedent for a given consequent, which they would have to produce as quickly as possible. In each case, they were then asked to produce another antecedent for the same consequent. In some cases, they were asked to produce a third or a fourth antecedent. They were given a practice trial and could ask the experimenter to repeat the instructions if they were in any doubt. The statements that were presented are given in the Appendix.

A short audible beep was heard directly after each of the statements. Two sorts of relations were used here. Four of the basic items were taken from a previous study and were chosen because they were causal conditionals for which only a single possible cause was highly associated with the given antecedent (see Quinn & Markovits, 1998, for details). These were (1) something that would make the pupil of an eye narrow, (2) something that would make a dog scratch a lot, (3) something that would make a developed film black, and (4) something that would make a finger bleed. The other items referred to class-based relations and were included as a control.

A second version of the audio tape was also prepared. This was identical to the first, except that the order of the items was inverted.

Procedure

The subjects were seen in individual sessions that lasted approximately 30 min. Both the relative order of the two tasks and the order of the items in the retrieval task were systematically counter-balanced. Half of the subjects received the reasoning task followed by the retrieval task, whereas the other half received the two tasks in the inverse order.

Reasoning task. The subjects were given the paper-and-pencil conditional reasoning task in a small quiet room. They were left

alone and were given as much time as they needed to complete the task.

Retrieval task. The subjects were in a small quiet room with a single experimenter. The entire session, including all statements and the subject's responses, were recorded on audio tape, using a microphone placed directly in front of the subject. The experimenter sat in front of and facing sideways to the subject. The experimenter ran the audio tape and, when necessary, explained the task if the instructions on the audio tape were not understood.

Subsequently, the content of the audio tapes were transferred onto a computer and were read, using a sound analysis program. Reaction times (RTs) were determined by blocking out the space between the end of the wave form corresponding to the audible beep and the beginning of the wave form corresponding to the answer. Note that the overall level of variability of this procedure was about 1% of the total time measured (e.g., for a latency of 1 sec, variability was 0.01 sec; for a latency of 3 sec, variability was .03 sec).

Results

We first performed a visual inspection of reaction times, which indicated that responses taking more than 8 sec were outliers. These were eliminated from calculation of mean response times. Table 2 indicates the mean number of correct responses to the four reasoning problems and the mean RTs in seconds for the first and second antecedents for both the causal and the class items.

In order to calculate correlations, all reaction times were converted to natural logarithms. (It should be noted that we also performed the same analyses without these manipulations. These gave similar results, although the correlations were generally stronger. In the following, we will present the more conservative analyses.) Our first analysis looked at the relation between responses on the reasoning problems and mean RTs for the causal items and for the class items. Table 3 gives the correlations between these means and performance on the reasoning items. An inspection of this table shows an overall pattern that is generally consistent with our global hypotheses. First, correlations between performance with MT and RT measures were either close to zero or slightly positive for both causal and class-based measures. Correlations between performance with MP and RTs were negative and of higher values than those for MT, but since performance with MP was close to ceiling, this was based on a small amount of variance. Correlations between performance with AC or DA and RTs were generally in the predicted direction. Correlations between AC and DA performance and RTs were also generally higher for causal items than for class items. In addition, correlations between AC or DA performance and time to produce a second cause were

Table 2
Mean Performance on Conditional Reasoning Problems and Mean Latencies for First and Second Causes Produced on the Causal Items and on the Class Items (With Standard Deviations)

Conditional Reasoning Problems								Retrieval Measures							
MP		MT		AC		DA		First Cause		Second Cause		First Class		Second Class	
<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
4.39	0.76	2.94	1.59	3.78	1.40	3.49	1.67	1.00	0.57	1.45	0.99	0.81	0.59	0.58	0.37

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

Table 3
Correlations Between Mean Reaction Times on Causal
and Class Items and Performance on the Four Logical Forms
With Concrete Content for Experiment 1

	MP	MT	AC	DA
Reaction times for first causal items	-.10	.20	-.10	-.15
Reaction times for second causal items	-.19	.11	-.23	-.43**
Reaction times for first class items	-.24	-.05	-.07	-.11
Reaction times for second class items	-.20	.10	-.21	-.14

Note—MP, modus ponens; MT, modus tollens; AC, affirmation of the consequent; DA, denial of the antecedent. ** $p < .01$.

higher than correlations between AC or DA reasoning and time to produce the first cause, although only the correlation between performance with DA and second-cause RT was significant.

