

Different developmental patterns of simple deductive and probabilistic inferential reasoning

HENRY MARKOVITS

Université du Québec à Montréal, Montréal, Québec, Canada

AND

VALERIE THOMPSON

University of Saskatchewan, Saskatoon, Saskatchewan, Canada

In three studies, we examined simple counterexample-based and probabilistic reasoning in children 6, 7, and 9 years of age. In the first study, participants were asked to make conditional (*if-then*) inferences under both categorical (certain or uncertain) and probabilistic instructions. Results showed that 6-year-olds respond to both forms of inference in similar ways, but whereas probabilistic conditional inferences showed little development over this period, categorical inferences clearly improved between 6 and 7 years of age. An analysis of the children's justifications indicated that performance under categorical instructions was strongly related to counterexample generation at all ages, whereas this was true only for the younger children for inferences under probabilistic instructions. These findings were replicated in a second study, using problems that referred to concrete stimuli with varying probabilities of inference. A third study tested the hypothesis that children confused probability judgments with judgments of confidence and demonstrated a clear dissociation between these two constructs. Overall, these results show that children are capable of accurate conditional inferences under probabilistic instructions at a very early age and that the differentiation between categorical and probabilistic conditional reasoning is clear by at least 9 years of age. These results are globally consistent with dual-process theories but suggest some difficulties for the way that the analytic-heuristic distinction underlying these theories has been conceptualized.

The capacity to make deductive inferences from premises that are assumed to be true is one of the more important abilities characterizing the human cognitive system. A great deal of research has examined the kinds of inferences that are made under standard deductive instructions, which require a dichotomous evaluation of the certainty of a putative conclusion. Studies in which both children and adults have been looked at (e.g., Cummins, 1995; Janveau-Brennan & Markovits, 1999; Thompson, 1994, 2000) have consistently shown a great deal of variation, due to both content and context, in this kind of reasoning. At least two kinds of processes have been posited to explain the observed variation. The first relies on explicit use of counterexamples to judge conclusions as certain or not (Johnson-Laird & Byrne, 1991; Markovits & Barrouillet, 2002; Verschueren, Schaeken, & d'Ydewalle, 2005). Such models assume that retrieval and use of at least one potential counterexample are sufficient to deny a conclusion. The second entails a more probabilistic approach (e.g., Evans, Handley, & Over, 2003; Oaksford, Chater, & Larkin, 2000), which requires an evaluation of the probability of a putative conclusion. This involves some process by which the relative frequency of a potential conclusion is compared with the relative frequency of potential counterexamples. The goal of the present article is to examine the

development of these two types of inferential reasoning in young children. Our basic hypothesis is that inferences using judgments of probability emerge early in childhood; explicit inferential reasoning based on explicit use of counterexamples, however, is posited to develop later.

Studies with adults support the hypothesis that when inferences are made on the basis of conditional statements that are familiar to reasoners, counterexample-based inferential reasoning is more demanding than inferential reasoning based on the assessment of probabilities. Verschueren et al. (2005) recently suggested that there are two distinct forms of inferential reasoning, corresponding to two ways of determining whether a putative conclusion is valid. The first relies on likelihood judgments about conclusion plausibility that are fairly intuitive and do not require many cognitive resources. Thus, if the premises suggest that a conclusion is highly likely when compared with potential alternatives that are consistent with the premises (i.e., counterexamples), this conclusion will be judged to be valid. The second form of reasoning requires explicit use of counterexamples in order to determine whether a conclusion is valid or not. A conclusion will be judged to be valid if a reasoner cannot generate at least one counterexample that is consistent with the premises. Verschueren et al. presented evidence that reasoners with higher working memory capacity show a greater de-

H. Markovits, markovits.henry@uqam.ca

gree of explicit counterexample use while making deductive judgments, whereas low working memory capacity reasoners tend to use likelihood judgments about conclusion plausibility. In other words, a deliberate, analytic strategy based on counterexample use, so that a single counterexample was explicitly considered sufficient to invalidate a conclusion, was associated with high working memory capacity, whereas reliance on a presumably more automatic, heuristic process that generated the likelihood of a conclusion was associated with low working memory capacity. The major goal of this study was to test the hypothesis that children also exhibit two separate modes of reasoning that parallel the ones observed in adults, in the context of conditional reasoning, and that these two modes of reasoning show different developmental trajectories. Specifically, the less-resource-demanding probabilistic form of reasoning is hypothesized to develop earlier than resource-intensive counterexample reasoning.

Conditional reasoning is a cornerstone of thinking and is probably one of the most extensively studied forms of logical reasoning, in both the adult and the developmental literatures. Conditional reasoning involves making inferences on the basis of a given *if-then* relation. Most studies of conditional reasoning have examined children's abilities to make inferences on the four basic argument forms. Modus ponens (MP) is the logical 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 forms, since both of them lead to a single, logically necessary 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*. For the latter forms, there is more than one conclusion that is logically consistent with the premises; we will subsequently refer to these as uncertain argument forms. In these cases, the logically correct response is to deny a putative conclusion.

Developmental studies of conditional reasoning show some very clear patterns that can be summarized in the following way: Young children tend to make the MP and MT inferences very reliably when reasoning with familiar premises (Markovits, 2000; Markovits et al., 1996; but see Barrouillet & Lecas, 1999, for evidence that the MT inference is not made when reasoning with premises for which the relation between the terms is arbitrary), but there is a well-documented tendency for them to also accept the invited AC and DA inferences (e.g., in the case of the AC inference, reasoners will often claim that *P implies Q, Q is true* allows the conclusion that *P is true*). Developmental change is characterized mainly by a decrease in the rate of accepting the AC and DA inferences (e.g., Janveau-Brennan & Markovits, 1999; Markovits & Vachon, 1990; O'Brien & Overton, 1980). There are some less clear-cut patterns that suggest possible developmental trends toward refusing the MP and MT inferences, at least in some cases (e.g., Janveau-Brennan & Markovits, 1999; O'Brien & Overton, 1980; Romain, Connell, & Braine, 1983), but these are inconsistent and more difficult to interpret. The

clearest index of developmental change in conditional reasoning is thus the increasing tendency to reject the invited inferences in the AC and DA forms.

In the following study, we will specifically focus on the AC inference. Although the same general patterns are found for the DA inference, the use of a negation in the formulation of this inference makes its use in young children more difficult, particularly in the present context, and we have not examined it here.

A great many empirical studies have examined this kind of reasoning and have made possible some clear understanding of the factors that influence the kinds of AC inferences that are made by both children and adults under standard deductive instructions. These studies have focused on the fact that reasoners apply background knowledge about the conditional relationship to the evaluation of inferences. Thompson (1994, 2000) suggested that the dimension most relevant to evaluating the AC and DA inferences is the extent to which the antecedent of the conditional is perceived to be a necessary condition for the consequent. Specifically, when reasoners believe that P must happen in order for Q to happen (e.g., *if a figure has four sides, it is a square*), correct responses to the AC and DA inferences are low; in contrast, when P is not believed to be a necessary condition for Q (e.g., *if a figure has four sides, it is a polygon*), the rate of correct responses is high.

