



Article scientifique

Article

2008

Published version

Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

Mental Models and the Suppositional Account of Conditionals

Barrouillet, Pierre Noël; Gauffroy, Caroline; Lecas, Jean-François

How to cite

BARROUILLET, Pierre Noël, GAUFFROY, Caroline, LECAS, Jean-François. Mental Models and the Suppositional Account of Conditionals. In: Psychological review, 2008, vol. 115, n° 3, p. 760–771. doi: 10.1037/0033-295X.115.3.760

This publication URL: <https://archive-ouverte.unige.ch/unige:88298>

Publication DOI: [10.1037/0033-295X.115.3.760](https://doi.org/10.1037/0033-295X.115.3.760)

THEORETICAL NOTE

Mental Models and the Suppositional Account of Conditionals

Pierre Barrouillet and Caroline Gauffroy
Université de Genève

Jean-François Lecas
Université de Bourgogne

The mental model theory of conditional reasoning presented by P. N. Johnson-Laird and R. M. J. Byrne (2002) has recently been the subject of criticisms (e.g., J. St. B. T. Evans, D. E. Over, & S. J. Handley, 2005). The authors argue that the theoretical conflict can be resolved by differentiating 2 kinds of reasoning, reasoning about possibilities given the truth of assertions and reasoning about the truth of assertions given possibilities. The standard mental model theory accounts for the former kind of reasoning but does not adequately account for the latter, contrary to the suppositional approach favored by J. St. B. T. Evans et al. (2005). The authors thus propose a modified mental model theory of conditionals that reconciles the 2 theoretical approaches. It is demonstrated that this theory is able to explain the key findings that have been opposed to the standard theory by J. St. B. T. Evans et al. and makes new predictions that are empirically verified.

Keywords: reasoning, conditional, mental models, suppositional theory, development of reasoning

Because conditional statements of the form *If p then q* can convey hypothetical thinking, they are essential to mundane as well as scientific reasoning and decision making. Accordingly, the way people use, understand, and reason from these sentences is one of the main questions addressed by cognitive psychologists in the field of human thinking and reasoning. The most influential and popular psychological account of conditional reasoning is probably the mental model theory (Johnson-Laird & Byrne, 1991, 2002; Johnson-Laird, Byrne, & Schaeken, 1992). According to Johnson-Laird and Byrne (2002), individuals reason by imagining the states of affairs that can occur when the conditional is true, that is, by constructing mental models of possibilities. However, from reasoning performance on tasks that require the evaluation of the truth or of the probability of conditionals, prominent researchers have recently argued that this theory is fundamentally mistaken and should be abandoned to adopt a suppositional and probabilistic account of conditional reasoning (Evans, Handley, & Over, 2003; Evans & Over, 2004; Evans, Over, & Handley, 2005).

This note presents a revised version of the mental model theory for basic¹ conditionals aimed at reconciling the two theoretical perspectives. We assume that there are two fundamentally different kinds of reasoning, which are (a) reasoning about possibilities given the truth of the conditional sentence and (b) evaluating the truth of a conditional sentence given some current state of affairs. The standard mental model theory proposed by Johnson-Laird and

Byrne (2002) has been designed to account for reasoning about possibilities, but its present form does not permit it to account for reasoning about truth values. On the other hand, the suppositional theory put forward by Evans and Over (2004) accounts perfectly well for this latter form of reasoning but still needs some specifications concerning reasoning about possibilities. Reconciling the two theories is made possible by augmenting the core principles of the mental model theory with postulates concerning the truth value of the conditional. We demonstrate that the resulting modified theory not only accounts for the facts at odds with the standard version but predicts new phenomena that are empirically verified.

The Standard Mental Model Theory and the Two Kinds of Reasoning

According to the mental model theory, people reason from conditional assertions of the form *If p then q* by constructing and manipulating mental models of the states of affairs to which these assertions refer. Within this theoretical framework, a mental model of an assertion represents a possibility given the truth of this assertion (Johnson-Laird & Byrne, 2002). This makes clear that the theory is mainly concerned with the psychological processes by which people construct the referents of assertions: Understanding and reasoning are processes by which people represent those state of affairs that can occur (i.e., possibilities) when these assertions are true. For example, a complete representation of a basic *If p then q* conditional would involve a representation of the following form:

Pierre Barrouillet and Caroline Gauffroy, Faculté de Psychologie et des Sciences de l'Éducation, Université de Genève, Genève, Switzerland; Jean-François Lecas, LEAD-CNRS, Université de Bourgogne, Dijon, France.