In order to look at this latter effect more globally, we combined performances on the two uncertain logical forms, AC and DA. The correlation between RTs and number of uncertainty responses to AC and DA combined for the first cause was indeed globally smaller ($-.139$) than that observed for the second cause ($-.366$). In order to provide a clearer picture of the relations between reasoning and the two RT measures, we also computed partial correlations for both RT measures. The correlation between RT for the first cause and performance on the two uncertain forms, with RT for the second cause partialled out, was $.053$. The correlation between RT for the second cause and performance on the two uncertain forms, with RT for the first cause partialled out, was $.345$. These results are consistent with the idea that time to produce a second cause is a better predictor of reasoning with the two uncertain forms than time to produce a first cause, with the kinds of items used in this study.

EXPERIMENT 2

The results of the previous experiment are generally consistent with the predictions made. In this experiment, we replicated these results, using two modifications. In the previous experiment, the subjects received fairly long descriptions of the causes that they were to produce in the RT task. The length of these requests could increase variance, owing to the possibility that some subjects could anticipate the request and start retrieval early, to an extent that is impossible to evaluate. In order to reduce any such effect, we simplified the presentation of the stimuli. The subjects were given a short description of the effect and then were asked to produce three possible causes as quickly as possible. We also added some extra items having a single strongly associated cause to the four that were used in the first experiment. Finally, as control items, we used causal conditionals for which there were many possible causes that are easily accessible.

Method

Subjects

A total of 57 university students, with French as their first language (48 females, 9 males; average age, 23 years 7 months) took part in

this experiment. Each subject was a volunteer and was offered the chance to participate in a lottery with a \$200 prize as recompense.

Materials

Reasoning task. The same test of conditional reasoning as that in the previous study was used.

Retrieval task. A recording on audio tape was used to present the items in the retrieval task. At the beginning of the tape, the subject heard an introduction in which the basic task was presented. The subjects were told that they would be given a statement corresponding to an effect. They were then told that they were to produce, as rapidly as possible, three possible causes for this effect, immediately after having heard the effect. The subjects were also told that they were to produce realistic causes and that they were not to repeat the same cause twice, even if the language used was different. They were given a practice trial and could ask the experimenter to repeat the instructions if they were in any doubt. The items used are presented in the Appendix.

A short audible beep was heard directly after each of the statements. Items 1 (yawns), 3 (dog scratches), 5 (finger bleeds), 7 (sneezes), 9 (film), 11 (room lit), 13 (weight), and 15 (pupil) were chosen to have a single strongly associated cause (see also Quinn & Markovits, 1998), whereas the other items were chosen to have several possible causes that were readily available. A second version of the audio tape was also prepared. This was identical to the first, except that the order of the items was inverted.

Procedure

The procedure was identical to that used in Experiment 1.

Results

We first looked at the relation between RTs to the retrieval items and performance on the reasoning problems. As in the first experiment, we eliminated responses that took more than 8 sec to be produced. In addition, all RTs were converted to natural logarithms. Table 4 indicates the mean number of correct responses to the four reasoning problems and the mean RTs in seconds for the first and second causes for both the single- and the multiple-cause items. (The RT for the third cause in the sequence was not analyzed, since observation of responses indicated that, for many subjects, production of a third cause was highly variable, with repetitions of previous responses and hesitations frequently produced.) Table 5 gives the correlations between these means and performance on the reasoning items. As can be seen, the pattern of correlations was very similar to that obtained in the first experiment. First, correlations between performance on MP or MT and RT measures were positive and nonsignificant. Correlations between AC and DA and RTs were in the predicted direc-

Table 4
Mean Performance on Conditional Reasoning Problems and Mean Latencies for First and Second Causes Produced on the Single-Cause Items and on the Multiple-Cause Items (With Standard Deviations)

Conditional Reasoning Problems								Retrieval Measures							
MP		MT		AC		DA		First Single		Second Single		First Multiple		Second Multiple	
<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
4.60	0.75	3.54	1.56	3.16	1.73	3.07	1.77	2.14	0.69	1.91	0.80	2.22	0.72	1.72	0.82

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

tion, with statistically significant correlations between reasoning and second-cause RTs.