In fact, there is a great deal of evidence that when making an inference with concrete content, both children and adults may attempt to access their knowledge about the premises; the nature of the retrieved information affects the kind of inferences they will then make. Several studies have shown that the relative number of alternative antecedents that can be readily accessed for any given *if-then* premise with familiar content will affect the kinds of inferences that are made to the AC (and DA) forms (Bucci, 1978; Cummins, 1995; Cummins, Lubart, Alksnis, & Rist, 1991; Janveau-Brennan & Markovits, 1999; Markovits & Vachon, 1990; Thompson, 1994, 2000). For example, take the following AC inference: *If a brick is thrown at a window, then the window will break. A window is broken. Has a brick been thrown at the window?* This inference will lead to a relatively high rate of rejecting the conclusion that a brick has been thrown at the window, since there are many ways of breaking a window other than throwing a brick at it. In comparison, the AC inference *If a tree is cut down, then the tree will fall. A tree has fallen. Has the tree been cut down?* will lead to a relatively low rate of rejecting the conclusion that the tree has been cut down, since there are few ways of making a tree fall other than cutting it down (Bucci, 1978; Cummins, 1995; Markovits & Vachon, 1990). Similarly, the strength of association between the consequent of the major premise (Q) and the most strongly associated alternative antecedent (Barrouillet, Markovits, & Quinn, 2002; Markovits, Fleury, Quinn, & Venet, 1998; Quinn & Markovits, 1998) also affects responses to the AC inference. That is, reasoners are less likely to draw the AC inference when there is a strongly associated, as opposed to a weakly associated, antecedent available.

This idea is reinforced by studies that have shown a correlation between efficiency of information retrieval (as mea-

sured by the quantity and speed of retrieval of information from long-term memory) and responding to the AC (and DA) forms in both children (Janveau-Brennan & Markovits, 1999) and adults (Markovits & Quinn, 2002) and also have shown differential latencies for AC inferences made with premises of differing degrees of associative strength (Grosset, Barrouillet, & Markovits, 2005). Recent studies in which patterns of responses to other inferential forms have been looked at have come to similar conclusions (e.g., De Neys, Schaeken, & d'Ydewalle, 2002).

There are at least two types of processes that can explain these patterns. One entails a process of retrieval and the use of counterexamples, such that a single counterexample is sufficient to invalidate a putative conclusion. One recent instantiation of such a process was presented by Markovits and Barrouillet (2002), who proposed that if at least a single alternative antecedent is retrieved and can be incorporated into an explicit mental model of the premises, the AC (and DA) inferences will be denied.

A second type of process involves a less cognitively demanding assessment of likelihood, as suggested by Verschueren et al. (2005) and by Thompson (2000). Although some forms of probabilistic reasoning certainly involve complex, working-memory-intensive processes, this hypothesis suggests that making probabilistic inferences on the basis of familiar conditional premises is often the result of a likelihood judgment that capitalizes on a reasoner's intuitive knowledge. Note that there are other models of probabilistic reasoning (e.g., Evans et al., 2003; Oaksford et al., 2000) that differ in their details but retain the same basic notion. Specifically, a conclusion will be evaluated as more or less valid if the likelihood that the conclusion is true, assuming the truth of some part of the premises (depending on the specific model), is high. This process is generally conceived of as a fairly effortless evaluation that relies on the associative structure of stored knowledge about the premises and the conclusion. In other words, when evaluating the likelihood of a conclusion, a reasoner can simply activate a network of consequences related to the premises, including potential counterexamples, and estimate the relative weights of the conclusion, as compared with the counterexamples.

At this point, it is useful to examine this distinction in the context of a specific example. Consider the following premises: *If a rock is thrown at a window, then the window will break. Suppose that a window is broken.* Suppose that a reasoner is asked to judge the validity of the putative conclusion *a rock was thrown at the window.* According to Markovits and Barrouillet's (2002) model, counterexample-based reasoning involves two processes. First, the reasoner must retrieve one or more alternative antecedents—that is, *things that can break a window that are not rocks.* Then the reasoner must add an explicit representation of this alternative to the initial representation of the major premise. If such a representation is indeed generated, the conclusion will be judged to be invalid. This corresponds to reasoning of the sort, *it is not certain that a rock was thrown at the window, since maybe a storm broke the window.* If such a representation is not generated, the conclusion will be judged to be valid. We will

refer to these types of inferences as *categorical*, since they result in a binary response.

However, a reasoner might also use a probabilistic approach to evaluating the inference. A probabilistic inference will require a reasoner to estimate what the probability would be that a rock was thrown at the window. This would require some evaluation of the relative frequency of windows being broken by rocks—that is, the conditional probability that if a window is broken, a rock was thrown at it—leading to reasoning of the sort, “it's fairly likely that if the window is broken, a rock was thrown.” Available evidence indicates that this generally involves a rapid judgment based on simple activation of *things related to breaking windows* and an intuitive estimation of the relative weights of rocks and other things. It is this form of judgment that is assumed to underlie current models of probabilistic reasoning, and frequent use of such a process is implied by the results of Verschueren et al. (2005). We will refer to this as *probabilistic* reasoning in the following.

The dissociation between categorical inferences based on a search for, and the explicit representation of, counterexamples and probabilistic inferences that are based on a more intuitive assessment of likelihood is consistent with the broader outlines of dual-process theories (Verschueren et al., 2005). Although there are differences between various formulations (e.g., Evans, 2007; Evans & Over, 1996; Klaczynski, 2001; Stanovich & West, 2000), these theories agree that making inferences can be guided by very direct associative processes that exploit people's experience and knowledge of the world in order to allow them to make inferences rapidly and automatically. In contrast to this, a second, analytic system is postulated that requires conscious effort and an explicit use of information, such as that required for counterexample generation and use. The latter kind of processing may ultimately allow the making of inferences that are abstract and decontextualized; however, in the present study, we were more concerned with the development of the underlying processes.

Common to these theories is the idea that associative processes are present at a very early age, whereas the more explicit, and working-memory-intensive processes develop later. In other words, if dual-process theories are true, young children should find it easier, when reasoning with familiar premises, to make probabilistic inferences, using some form of likelihood evaluation, than to make explicit categorical inferences based on counterexample generation and use. There are, in fact, some recent results that are consistent with the idea that cognitive development in adolescence can be understood by the interplay between an early-developing associative system and a later developing analytic system (e.g., Klaczynski, 2001; Kokis, Macpherson, Toplak, West, & Stanovich, 2002). More specifically, developmental studies in which children's conditional reasoning with familiar premises under standard deductive instructions has been looked at have suggested that children's ability to use an explicit counterexample to deny the AC inference is very limited at 6 or 7 years of age and undergoes rapid change subsequently (Janveau-Brennan & Markovits, 1999; Markovits, 2000). Although there are no known direct studies of probabilis-

tic conditional inferences, there is evidence that children as young as 6 years of age understand the basic notion of probabilistic uncertainty with familiar content (e.g., Acredolo, O'Connor, Banks, & Horobin, 1989; Brainerd, 1981). In fact, even infants have basic probabilistic intuitions (Téglás, Girotto, Gonzalez, & Bonatti, 2007). A recent study by Girotto and Gonzalez (2008) has shown that children this young can also understand the more complex notion of posterior probability, which involves updating judgments in uncertain situations on the basis of new evidence, an ability that must be a key component in the making of probabilistic inferences. These results indicate that very young children (around 6 years of age) possess the kind of understanding of probabilities that is required to understand the task demands involved in making a probabilistic conditional inference. There is, however, no study that we know of that has directly tested the hypothesis that 6-year-old children can adequately make probabilistic conditional inferences and later develop inferential strategies based on the deliberate use of counterexamples. This was the purpose of the following studies.

STUDY 1

In this first study, we decided to look specifically at the way that young children make conditional inferences with simple verbal propositions, corresponding to the MP and AC forms on two different tasks. The first used standard deductive instructions and requested categorical inferences (we will refer to these as *categorical* instructions). These instructions were chosen to promote use of counterexample-based reasoning. We hypothesized that younger children would show relatively more difficulty on this task than would older children and, conversely, that older children would be more likely to make explicit use of counterexamples than would younger children on this task. The second task asked children to assess the degree to which a putative conclusion is likely (these will be referred to as *probabilistic* instructions). This was designed to elicit probabilistic inferences. Under the hypothesis that probabilistic reasoning is relatively well developed in young children, when compared with counterexample-based categorical reasoning, we expected relatively little developmental change on the probabilistic task, as compared with developmental change with the categorical inferences. Finally, under the hypothesis that counterexample-based reasoning develops later than probabilistic reasoning, we should observe that younger children tend to treat the probabilistic and categorical problems similarly; however, as children grow older, they should differentiate more clearly between the two.