Correspondence concerning this article should be addressed to Pierre Barrouillet, Université de Genève, Faculté de Psychologie et des Sciences de l'Éducation, 40 bd du Pont D'arve, 1205 Genève, Switzerland. E-mail: Pierre.Barrouillet@pse.unige.ch

¹ According to Johnson-Laird and Byrne (2002), basic conditionals are conditionals in which the antecedent and the consequent have no semantic or referential relations, or relations based on knowledge.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

p	q
$\neg p$	q
$\neg p$	$\neg q$

in which each line denotes a mental model representing a possible state of affairs. The first line denotes a model referring to the possibility in which both propositions p and q are satisfied, whereas the second line denotes a model referring to the possibility in which q is satisfied but p is not, the \neg sign standing for negation. As noted by Johnson-Laird and Byrne (2002), each model corresponds to a true row in the truth table of a material implication (see Table 1), but mental models would differ from the lines of a truth table because they do represent not truth values but possibilities:

Each entry in a truth table represents the truth or falsity of an assertion given a particular possibility. In contrast, each mental model in a set represents a possibility. A corollary is that possibilities are psychologically basic, not truth values. Discourse about the truth or falsity of propositions is at a higher level than mere descriptions of possibilities. (Johnson-Laird & Byrne, 2002, p. 653)

Our claim is that such a distinction between possibilities and truth values delineates two different kinds of reasoning. In the first kind, people reason from assertions they consider as true and try to find out what must, can, or cannot occur in the world described by these assertions. We refer to this kind of reasoning as *reasoning about possibilities*. In the second kind of reasoning, people start from a given state of affairs they consider as existing, and they judge if a given assertion is true or false in this case. This second line of reasoning runs from states of affairs to language and consists of *reasoning about truth values*. The former kind would be psychologically basic, whereas the latter is more complex and difficult, involving some form of metalogic, a meta-ability that requires a grasp of the relations between assertions and the world through the predicates true and false (Johnson-Laird & Byrne, 2002; Moshman, 1990).

This dichotomy permits us in turn to distinguish between two kinds of tasks used in conditional reasoning studies. The first category involves reasoning about possibilities and consists of those tasks in which participants are given some conditional premise (and eventually additional information) they have to consider as true and are asked to imagine either the resulting states of affairs or those that cannot occur. Within this category are the production and evaluation of conditional inferences (see Evans, Newstead, & Byrne, 1993, for a complete description), as well as those tasks in which participants are asked to identify or to produce those cases that are either compatible or incompatible with a conditional statement given as a rule to be obeyed (e.g., Barrouillet & Lecas, 1998, 1999; Marcus & Rips, 1979). The second category

does not involve reasoning from but reasoning about a conditional statement the truth of which is under inquiry and must be evaluated. This category comprises the truth-testing task in which participants are given different states of affairs and asked to evaluate for each of them the truth of a conditional statement, the probability tasks in which participants are asked to evaluate the probability that a conditional is true or false, and also Wason's (1966) selection task in which participants are asked to indicate what kind of evidence they need to know if a conditional rule is either true or false.

The difference between these two forms of reasoning and the two types of tasks is often overlooked, as frequent confusions between two categories of terms testify. The first denotes the truth value of assertions that can be said to be true or false, the truth value remaining in some occasions indeterminate (e.g., we show that for a majority of adults, the truth value of a conditional remains indeterminate when the sole available evidence consists of *not p* cases). The second category of terms denotes the status of cases, states of affairs, regarding some assertion. States of affairs can be said compatible or incompatible with a given assertion. When the assertion is known to be true, they are considered possible or impossible, respectively (e.g., the $p \neg q$ case is incompatible with the conditional *If p then q* and impossible when it is true). When evaluating an assertion, a given state of affairs can be said to be irrelevant if it does not convey information about the truth or falsity of the assertion (see Sevenants, Schroyens, Dieussaert, & Schaeken, 2006, for an interesting discussion about the term *irrelevant*). We are aware that this distinction might be considered trivial, but unfortunately, it is not. Indeed, confusions and misuses are frequent. For example, Evans et al. (2003, p. 322) evoked a task used by Evans (1972) and stated, "If asked exhaustively to construct cases that conform to the conditional and cases that contradict it, people tend to omit false antecedent cases in both tasks." This description clearly refers to a task involving reasoning about possibilities: People have to construct cases that conform to a conditional or that contradict it, this conditional being considered as a rule to be obeyed. Now, consider how Evans et al. (1993, p. 49) presented the same task: "Subjects were shown an array of coloured shapes and were asked exhaustively to identify cases which made the rule true and cases which made the rule false." This description clearly refers to a truth table task, that is, a task involving reasoning about truth values. From one description to another, it appears that "cases that conform to the conditional" are confused with "cases which made the rule true." This is of importance because there is ample evidence that the two kinds of tasks do not produce the same patterns of responses (Sevenants et al., 2006). Contrary to Evans et al.'s (2003) claim, we have demonstrated in many studies that when adolescents and adults are asked to construct cases that conform to a conditional rule, that is, cases that are compatible with the conditional, they do not omit false-antecedent cases (Barrouillet & Lecas, 1998; Lecas & Barrouillet, 1999; see Figure 1). It is only when they are asked to indicate those cases that make the rule true that people omit false-antecedents cases, as several studies demonstrated (see Evans et al., 1993, for a review). Thus, empirical evidence indicates that, for logically untrained individuals, tasks involving reasoning about possibilities and tasks involving reasoning about truth values elicit different processes and should be distinguished.