Analysis of Combined Results

In order to provide a stronger statistical base for the basic pattern of correlations, we combined results across the two RT studies. For the RT measures, this was done by using only the four causal premises that were used in the first experiment and then normalizing the RTs within each experiment in order to account for differences in problem presentation. Table 6 presents the correlations between the normalized RT measures for production of the first and the second causes to the four causal premises and performance on the uncertain logical forms. The overall correlations clearly showed the predicted pattern. Correlations between RTs and performance on MP and MT were close to zero. Correlations between RTs and performance on the AC and DA forms were in the predicted direction and were significant only for second-cause generation. The correlation between performance on the AC form and mean RTs was significantly higher for the second cause than for the first cause [$t(103) = 1.87, p < .05$, one-tailed]. This was also true for the DA form [$t(103) = 1.84, p < .05$, one-tailed]. Finally, as was predicted, the correlation between mean RTs for the second cause and performance on the DA form was significantly higher than that for performance on the AC form [$t(103) = 1.77, p < .05$, one-tailed].

In order to provide a more synthetic view of what these correlations involved, we examined subjects whose mean retrieval times for the second cause were either one standard deviation or more above the mean (the slow group) or one standard deviation or more below the mean (the fast group). Mean percentage of correct responses to the four logical forms for the slow group were MP = 90.0%, MT = 60.0%, AC = 54.4%, and DA = 43.4%. The cor-

responding values for the fast group were MP = 90.6%, MT = 54.6%, AC = 80.0%, and DA = 85.4%. The difference between the slow and the fast groups was significant for both AC [$t(31) = 2.31, p < .03$] and DA [$t(31) < 4.27, p < .001$], but not for the MP and MT forms. The difference in correct responding between the slow and the fast groups was indeed concentrated in the two uncertain forms, AC and DA. More interesting, the relative difference between fast and slow retrievers on these forms was very large, in both cases.

We then looked at one further factor. In these two experiments, reasoners were asked to make inferences with both concrete and abstract premises. The basic hypothesis of the existence of a relation between retrieval speed and reasoning applies specifically to premises with concrete content. Although the processes involved in reasoning with abstract premises are not well understood, there is evidence that they are quite different (Markovits, Doyon, & Simoneau, 2002; Venet & Markovits, 2001). Unless some form of analogical transfer is used as a strategy with concrete premises, it would seem reasonable to suppose that the kind of knowledge-based retrieval that is postulated for concrete reasoning should not apply to abstract premises. Generally, performance on the abstract items was lower than that on the concrete items. Percentage of correct responses for the MP, MT, AC, and DA forms for the abstract items were 76%, 57%, 44%, and 54%, respectively, whereas for the concrete items, the corresponding values were 90%, 65%, 69%, and 65%. Performance on the two forms of each problem was generally highly correlated (these correlations varied between .55 for the MP form and .73 for the DA form). We thus looked at correlations between RT measures for second causes and performance on the abstract problems when performance on the concrete problem was partialled

Table 5
Correlations Between Mean Reaction Times on Single-Cause and Multiple-Cause Items and Performance on the Four Logical Forms With Concrete Content in Experiment 2

	MP	MT	AC	DA
Reaction times for first single-cause items	.17	.20	-.19	-.28*
Reaction times for second single-cause items	-.01	.11	-.30*	-.46**
Reaction times for first multiple-cause items	.05	.16	-.19	-.36**
Reaction times for second multiple-cause items	.13	.15	-.15	-.38**

Note—MP, modus ponens; MT, modus tollens; AC, affirmation of the consequent; DA, denial of the antecedent. * $p < .05$. ** $p < .01$.

Table 6
Correlations Between Normalized Mean Reaction Times
on Causal Items and Performance on the Four
Logical Forms With Concrete Content Combined
Over Both Reaction Time Experiments

	MP	MT	AC	DA
Reaction times for first cause	-.04	.09	-.05	-.17
Reaction times for second cause	-.08	-.02	-.26**	-.37**

Note—MP, modus ponens; MT, modus tollens; AC, affirmation of the consequent; DA, denial of the antecedent. ** $p < .01$.

out, for each of the four logical forms. For AC and DA, the partial correlations were $-.06$ and $.08$, respectively. For MP and MT, these correlations were $.07$ and $.02$, respectively. As a comparison, correlations between RTs for second causes and performance on the concrete problems when performance on the abstract problem was partially out, for AC and DA, were $-.16$ and $-.32$, respectively. Thus, as might be expected, performance on the abstract problems was not related to retrieval speed.