To test these hypotheses, we examined the inferences that the children drew for MP and AC problems, and we also asked them to provide justifications for those inferences. There is very good evidence that young children's justifications are good indicators of the kinds of information that are used to make an inference (Janveau-Brennan & Markovits, 1999; Markovits, 2000). By asking for justifications in addition to inferential judgments, we could then examine patterns of variation in both categorical and

probabilistic inferences as a function of information use. The specific premises used here were chosen to provide relatively easy access to potential alternative antecedents for children of the age examined (Markovits et al., 1998).

We can then make some specific predictions about children's inferential performance. First, we know from developmental studies that even young children can respond correctly to the MP inference under standard deductive instructions with the kinds of plausible premises used in this study (Markovits, 2000; Markovits et al., 1996). Adults appear to reason similarly on MP with both standard deductive and probabilistic instructions, and there was no reason to suppose that probabilistic evaluation of MP inferences in children would be difficult. Thus, we predicted that children of all the age groups should accept the MP inference under both probabilistic and deductive instructions.

It should be noted that when premises allow potential retrieval of disabling conditions, performance on the MP inference is more difficult (these are conditions that allow the antecedent to be true but suggest that the consequent may not be true—e.g., *if a rock is thrown at a window, the window will certainly break, unless the window is made of Plexiglas*; see Cummins et al., 1991). Available evidence suggests that retrieval of disabling conditions during reasoning will lead to rejection of the MP inference (e.g., Cummins et al., 1991), unless the reasoner manages to inhibit these (De Neys, Schaeken, & d'Ydewalle, 2005b; Markovits & Doyon, 2004). In the present context, we did not wish to examine the effects of disabling conditions and, thus, chose premises for which few disabling conditions were available. We thus expected that all the children in the present study would show high levels of acceptance of the MP inference under categorical instructions and high ratings of the MP inference under probabilistic instructions, irrespective of the underlying processes.

In the case of the AC inference, however, probabilistic and categorical instructions should produce different patterns. Although younger children may not show a clear distinction between the forms of reasoning used with the two instructions, as reasoning by counterexample develops, the two forms should diverge. In particular, the reasoning-by-counterexample model supposes that production of at least one alternative antecedent (of any kind) should be strongly correlated with rejection of the AC inference. In contrast, when probabilistic inference is made, the relation between alternatives and ratings of the likelihood of a putative conclusion is weaker (see Markovits & Handley, 2005, for evidence of this difference in adults).

A simple example of this is a situation in which the alternative antecedent that is retrieved is one whose real probability of occurrence is low; in this situation, the likelihood of the invited conclusion will remain high. For instance, suppose that a reasoner is asked to evaluate the following AC inference: *If an animal is a duck, then it has wings. An animal has wings. It is a duck.* Suppose that Reasoner A retrieves the example of a sparrow, as an alternative counterexample, whereas Reasoner B does not do so. If these reasoners are given standard deductive instructions and are capable of reasoning by counterexample, Reasoner A would deny the invited conclusion *it*

is a duck, and Reasoner B would accept this conclusion. If these two reasoners are given probabilistic instructions, they would have to evaluate the probability that something that has wings is a duck. Reasoner B would rate this as quite low, given general knowledge about things having wings. Reasoner A would add to general knowledge the specific example of a sparrow. However, this would not appreciably change A's initial evaluation, since the probability that something having wings is a duck is low even when sparrows are considered. Thus, we would predict that the relation between alternative antecedent use and inferential performance would be strong with categorical instructions and weak with probabilistic instructions, if the processes used to make these two forms of inference were clearly differentiated, as is the case with adults.

In sum, we hypothesized that counterexample-based categorical reasoning develops later than probabilistic reasoning. Thus, we expected a stronger developmental trend in the tendency to reject AC inferences under categorical instructions than under probabilistic instructions. We expected older children to generate more counterexamples than would younger children. We expected the relation between counterexample generation and AC reasoning to be stronger under categorical instructions than under probabilistic instructions among the older children, but not among the younger children. Moreover, whereas older children should be better than younger ones at using counterexamples to make categorical inferences, this trend was not expected for probabilistic inferences.

Method

Participants. A total of 56 elementary school children participated in the study. Of these, 17 were in Level 1 (average age: 6 years 5 months; 2 girls, 15 boys), 13 were in Level 2 (average age: 7 years 1 month; 5 girls, 8 boys), and 26 were in Level 4 (average age: 8 years 7 months; 13 girls, 13 boys). The children attended various rural state schools in southwest England.

Materials. A set of four videotaped sequences were prepared. The first showed a cartoon character with a black background. The character greeted the child (although not by name!) and then said that he would have some questions to ask the child. Then the following instructions were presented for the categorical condition:

I will first read to you a short sentence which you must suppose to be true. Then I will ask you to use the first sentence to decide whether or not you can say that something must be certainly true or not.

The child was then presented with an example of an MP inference, using the premise *If an animal is a duck, then it has wings*. After this, the following series of inferences were presented:

Inference 1 (MP): Suppose that it is true that: If something is a car, then it has a motor. Now, suppose that you see a car. Is it certain that it has a motor?

Inference 2 (AC): Now, suppose that you see something that has a motor. Is it certain that it is a car?

Inference 3 (MP): Suppose that it is true that: If an animal is a hedgehog then it has four legs. Now, suppose that you see an animal that is a hedgehog. Is it certain that it has four legs?

Inference 4 (AC): Now, suppose that you see an animal with four legs. Is it certain that it is a hedgehog?

For each of these inferences, the children were asked to justify their responses by answering the question, "Why do you think this?"

In all cases, the MP inference preceded the AC inference for the same premise. The MP inference was given first in order to allow the participants to establish the basic premise clearly before being given the more complex AC inference. It should also be noted that there is no evidence of the existence of order effects between the MP and the AC inferences. Finally, pretesting showed that even 5-year-old children were able to reliably generate counterexamples required by the two AC inferences (i.e., animals who are not hedgehogs that have four legs, things that are not cars that have motors) when explicitly asked to do so.

Following this first set of questions, the participants were shown the certainty scale by the experimenter. This consisted of a sequence of five vertical red bars on a white sheet of cardboard, with the bar on the left being close to 0 cm. Each successive bar was longer than the one to the left of it, making an ascending sequence. Each child was told that the extreme points represented *totally likely* and *totally unlikely* possibilities, with intermediate points representing variable degrees of likelihood. The exact formulation used the notion of "how sure it is." The points on the scale were indicated by the following sequence: (1) *not sure*, (2) *a little sure*, (3) *medium sure*, (4) *very sure*, (5) *very, very sure*. These categories were converted to a 5-point scale for the purpose of analysis, with 1 used to denote *not sure* and 5 used to denote *very, very sure*.

The children were then shown three situations, using a set of green and red trees. In the first, there was one red tree and nine green trees; in the second, five red and five green trees; and in the third, nine red and one green tree. For each of these, the participants were asked to show on the scale the likelihood of choosing a red tree from among all the trees. Specifically, they were asked, "How sure is it that if you chose 1 tree from among all these trees, the tree would be red?" If the child had problems understanding how to use the scale appropriately, the experimenter repeated the explanation. Most children had no difficulty in understanding how to use the scale in these conditions. The exact formulation used was derived after pretesting the children in the youngest age group examined here.

Following this explanation, the children were asked the same inferential questions as in the categorical condition, in the same order. This time, however, they were asked how likely the presented conclusion was. For example, the MP inference was presented in the following way:

Suppose that it is true that: If something is a car, then it has a motor. Now, suppose that you see a car. How sure is it that it has a motor?

A second version of this videotaped sequence was constructed by starting with Inferences 3 and 4 and then presenting Inferences 1 and 2. Finally, for each of these two sequences, a second set was constructed by asking the probabilistic inferences first, followed by the categorical inferences, thus giving a total of four sequences.