Table 1
The Truth Table for a Material Implication p Implies q and the Truth Table Known as Defective for an If p Then q Conditional

p	q	p implies q	Defective table for <i>If p then q</i>
True	True	True	True
True	False	False	False
False	True	True	Indeterminate
False	False	True	Indeterminate

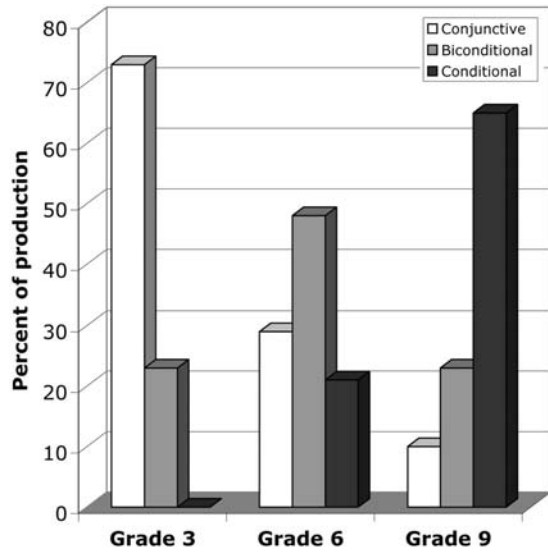


Figure 1. Developmental changes in the rate of response patterns when children are asked to construct those cases that are compatible with a basic conditional. Conjunctive: construction of the sole $p q$ case; biconditional: construction of $p q$ and $\neg p \neg q$; conditional: construction of $p q$, $\neg p \neg q$, and $\neg p q$. Data from Lecas and Barrouillet (1999, p. 378, contingency context).

We claim that the contrast between the two forms of reasoning delineates a contrast between theories and their respective strengths and weaknesses. Because the mental model theory is mainly a description of the psychological processes by which individuals construct and manipulate representations of possibilities, it accounts quite well for behaviors induced by the former set of tasks. Accordingly, advocates of the mental model theory put forward empirical evidence mainly issuing from reasoning about possibilities. As a corollary, the strongest criticisms come from authors who privilege tasks about truth value in studying conditional reasoning. As stressed by Evans et al. (2005), how people decide to what extent to believe ordinary conditionals is a fundamental question. It is thus not surprising that the suppositional theory defended by Evans et al. mainly addresses reasoning about truth value. We in turn present the two theoretical realms and the empirical evidence on which they are based.

Mental Models and Reasoning About Possibilities

As we noted earlier, most of the behavioral evidence put forward by mental model theorists comes from reasoning about possibilities. The main predictions of the theory result from the so-called *principle of implicit models* stating that basic conditionals have mental models representing the possibilities in which their antecedents are satisfied but have only implicit mental models for the possibilities in which their antecedents are not satisfied. Thus, individuals construct an initial model representing both p and q satisfied, the other possibilities (i.e., $\neg p q$ and $\neg p \neg q$) remaining implicit (the three dots in the following diagram):

$p \quad q$
 ...

Implicit information can be made explicit through a demanding and time-consuming fleshing-out process (Johnson-Laird & Byrne, 1991; Schroyens, Schaeken, & Handley, 2003). Many studies designed to test specific predictions issuing from this account supported the theory. It was demonstrated that when the order of the premises reduces the number of models to be held, reasoning improves in a conditional syllogisms task (Giroto, Mazzocco, & Tasso, 1997). Chronometric studies established that the time course of conditional inferences is better predicted by the mental model theory than by alternative accounts (Barrouillet, Grosset, & Lecas, 2000). Compelling but illusory conditional inferences are predicted by the theory and indeed have been observed (Johnson-Laird & Savary, 1999). Finally, large-scale meta-analyses on conditional inferences established that the mental model theory provides a better account of the data than, for example, probabilistic accounts of conditional reasoning (Schroyens & Schaeken, 2003).