GENERAL DISCUSSION

Our basic idea was that for the relatively straightforward conditional inferences examined here, individual variation in performance could be explained (at least partly) by efficiency of information retrieval from long-term memory. Specifically, in order for reasoners to produce an uncertainty response to the two invalid forms, AC and DA, they must be able to retrieve potential alternative antecedents from memory. This is made difficult by the necessity of retaining premises in working memory and the particular nature of causal premises, which require searching for alternatives in a semantic field that often differs from that suggested by the premises. Following Rosen and Engle (1996), we predicted that when reasoners were given effects for which there is a single strongly associated cause and were asked to retrieve two potential causes, speed of retrieval of the second cause would be a better indicator of retrieval efficiency than would speed of retrieval of the first cause. The results of these experiments were quite consistent with these predictions. They showed that there was a clear correlation between speed of retrieval of alternate causes from long-term memory and performance on the two invalid forms, AC and DA, and that this correlation was higher for retrieval of second causes than for retrieval of first causes. Specifically, as was predicted, reasoners who were able to more quickly retrieve information concerning alternative causes for a given effect, subsequent to having generated the most highly associated one, also responded to these logical forms more often with a "correct" uncertainty response. Responses to the two certain forms, MP and MT, were not related to speed of retrieval of this form of information.

There is one important question that is raised by this conclusion. This concerns whether it is really useful to specifically distinguish retrieval effects from those related to the capacity of working memory to manipulate more or

less complex representations or algorithms. After all, it could be claimed that these results might simply reflect basic limitations in working memory capacity and are thus consistent with the kinds of analyses that are already made by both mental model and rules theories. However, this cannot account for the fact that these results show a correlation between retrieval speed and performance on the AC and DA forms, but not on the MT form. If the observed retrieval effects were simply artifacts of global limitations in working memory capacity, one would expect a correlation with the MT form also, particularly since this form has a success rate comparable to that for the AC and DA forms. In fact, both rules theories and mental model theory consider that reasoning correctly with the MT form involves greater processing demands than with the other forms, which would imply that there should be a stronger relation between working memory capacity and MT performance. The latter conclusion is indeed supported by a study by Toms, Morris, and Ward (1993), which found that a decrease in available working memory (via a concurrent memory load task) affected performance on MT, but not on the other logical forms. Thus, the lack of a correlation between retrieval speed and MT performance clearly shows that the obtained effects were not simply artifacts of more basic information-processing constraints.

More generally, both our theoretical context and the results of the present study support a two-stage model of conditional reasoning. The first stage involves analysis of the premises and concurrent retrieval of information concerning complements and alternatives. This results in the generation of a model set with variable numbers of elements. The second stage involves applying a decision algorithm to the model set. Retrieval limitations will principally affect the first stage, whereas information-processing constraints will principally affect the second stage. Our results indicate that one of the important characteristics of a "good" reasoner is the efficiency with which he or she can consider and retrieve alternative antecedents, thus allowing him or her to more easily reject conclusions that are erroneously accepted by less efficient reasoners.

REFERENCES

- ANDERSON, J. R., REDER, L. M., & LEBIERE, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, **30**, 221-256.
- BARROUILLET, P. (1996). Transitive inferences from set-inclusion relations and working memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1408-1422.
- BARROUILLET, P., & LECAS, J. F. (1999). Mental models in conditional reasoning and working memory. *Thinking & Reasoning*, **5**, 289-302.
- BARSALOU, L. W. (1983). Ad hoc categories. *Memory & Cognition*, **11**, 211-227.
- BATCHELDER, W. H., & RIEFER, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, **6**, 57-86.
- BRAINE, M. D. S., & O'BRIEN, D. P. (1991). A theory of if: Lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, **98**, 182-203.
- 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.