Procedure. The children were tested individually in sessions lasting about 10 min. The children were introduced to an experimenter, who told them that they would watch a single videotape and that, during the tape, they would be asked some questions. They were asked to listen to the tape attentively. After that, the experimenter started the tape. The tape was paused after each question. The children's responses were recorded and analyzed subsequently.

The four videotaped sequences were varied systematically from one child to the next.

Results

We first examined the mean responses to inferences under the categorical and probabilistic instructions. Since an initial analysis showed that there were no order effects, subsequent analyses did not take this factor into consideration. We also compared AC performance with the car and hedgehog premises. No significant differences were found between these two AC inferences, under either cat-

egorical or probabilistic instructions, so we combined scores across these two contents.

Acceptance of the MP and AC inferences. The children’s response was “yes,” “no,” or “not sure.” The first was classed as accepting the inference, whereas “no” and “not sure” responses were classed as not accepting the inference (Markovits et al., 1998). Table 1 gives the mean percentages of acceptance of the MP and AC inferences under categorical instructions (out of two) and the mean probability ratings for the MP and AC inferences (on a scale from 1 to 5) under probabilistic instructions. As can be seen from this table, performance on the MP inference showed very high levels of acceptance under categorical instructions and correspondingly high levels of the evaluated probability of this inference. This indicated that even the youngest children had no real difficulty in understanding the instructions to suppose that the premise was true, with both forms of instruction.

We initially examined the way that inferential performance under probabilistic instructions varied as a function of age. An ANOVA with mean conclusion ratings on the MP and AC inferences as dependent variables, with argument form as a repeated measure, and with grade level as the independent variable was performed. This showed only an effect of argument form [$F(1,53) = 32.27, p < .001$]. No other effects were significant. Mean ratings on the AC inference were lower than mean ratings on the MP inference at all grade levels. This shows that the participants were generally able to differentiate the AC from the MP inference in the probabilistic condition. More important, consistent with our hypothesis, there was little developmental change in how either the AC or the MP inferences were evaluated under probabilistic instructions.

We also examined how the probabilistic scale was used. At each grade level, we calculated the proportion of responses that were at either extreme of the probability scale. These represented 80%, 62%, and 50% of all the responses among the 6-year-olds, the 7-year-olds, and the 9-year-olds, respectively. This shows that the 6-year-olds did sometimes employ the intermediate points on the probability scale but had, nonetheless, a strong tendency to dichotomize their responses. Use of an intermediate response clearly increased with age.

We subsequently looked at inferential performance under categorical instructions. We first looked at whether these children were able to distinguish between the MP and the AC inferences. In order to do this, we calculated

the number of times that each participant accepted the MP and AC inferences (this gave scores varying between 0 and 2 for each inference type). We then compared these, using a Wilcoxon signed-rank test for each grade level. This showed that acceptance of the MP inference was significantly higher than acceptance of the AC inference for the 6-year-olds [$T(17) = 12, p < .05$], the 7-year-olds [$T(13) = 39, p < .001$], and the 9-year-olds [$T(26) = 95, p < .001$]. We then looked at the change in performance on the two forms of inference as a function of age. We performed a log-linear analysis on the number of MP acceptances, with grade as a dependent variable. This showed no effect of grade. We then performed a log-linear analysis on the number of AC acceptances, with grade as a dependent variable. This showed a significant effect of grade [$\chi^2(2) = 9.78, p < .01$]. There was a significant decrease in AC acceptance between the 6-year-olds and the 7-year-olds [$\chi^2(1) = 10.18, p < .01$], with no difference between the 7-year-olds and the 9-year-olds [$\chi^2(1) = 1.70, n.s.$].

In sum, consistent with our predictions, acceptance of the AC inference under deductive instructions declined with increasing age. No other developmental trends were noted. Thus, although there was a very strong decrease in the percentage of AC acceptances under categorical instructions between 6 and 7 years of age (from 61.8% to 11.5%), this was not mirrored under probabilistic instructions, for which there was no significant difference between these same ages (from 3.57 to 3.38). Moreover, a basic acceptance of the premise, as shown by the high rates of acceptance of the MP inferences with categorical instructions and the high rating of the MP inference with probabilistic instructions, was evident at all age levels. Note that although the relatively small number of participants precludes any absolute judgment as to the developmental pattern underlying probabilistic inferences, we can nonetheless conclude that the change in performance under categorical instructions, between 6 and 7 years of age, was clearly stronger than that observed with probabilistic inferences for the same premises.

Relation between justifications and inferential performance. A second set of analyses then examined the relation between the children’s explicit justifications and their inferential performance. This was done in order to test our prediction that the relation between counterexample generation and inferential performance would be more closely tied to categorical than to probabilistic inferences among the older children.

We started by examining developmental patterns in the children’s justifications for the AC inferences. We were specifically interested in justifications that referred to potential alternative antecedents—either those that stated a specific alternative (e.g., *because cats and dogs have legs*) or those that stated a general alternative (e.g., *other things have legs*). These categories have been used previously (e.g., Janveau-Brennan & Markovits, 1999) and are very clearly defined. Justifications were coded separately by two experimenters. There were disagreements on fewer than 5% of the codes, and these were resolved by subsequent discussion between the two coders.

Table 2 gives the percentage of justifications that referred to alternative antecedents for the categorical and

Table 1
Mean Percentages of Acceptance of the Invited Inference in the Categorical Condition and Mean Ratings of the Probability of the Invited Inference (on a Scale From 1 to 5) in the Probabilistic Condition for the Modus Parens (MP) and Affirmation of the Consequent (AC) Forms Combined Over the Two Premises

Grade Level	n	Categorical Inference		Probabilistic Inference	
		MP	AC	MP	AC
1	17	88.2	61.8	4.56	3.57
2	13	88.5	11.5	4.50	3.38
4	26	84.6	25.0	4.56	2.72

Table 2
Mean Percentages of All Justifications That Referred to
Alternatives on the Affirmation of the Consequent Inferences by
Grade Level for Categorical and Probabilistic Inferences

Grade Level	<i>n</i>	Categorical	Probabilistic
1	17	38.2	35.3
2	13	80.8	50.0
4	26	76.9	51.9

probabilistic inferences as a function of grade level. We first looked at whether the number of justifications referring to alternatives differed between the two types of inferential problems. We performed an ANOVA for the number of justifications referring to alternatives on the categorical inferences and on the probabilistic inferences, with problem type as a repeated variable and grade level as the independent variable. This analysis was repeated using rank order and gave the same pattern of results (see, e.g., Conover, 1980). This showed a significant effect of grade [$F(1,53) = 5.99, p < .01$] and of problem type [$F(1,53) = 11.49, p < .001$]. More justifications referring to alternatives were produced under categorical instructions than under probabilistic instructions. Also, as was predicted, there was an overall increase in the relative production of such justifications with age, with contrast analyses indicating a significant increase between 6 and 7 years of age only. These data provide support for our hypothesis that counterexample reasoning develops with age.

Just as important, the data in Table 3 confirm that the use of counterexamples develops differently for categorical and probabilistic inferences. Table 3 shows the percentage of acceptances of the AC inference under categorical instructions and the mean ratings for AC inferences under probabilistic instructions as a function of production of an alternative antecedent as a justification. An examination of Table 3 shows that the percentage of acceptances of the AC inference under categorical instructions when an explicit alternative antecedent was produced as a justification was less than the percentage of acceptances when an explicit alternative antecedent was not produced, at all age levels. Under probabilistic instructions, the mean AC rating was greater when an alternative antecedent was not produced than when it was produced, for the two younger age levels. However, for the oldest participants, this pattern was reversed.

We corroborated these conclusions by calculating the correlations between the rate of production of an alterna-

tive antecedent and responses to the AC inferences for both the categorical and the probabilistic conditions. In order to do this, we calculated the number of justifications to the two AC problems that involved at least one alternative antecedent (out of a total of two), separately for the categorical and the probabilistic conditions. We then calculated Pearson correlations between justifications with alternative antecedents and responses to the AC inferences for the categorical and the probabilistic instructions. Correlations between justifications and number of acceptances of the AC inference under categorical instructions were significant for the 6-year-olds [$r(17) = -.69, p < .01$], the 7-year-olds [$r(13) = -.61, p < .05$], and the 9-year-olds [$r(26) = -.75, p < .001$]. The children who produced more alternative antecedents as justifications rejected the AC inference more often under categorical instructions, at all ages. The correlation between justifications and mean ratings of the AC inferences under probabilistic instructions was significant for the 6-year-olds [$r(17) = -.56, p < .05$] and for the 7-year-olds [$r(13) = -.59, p < .05$], but not at all for the older children [$r(26) = -.01, n.s.$]. In other words, younger children who produced more alternative antecedents gave lower ratings to the AC inferences under probabilistic instructions, whereas older children showed no relation between justifications and ratings. For the 9-year-olds, correlations between justifications and AC inferences were significantly higher under categorical instructions than under probabilistic instructions [$t(25) = 5.35, p < .001$].

Discussion

The results of this study provided a focused look at the developmental patterns and relations between simple conditional reasoning under *deductive* instructions and under probabilistic instructions. The children of all ages and in both instruction conditions accepted the MP inference at a high rate. Moreover, even the 6-year-olds accepted the AC inferences less often than the MP inferences under categorical instructions and evaluated the AC inference as less likely than the MP inference under probabilistic instructions. More important, unlike for the MP inference, there were changes in the tendency to reject the AC inference as a function of age and differences in the development trends for categorical and probabilistic inferences.

Specifically, there was a strong tendency for the 6-year-old children to accept the AC inference under categorical instructions, with acceptance rates around 60%, even

Table 3
Mean Percentages of Acceptance of the Invited Affirmation of the Consequent Inference in the
Categorical Condition and Mean Probability Ratings (on a Scale From 1 to 5) in the Probabilistic
Condition As a Function of Production or Lack of Production of an Alternative Antecedent on
Justifications Generated for the Corresponding Inferences

Grade Level	Categorical Condition				Probabilistic Condition			
	Alternative Produced		No Alternative		Alternative Produced		No Alternative	
	<i>n</i>	% Acceptance	<i>n</i>	% Acceptance	<i>n</i>	Mean Rating	<i>n</i>	Mean Rating
1	13	30.8	21	81.1	12	2.90	22	3.85
2	21	5.8	5	40.0	13	3.00	13	3.77
4	40	10.0	12	75.0	25	2.96	27	2.54

with premises that allowed very easy access to potential alternative antecedents. There was a very steep drop in acceptances of the AC inference between 6 and 7 years of age (from 62% to 12%). This was accompanied by similar patterns of change between these two age levels involving both increasing production of explicit alternative antecedents in both conditions and an increased use of intermediate scale points in the probabilistic condition. These different factors indicate that 6 years of age may be the lower limit for the ability to differentiate AC and MP inferences with the very simple kinds of verbal premises used in this study.

What do these results indicate about the reasoning processes used under categorical and probabilistic instructions? First, these two forms of reasoning have different developmental patterns. *Categorical* reasoning showed a steep increase in AC refusals between 6 and 7 years of age. There was no such effect with probabilistic reasoning, which overall showed a much flatter pattern of developmental change. Second, the correlation between responses to the AC inferences and numbers of justifications referring to alternatives was similarly high among the 6- and 7-year-old children under both categorical and probabilistic instructions. However, among the 9-year-olds, there was no correlation between justifications and ratings of the AC inference under probabilistic instructions, whereas the same correlation was very high under categorical instructions. This is exactly the pattern that is found in adults (Markovits & Handley, 2005) and is also consistent with Verschueren et al.'s (2005) results.

These results clearly are consistent with the hypothesis that children as young as 9 years of age have access to two distinct modes of inferential reasoning, one of which corresponds to a reasoning-by-counterexample model, whereas the other corresponds to a form of probabilistic inference. In contrast, the 6-year-olds showed difficulty in using an explicit process of reasoning by counterexample, since they were less likely to generate counterexamples than were their older counterparts. In contrast, when a counterexample was produced, likelihood ratings of the AC inference were relatively constant across grades, suggesting that even the youngest children had access to a probabilistic form of inference, which showed a weaker pattern of developmental change in the age range between 6 and 9 years than was shown by reasoning by counterexample.

STUDY 2

The results of Study 1 suggested that even 6-year-old children can make adequate conditional inferences under probabilistic instructions with simple verbal premises, whereas basic counterexample-based reasoning is still difficult at this age—specifically, because 6-year-olds are less likely to retrieve and use a counterexample under categorical instructions, as compared with older children. Retrieval, when not automatic, uses working memory capacity (Rosen & Engle, 1997) and presents an additional level of difficulty for very young children. Thus, we hypothesized that very young children's counterexample-based

reasoning would improve if they were given problems for which counterexamples were directly accessible.

We thus decided to replicate the initial study by constructing inferential problems for which the universe of alternatives was directly visible. In order to do this, we presented children with a series of red and blue shapes. In one situation, the red shapes consisted of four squares and one circle, and the premise referred to the fact that all the squares were red. In this case, since the probability of a red shape's being a square was .8, the AC inference was highly probable. In the second situation, the red shapes consisted of one triangle and four circles, and the premise referred to the fact that all the triangles were red. In this case, since the probability of a red shape's being a triangle was .2, the AC inference was highly unlikely.

We made two specific hypotheses. First, the results of the previous study had indicated that even 6-year-olds were able to make adequate conditional inferences under probabilistic instructions with purely verbal premises with very familiar content. Thus, we anticipated that the 6-year-olds in the present study would be able to distinguish between the high- and low-probability AC inferences under probabilistic instructions. The results of the previous study had also shown that whereas 7-year-olds were able to reject the AC inference under categorical instructions at a very high level, 6-year-olds were not able to do so. We predicted that providing 6-year-old children with these kinds of problems would facilitate counterexample-based reasoning and would allow them to reject the AC inference under categorical instructions at a higher level than had been observed in the initial study. Finally, given the way that the stimuli were presented, we did not ask the children for justifications in this case but simply examined responses to the inferential questions.

Method

Participants. A total of 53 elementary school children participated in the study. Of these, 21 were in Level 1 (average age: 6 years 5 months; 11 girls, 10 boys), 18 were in Level 2 (average age: 7 years 7 months; 8 girls, 10 boys), and 14 were in Level 4 (average age: 9 years 8 months; 8 girls, 6 boys). The children attended various rural state schools in southwest England.

Materials. A set of four videotaped sequences were prepared that were identical to those used in the first study, with one difference. When the inferential problems were presented, a set of colored shapes appeared on the screen. For the first set of inferences, these were four red squares, one red circle, and four blue circles, in that order (we will refer to this as the *high-probability* problem). The corresponding inferential problems in the categorical condition referred to these shapes and were

Inference 1 (MP): It is true that: If a shape is a square, it is red. Now, suppose that a shape is a square. Is it certain that it is red?

Inference 2 (AC): Now, suppose that a shape is red. Is it certain that it is a square?

The shapes used in the next two problems were one red triangle, four red circles, and four blue circles, in that order (we will refer to this as the *low-probability* problem). The corresponding inferential problems in the categorical condition referred to these shapes and were

Inference 3 (MP): It is true that: If a shape is a triangle, it is red. Now, suppose that a shape is a triangle. Is it certain that it is red?

Inference 4 (AC): Now, suppose that a shape is red. Is it certain that it is a triangle?