- EVANS, J. ST. B. T. (1993). The mental model theory of conditional reasoning: Critical appraisal and revision. *Cognition*, **48**, 1-20.
- 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.
- JOHNSON-LAIRD, P. N., BYRNE, R. M. J., & SCHAEKEN, W. (1992). Propositional reasoning by model. *Psychological Review*, **99**, 418-439.
- KYLLONEN, P. C., & CHRISTAL, R. E. (1990). Reasoning ability is (little more than) working memory capacity? *Intelligence*, **14**, 389-433.
- MARCUS, S. L., & RIPS, L. J. (1979). Conditional reasoning. *Journal of Verbal Learning & Verbal Behavior*, **18**, 199-223.
- MARKOVITS, H. (2000). A mental model analysis of young children's conditional reasoning with meaningful premises. *Thinking & Reasoning*, **6**, 335-348.
- MARKOVITS, H., & BARROUILLET, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review*, **22**, 5-36.
- MARKOVITS, H., DOYON, C., & SIMONEAU, M. (2002). Individual differences in working memory and conditional reasoning with concrete and abstract content. *Thinking & Reasoning*, **8**, 97-107.
- MARKOVITS, H., FLEURY, M.-L., QUINN, S., & VENET, M. (1998). The development of conditional reasoning and the structure of semantic memory. *Child Development*, **69**, 742-755.
- MARKOVITS, H., & POTVIN, F. (2001). Suppression of valid inferences and knowledge structures: The curious effect of producing alternative antecedents on reasoning with causal conditionals. *Memory & Cognition*, **29**, 736-744.
- O'BRIEN, D. P., COSTA, G., & OVERTON, W. F. (1986). Evaluations of causal and conditional hypotheses. *Quarterly Journal of Experimental Psychology*, **38A**, 493-512.
- 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.
- RIPS, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, MA: MIT Press.
- ROSEN, V. M., & ENGLE, R. W. (1996). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, **126**, 211-227.
- RUMAIN, B., CONNELL, J., & BRAINE, M. D. S. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults. *Developmental Psychology*, **19**, 471-481.
- THOMPSON, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition*, **22**, 742-758.
- TOMS, M., MORRIS, N., & WARD, D. (1993). Working memory and conditional reasoning. *Quarterly Journal of Experimental Psychology*, **46A**, 679-699.
- TORRENS, D., THOMPSON, V. A., & CRAMER, K. M. (1999). Individual differences and the belief bias effect: Mental models, logical necessity, and abstract reasoning. *Thinking & Reasoning*, **5**, 1-28.
- VENET, M., & MARKOVITS, H. (2001). Understanding uncertainty with abstract conditional premises. *Merrill-Palmer Quarterly*, **47**, 74-99.

APPENDIX
Items Used in the Reasoning Test and Those Used in the Retrieval Phases of Experiments 1 and 2

Reasoning Test Items	Retrieval Items, Experiment 1	Retrieval Items, Experiment 2
If a person has a cavity, then they will have a toothache.	Something else that would make the pupil of an eye narrow.	A person yawns.
If a camel sleeps, then they will troublette.	Something that would make the pupil of an eye narrow.	There is a noise.
If a person is diabetic, then they will take insulin.	An animal that has wings.	A dog scratches a lot.
If a dichot passes, then it will be virdin.	Another animal that has wings.	A person is tired.
If a rail is displaced in front of a train, then the train will be derailed.	Another animal that has wings.	A finger bleeds.
If Lucy is sick, then she will be vindille.	Something that would make a dog scratch a lot.	A window is broken.
If a person puts their finger on something very hot, then they will burn themselves.	Something else that would make a dog scratch a lot.	A person sneezes.
If the malou falls, then it will be dredon.	Something that has legs.	A person is ill.
If it is autumn, then the leaves will fall from the trees.	Something else that has legs.	A developed film is black.
	Something else that has legs.	A person is wet.
	Something else that has legs.	A room is lit up.
	Something that would make a developed film black.	A worker is late.
	Something that would make a developed film black.	A person gains weight.
	Something that has wheels.	A person sweats.
	Something else that has wheels.	The pupil of an eye narrows.
	Something that would make a finger bleed.	
	Something else that would make a finger bleed.	
	Something that is a tool.	
	Something else that is a tool.	

(Manuscript received June 29, 2001;
revision accepted for publication February 14, 2002.)