Probabilistic versions of these problems used the same scale and wording as those in the initial study (e.g., *It is true that: If a shape is a triangle, it is red. Now, suppose that a shape is a triangle. How sure is it that it is a triangle?*). Note that the *if-then* relations in Study 1 were presented by the formulation *suppose that it is true*, whereas in the present study, the formulation *it is true* was used. This was done simply because it was felt that using a suppositional term when all of the problem stimuli were, in fact, directly visible, might lead the children to suspect that part of the problem was not actually presented.

Procedure. The children were examined individually in sessions lasting about 10 min. The children were introduced to an experimenter, who told them that they would watch a single videotape and that, during the tape, they would be asked some questions. They were asked to listen to the tape attentively. After that, the experimenter started the tape. The tape was paused after each question. The children's responses were recorded and analyzed subsequently.

The four videotaped sequences were varied systematically from 1 participant to the next.

Results

Since all the participants consistently accepted the MP inference under categorical instructions at a very high rate and consistently rated these inferences very highly under probabilistic instructions, we concentrated our subsequent analyses on the AC inferences.

Table 4 indicates the mean percentage of participants accepting the AC inferences by age and order for the two inferential problems under categorical instructions and the mean ratings given to the AC inferences under probabilistic instructions. The former showed a clear effect of order, which we thus included in subsequent analyses.

We first examined AC performance on the two problems under probabilistic instructions. We ran a repeated measures ANOVA with AC ratings on the high-probability and the low-probability problems as dependent variables and grade and order as independent variables. This showed a significant difference between the two problems [$F(1,47) = 22.64, p < .001$]. No other effect was significant. The mean AC rating for the low-probability problem ($M = 2.56$) was lower than that for the high-probability problem ($M = 3.87$). These results clearly indicate that even the youngest children were able to make probabilistic ratings of the AC inferences that accurately reflected differences in the real characteristics of the problems. Sec-

ond, inferential performance under probabilistic instructions was quite stable and was not influenced by problem order.

We then looked at the degree to which the children used only the end points of the probability scale. The percentages of judgments that did this were 62%, 36%, and 36% at Levels 1, 2, and 4, respectively. These rates are lower than those observed in the first study (where 80% of Level 1 children used end points) and indicate an overall increase in use of intermediate points on the probabilistic judgments.

We then examined AC performance on the two problems under categorical instructions. Given the dichotomous nature of the scores in this condition and the complex pattern of results, we analyzed these results in the following way. We first looked at the relative difference between the two problem types globally. We combined scores across grades and used a McNemar test with a correction for continuity comparing percentage of acceptance of the AC inference for each problem. This showed a significant difference between the low-probability problem ($M = 28.3\%$) and the high-probability problem ($M = 49.1\%$) [$\chi^2(1) = 6.67, p < .01$]. Thus, categorical performance reflects the probability manipulation in the same way as probabilistic performance.

We then looked at overall inferential performance by combining the two problems into a single score. We performed a log-linear analysis on the combined score, with grade and order as dependent variables. This showed only an effect of order [$\chi^2(1) = 6.34, p < .01$]. The rate of acceptance of the AC inference was lower when categorical problems were given first ($M = 25.9\%$) than when these problems were preceded by probabilistic problems ($M = 51.9\%$).

Discussion

The aim of this second study was to examine probabilistic and categorical inferences in a situation in which retrieval demands were minimized and there was an observable difference in the real probabilities of AC inferences. There are some clear conclusions that emerge from these results.

The initial study suggested that young children might have a well-developed ability to make probabilistic conditional inferences, on the basis of their ability to distin-

Table 4
Mean Percentages of Acceptance of the Invited Affirmation of the Consequent (AC) Inference Under Categorical Instructions and Mean Ratings of the Invited AC Inference (on a Scale From 1 to 5) Under Probabilistic Instructions for the High-Probability and Low-Probability Problems

Order	Grade Level	n	Categorical Instructions		Probabilistic Instructions	
			High Probability	Low Probability	High Probability	Low Probability
Deductive first	1	10	40.0	30.0	3.20	2.80
	2	10	40.0	20.0	3.90	2.70
	4	7	14.3	0.0	4.00	2.14
Probability first	1	11	72.7	63.6	4.18	3.00
	2	8	37.5	12.5	3.63	2.25
	4	7	85.7	28.6	4.43	2.14

guish between MP and AC inferences. The present study allows the conclusion that even 6-year-old children have a clear ability to make AC inferences that accurately mirror key differences in real conditional probabilities. This reinforces the results of the previous study and strongly suggests that the ability to make accurate conditional inferences under probabilistic instructions is indeed present at a very early age.

The results for children's performance under categorical instructions are more complex and can be looked at in two different ways. When the categorical instructions were given first, the overall rate of acceptance of the AC conclusion for the 6-year-olds was 40.0% for the high-probability problem and 30.0% for the low-probability problem. Thus, a clear majority of 6-year-olds were correctly able to deny the AC inference under categorical instructions when these were given first, something that was not the case for the 6-year-olds in the first study with verbally presented premises. Presenting the inferential problems in the present format did improve performance under categorical instructions for the youngest children, as was hypothesized. These data indicate that part of the difficulty young children experience with categorical reasoning can be attributed to the failure to retrieve counterexamples.

However, despite this relative improvement, categorical performance nonetheless reflected the probability manipulation of the visible stimuli used in these problems, as is shown by the greater rate of acceptance of the AC inference with the high-probability problem, as compared with the low-probability problem. In other words, the children in all the age groups appeared to use probabilistic information to inform a categorical judgment, at least when this information was made explicit. This effect has not been found with adults who were given direct access to alternative antecedents (e.g., Markovits & Handley, 2005), and indicates that although young children's ability to reason with counterexamples might be present at an early age, particularly with the kinds of stimuli used in this study, their reasoning remains influenced by the probabilistic characteristics of the problems.

This influence is most clearly shown by the clear impact that making probabilistic inferences had on subsequent performance under categorical instructions in the present study. When the children started by making a conditional inference under categorical instructions, they rejected the AC inference at a very high rate, with even the 6-year-olds showing an overall rejection rate of 65%. However, when asked to first make conditional inferences under probabilistic instructions, the rejection rate of the AC inference went down significantly. In addition, there was no effect in the opposite direction: Probabilistic inferences were not affected by categorical ones.

On the surface, this finding is counterintuitive. One might anticipate that asking children to make probabilistic inferences initially should focus their attention more clearly on alternative antecedents, which should thus facilitate subsequent rejection of the AC inference under categorical instructions. However, the observed effect is exactly the opposite of what would be predicted by this

analysis. Put more directly, the results of this study indicate that the way in which children process the concrete parameters that allow them to make relatively accurate probabilistic inferences, such as *if a shape is red, then it is relatively improbable that it will be a triangle*, makes it more difficult for them to conclude that it is not certain that *if the shape is red, it is a square*. We will examine this point further on. The most direct interpretation of this result is that once children have been induced to think probabilistically, it is difficult for them to initiate the processes required for successful categorical reasoning. In other words, categorical inferences appear to be somewhat fragile; probabilistic inferences, in contrast, appear to be robust and not influenced by categorical inferences. These data are consistent with our hypothesis that probabilistic reasoning is an earlier developing mode of thinking than is categorical inference.

These data provide additional confirmation that the way that relevant information is processed when children make conditional inferences under probabilistic instructions is different from the way that this information is processed under categorical instructions. Thus, these data are consistent with the results of the first study showing that, at least among the older children, the relationship between explicit counterexample production and inferential performance was significantly different under probabilistic and under deductive instructions.

The results of this study reinforce those of the previous one in several ways. First, the results of both studies strongly suggest that probabilistic reasoning is quite well established as early as 6 years of age. Second, the present results suggest that the basic processes involved in counterexample-based reasoning are clearly observable at 6 years of age, when retrieval demands are limited. Third, they provide strong evidence that early counterexample-based reasoning can be differentiated from probabilistic reasoning but that this distinction remains fragile in young children.

STUDY 3

Finally, we wished to examine one further question that might affect the interpretation of the results of the first two studies. When any kind of inference is made, one must distinguish between the nature of the inference and the reasoner's confidence in his or her judgment. The response scale used for the probabilistic inferences is one that also suggests the degree of certainty that participants might have about a specific response. This, in turn, suggests the possibility that the children in these studies might have been confusing degree of certainty with probability.

If this were the case, it would imply that the children were consistently making counterexample-based inferences but were responding in terms of their confidence when making probabilistic judgments. This is not consistent with the results of the first study, which showed varying patterns of correlations between counterexample generation and reasoning under deductive and probabilistic instructions. This is also inconsistent with the results of the second study, since it is unclear how the observed asymmetric order effect could be explained if children

were using the same basic processes under both deductive and probabilistic instructions.

Nonetheless, this raises an interesting general question, which is the ability of young children to distinguish between a probabilistic judgment and a confidence judgment. Certainly, our previous results allow us to hypothesize that children can indeed make this distinction. We thus decided to address this question directly by asking a separate sample of children both to give a simple probabilistic inference and, at the same time, to express their degree of certainty about their response.

Method

Participants. A total of 26 elementary school children participated in the study. These were all in Grade 2 (average age: 7 years 3 months; 13 girls, 13 boys) in a public elementary school in Brussels, Belgium. All the children were native French speakers.

Materials. Two separate drawings were prepared. The first drawing showed a line of nine squares; the first eight were red, whereas the ninth was blue. The children were asked an initial, high-probability question, requiring them to estimate how probable it would be to choose a red square with their eyes closed. Following their response, the children were also asked to indicate how certain they were about the response. The same 5-point scale was used for both questions. After that, they were asked a second, low-probability question, requiring them to estimate how probable it would be to choose a blue square with their eyes closed. They were then asked to indicate how certain they were about this response.

A second drawing showed a line of eight circles followed by a triangle. The children were asked how probable it would be to choose a circle (high-probability question) and, subsequently, how probable it would be to choose a triangle (low-probability question). In both cases, the children were asked to indicate their degree of certainty about their response.

Procedure. The children were examined individually in sessions lasting about 5 min. The children were introduced to an experimenter, who started by showing the children how the two scales worked. Half of the children received the problem with the red and blue squares first, whereas the other half received the problem with the circles and the triangle first.

Results

Table 5 indicates the mean ratings for both a high- and a low-probability question for both problems and the corresponding degrees of certainty. Inspection of this table shows that although the ratings for the low-probability questions are lower than those for the high-probability questions, there is little variation in the degree of certainty across the four questions. We then ran an ANOVA on problem ratings, with problem type and high or low probability as repeated measures. This indicated only a main effect of high or low probability [$F(1,23) = 7.11, p < .01$]. Ratings for the high-probability questions were higher than those given to low-probability questions. We then performed the

same analysis on the degrees of certainty. No significant effects were found.

In order to examine the difference between ratings and certainty more directly, we conducted paired *t* tests between the four ratings and corresponding levels of certainty. No differences were found for the high-probability questions. However, for the low-probability questions, ratings were significantly lower than degree of certainty, both for the red and blue square question [$t(25) = 3.45, p < .001$] and for the circle and triangle question [$t(25) = 3.11, p < .01$].

Discussion

The children clearly rated the high-probability question as more probable than the low-probability question. They were equally certain about their responses for both these questions. In addition, their ratings of certainty were significantly higher than the probability ratings for the low-probability questions. These results thus show quite clearly that young children can, indeed, differentiate between a probabilistic evaluation of a potential outcome and their own certainty about their response.

GENERAL DISCUSSION

The results of these studies support the conclusion that children are in the process of developing two different forms of inference, corresponding to counterexample-based reasoning and probabilistic reasoning. One strong indication of this is the fact that the relation between inferential responses and justifications observed in the first study clearly varied with age. For the 6- and 7-year-olds, there were strong correlations between the two for both categorical and probabilistic reasoning. However, for the 9-year-olds, this relation was maintained for the categorical inferences but disappeared for probabilistic inferences, which pattern is the same as that observed in adults (Markovits & Handley, 2005; Verschueren et al., 2005). The latter findings were interpreted as indicating differential use of counterexample-based and probabilistic strategies, and their presence in children of this age is strongly suggestive of the same form of difference.

We also found support for the hypothesis that making probabilistic conditional inferences is an early-developing form of reasoning that is relatively well established even by 6 years of age, whereas counterexample-based reasoning shows a stronger developmental increase between 6 and 9 years. Even 6-year-olds were able to make accurate probabilistic inferences, with both verbal and visually presented premises. Note that the latter results both confirm and extend those of Girotto and Gonzalez (2008)

Table 5
Mean Ratings Given to the High- and Low-Probability Questions and Mean Ratings of Degree of Certainty for the Two Problems

Problem	High Probability		Low Probability	
	Rating	Certainty	Rating	Certainty
Eight red squares and one blue square	3.61	3.81	2.62	3.92
Eight circles and one triangle	3.54	3.69	2.88	3.96

and reinforce the idea that children at this age possess clear intuitions of conditional probabilities. Development of abilities related to probabilistic inference in this age range was seen most clearly in the relative increase of use of intermediate points on the probability scale. This shows that children develop a more nuanced ability to translate probabilistic intuitions into explicit judgments. Nonetheless, the children were able to make the key distinctions involved in probabilistic inferences at an early age, and their ability to do so showed a slower pattern of developmental change between 6 and 9 years of age than did that for comparable categorical inferences, for which there was very large improvement in explicit use of counterexample-based reasoning between 6 and 7 years of age. We interpret this to mean that in contrast to probabilistic reasoning, basic counterexample-based reasoning skills are less well developed by the age of 6, but do develop rapidly over the next few years.

The data from Study 2 suggest that part of the difficulty experienced by 6-year-olds is their ability to access counterexample information on an inference task. Specifically, the fact that 6-year-olds do better on categorical inferences when premise parameters are presented visually and are directly accessible, at least when these inferences are presented before probabilistic inferences, is, in turn, consistent with the idea that one of the limits of their ability to do this form of reasoning is related to information retrieval. This is reinforced by the results from Study 1, which show a clear developmental increase in explicit counterexample use with verbally presented premises (see Janveau-Brennan & Markovits, 1999, for related evidence among older children). Nonetheless, simple accessibility of counterexamples is not sufficient for adequate counterexample-based reasoning, since between 30% and 40% of 6-year-olds continue to accept the AC inference even when counterexamples are directly visible and when categorical problems are given first, whereas the equivalent percentages are 0% and 15% for the oldest children. This suggests that the ability to explicitly represent counterexamples in a way that allows rejection of a putative conclusion when counterexample-based reasoning is employed is also a component of developing reasoning skills in children.

This, in turn, suggests an explanation for the interaction between the two forms of reasoning observed in the second study. In this case, children of all ages were able to make relatively accurate ordinal judgments of the conditional probabilities required to evaluate the likelihood of the AC inference. However, when asked to make such probabilistic judgments first, the children then rejected the AC inference in the categorical task at a much lower rate. No such interaction was observed in the first study, where the premises were verbally presented. This implies that the locus of the interaction specifically concerned the way in which visually presented information was processed. Recall that models of probabilistic reasoning (e.g., Oaksford et al., 2000; Thompson, 2000; Verschuere et al., 2005) suggest that likelihood judgments are made by an intuitive, low-cost evaluation of the probability of a putative conclusion. In contrast, counterexample-based reason-

ing requires the retrieval and the explicit representation of at least one counterexample (Markovits & Barrouillet, 2002). When reasoning with verbal premises, reasoners must activate stored knowledge from long-term memory in similar ways, regardless of which type of reasoning is used. However, when cases are visually presented, as in the second study, the retrieval process can be very different. When making a likelihood judgment, children might scan the array of stimuli in order to arrive at a rapid and intuitive judgment about the relative numbers of objects consistent with a conclusion or not. If such a process were carried over to a subsequent categorical reasoning task, it would make it more difficult for children to explicitly attend to specific counterexamples in order to incorporate at least one of these into a representation of the premises. This would then lead to the observed tendency to accept the AC inference. In contrast, a more deliberate examination of the visible cases required for both accessing counterexamples and using them in a representation, which is necessary for counterexample-based reasoning, would still allow the kind of evaluation required for a likelihood judgment. Thus, the observed interaction is certainly consistent with the notion that the processes involved when children make probabilistic inferences can, indeed, differ from those used to make counterexample-based reasoning.

These results are at least partly consistent with dual-process theories (e.g., Evans & Over, 1996; Klaczynski, 2001; Stanovich & West, 2000). They indicate that the kind of intuitive probabilistic conditional reasoning that can be identified with the heuristic system does, indeed, develop earlier than basic counterexample-based reasoning. The results of the second study, where AC inferences were designed to have differing probabilities that were directly observable, show that this probability manipulation also influences categorical reasoning. The results of the second study also show that priming processes associated with probabilistic reasoning make counterexample-based reasoning more difficult, at least with visible stimuli. These results are consistent with the idea that heuristic processes can sometimes interfere with more working-memory-intensive analytic processes.

However, it is also important to note that although a dual-process interpretation of the present results is reasonable, the distinction between heuristic and analytic processes might not apply in a straightforward manner in the present case. Specifically, our analysis of counterexample-based reasoning suggests that it requires both retrieval of counterexamples and the ability to explicitly represent these. Although the latter process might be reasonably considered to require the kind of cognitive resources that characterize analytic processes, the nature of the retrieval component is more ambiguous. On the one hand, there is clear evidence that the counterexample retrieval used in conditional reasoning is sometimes automatic (e.g., De Neys, Schaeken, & d'Ydewalle, 2005a). On the other hand, there is evidence that when premises do not allow easy access to counterexamples, successful conditional reasoning is mediated by retrieval processes that do require cognitive resources—for example, retrieving weakly

associated alternatives (Markovits & Quinn, 2002). This suggests that counterexample-based reasoning can imply differing forms of processing related to counterexample retrieval and that the distinction between heuristic and analytic processes is not as clear-cut in the present case as dual-process theories would imply.

Finally, however they are interpreted, these results provide convincing evidence that the ability to make probabilistic conditional inferences based on very familiar verbal premises, or on premises using observable stimuli, is present by 6 years of age. The ability to make simple counterexample-based categorical inferences, although clearly present at this age, is more embryonic and shows a sharp developmental increase between 6 and 9 years of age. Although interactions between counterexample use and the kinds of inferences that are made suggest some clear commonality in their underlying processes, there is an increasing differentiation between them over this age range.

AUTHOR NOTE

Preparation of the manuscript was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to H.M. Correspondence concerning this article should be addressed to H. Markovits, Department of Psychology, Université du Québec à Montréal, C. P. 8888, Succ Centre-Ville, Montréal, QC, H3C 3P8 Canada (e-mail: markovits.henry@uqam.ca).

REFERENCES

- ACREDOLO, C., O'CONNOR, J., BANKS, L., & HOROBIN, K. (1989). Children's ability to make probability estimates: Skills revealed through application of Anderson's functional measurement methodology. *Child Development*, *60*, 933-945.
- BARROUILLET, P., & LECAS, J.-F. (1999). Mental models in conditional reasoning and working memory. *Thinking & Reasoning*, *5*, 289-302.
- BARROUILLET, P., MARKOVITS, H., & QUINN, S. (2002). Developmental and content effects in reasoning with causal conditionals. *Journal of Experimental Child Psychology*, *81*, 235-248.
- BRAINERD, C. J. (1981). Working memory and the developmental analysis of probability judgment. *Psychological Review*, *88*, 463-502.
- BUCCI, W. (1978). The interpretation of universal affirmative propositions. *Cognition*, *6*, 55-77.
- CONOVER, W. J. (1980). *Practical nonparametric statistics* (2nd ed.). New York: Wiley.
- 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.
- DE NEYS, W., SCHAEKEN, W., & D'YDEWALLE, G. (2002). Causal conditional reasoning and semantic memory retrieval: A test of the semantic memory framework. *Memory & Cognition*, *30*, 908-920.
- DE NEYS, W., SCHAEKEN, W., & D'YDEWALLE, G. (2005a). Working memory and counterexample retrieval for causal conditionals. *Thinking & Reasoning*, *11*, 123-150.
- DE NEYS, W., SCHAEKEN, W., & D'YDEWALLE, G. (2005b). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning*, *11*, 349-381.
- EVANS, J. ST. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgement*. Hove, U.K.: Psychology Press.
- EVANS, J. ST. B. T., HANDLEY, S. J., & OVER, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *29*, 321-335.
- EVANS, J. ST. B. T., & OVER, D. E. (1996). *Rationality and reasoning*. Hove, U.K.: Psychology Press.
- GIROTTO, V., & GONZALEZ, M. (2008). Children's understanding of posterior probability. *Cognition*, *106*, 325-344.
- 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). The development of reasoning with causal conditionals. *Developmental Psychology*, *35*, 904-911.
- JOHNSON-LAIRD, P. N., & BYRNE, R. M. J. (1991). *Deduction*. Hove, U.K.: Erlbaum.
- KLACZYNSKI, P. A. (2001). Analytic and heuristic processing influences on adolescent reasoning and decision-making. *Child Development*, *72*, 844-861.
- KOKIS, J. V., MACPHERSON, R., TOPLAK, M. E., WEST, R. F., & STANOVICH, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, *83*, 26-52.
- 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. (2004). Information processing and reasoning with premises that are empirically false: Interference, working memory, and processing speed. *Memory & Cognition*, *32*, 592-601.
- 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., & HANDLEY, S. (2005). Is inferential reasoning just probabilistic reasoning in disguise? *Memory & Cognition*, *33*, 1315-1323.
- MARKOVITS, H., & QUINN, S. (2002). Efficiency of retrieval correlates with "logical" reasoning from causal conditional premises. *Memory & Cognition*, *30*, 696-706.
- MARKOVITS, H., & VACHON, R. (1990). Conditional reasoning, representation, and level of abstraction. *Developmental Psychology*, *26*, 942-951.
- MARKOVITS, H., VENET, M., JANVEAU-BRENNAN, G., MALFAIT, N., PION, N., & VADEBONCOEUR, I. (1996). Reasoning in young children: Fantasy and information retrieval. *Child Development*, *67*, 2857-2872.
- OAKSFORD, M., CHATER, N., & LARKIN, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 883-899.
- O'BRIEN, D. [P.], & OVERTON, W. F. (1980). Conditional reasoning following contradictory evidence: A developmental analysis. *Journal of Experimental Child Psychology*, *30*, 44-61.
- 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.
- ROSEN, V. M., & ENGLE, R. W. (1997). 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: *If* is not the biconditional. *Developmental Psychology*, *19*, 471-481.
- STANOVICH, K. E., & WEST, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral & Brain Sciences*, *23*, 645-665.
- TÉGLÁS, E., GIROTTO, V., GONZALEZ, M., & BONATTI, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences*, *104*, 19156-19159.
- THOMPSON, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition*, *22*, 742-758.
- THOMPSON, V. A. (2000). The task-specific nature of domain-general reasoning. *Cognition*, *76*, 209-268.
- VERSCHEUREN, N., SCHAEKEN, W., & D'YDEWALLE, G. (2005). Everyday conditional reasoning: A working memory-dependent tradeoff between counterexample and likelihood use. *Memory & Cognition*, *33*, 107-119.

(Manuscript received November 26, 2007;
revision accepted for publication April 2, 2008.)