According to Johnson-Laird and Byrne (2002), the most compelling corroboration of the principle of implicit models comes from developmental studies. In several studies, we asked children to list the possibilities either compatible or incompatible with basic conditionals (Barrouillet, 1997; Barrouillet & Lecas, 1998, 1999, 2002; Lecas & Barrouillet, 1999). All the results converged toward a clear developmental trend. In a first developmental level, children are able to construct only one model that corresponds to the initial model constructed by adults but are unable to complete their representation with additional models. As a consequence, they exhibit a conjunctive interpretation based on the $p q$ model representing the sole possibility they consider as consistent with basic conditionals, the other possibilities (i.e., $\neg p q$, $\neg p \neg q$, and $p \neg q$) being considered as incompatible. Several studies have reported this conjunctive interpretation in young children (Kuhn, 1977; Paris, 1973; Taplin, Staudenmayer, & Taddonio, 1974). In a second level, usually observed in young adolescents, a second $\neg p \neg q$ model can be added, leading to a biconditional interpretation. At this level, the $p q$ and $\neg p \neg q$ possibilities are judged as compatible with the conditional, whereas the $\neg p q$ and $p \neg q$ possibilities are considered as incompatible. Finally, older adolescents and adults predominantly exhibit a complete conditional interpretation in which all the possibilities are considered as compatible with the conditional except the $p \neg q$ case, which is judged as incompatible. Such a developmental trend is particularly clear when children and adolescents are asked to construct cases compatible with a conditional statement (Lecas & Barrouillet, 1999, Figure 1). Barrouillet et al. (2000) demonstrated that this developmental trend was corroborated by the developmental changes in the production rates of the four canonical conditional inferences (see Table 2).

In summary, studies on reasoning about possibilities support the mental model theory and the three-model account of adult reasoning from basic conditionals, that is, those that are not modulated by meaning, reference, or knowledge. As we show, difficulties with the standard mental model theory arise when considering reasoning about truth values rather than reasoning about possibilities.

Reasoning About Truth Values and the Suppositional Account

As we noted above, the mental model theory has recently been the object of strong criticisms by those who consider that the theory is not only inaccurate but fundamentally mistaken (Evans &

Table 2
Developmental Levels of Interpretation of If p Then q Conditionals, Mental Models Constructed, and the Four Canonical Inferences That Should Be Produced (+) or Not (-) According to Barrouillet, Grosset, and Lecas (2000)

			Levels of interpretation and models constructed					
			Conjunctive		Biconditional		Conditional	
Inference	Minor premise	Conclusion	<i>p</i>	<i>q</i>	<i>p</i>	<i>q</i>	<i>p</i>	<i>q</i>
			$\neg p$	$\neg q$	$\neg p$	$\neg q$	$\neg p$	$\neg q$
MP	<i>p</i>	<i>q</i>	+		+		+	
AC	<i>q</i>	<i>p</i>	+		+		-	
DA	<i>not p</i>	<i>not q</i>	-		+		-	
MT	<i>not q</i>	<i>not p</i>	-		+		+	

Note. MP = modus ponens; AC = affirmation of consequent; DA = denial of antecedent; MT = modus tollens.

Over, 2004; Evans et al., 2005). Evans and colleagues (2005) assumed first that the mental model theory gives a psychological account of the conditionals that is equivalent to the material implication (see Oberauer & Wilhelm, 2003, for a related claim), then argued that many empirical findings contradict this conception, and finally concluded that the mental model theory is mistaken. According to Evans and Over (2004), the most direct evidence against the mental model theory comes from studies of the how people judge the truth value or the probability of conditional statements. Indeed, many studies reported that, when asked to evaluate a conditional statement from the four cases defined by either the affirmation or the negation of the antecedent and the consequent, individuals frequently indicate a defective truth table in which *not p* instances (i.e., $\neg p q$ and $\neg p \neg q$) are described as irrelevant to the truth of the conditional, whereas a material implication interpretation should lead to consider the conditional as true in these cases (see Table 1).

In the same way, recent studies demonstrated that when asked to evaluate the probability of conditional statements, people rarely produce evaluations compatible with the material implication interpretation but rather adopt a suppositional interpretation. For example, in Evans et al. (2003), participants were given frequency information about the cases *p q*, *p $\neg q$* , *$\neg p q$* , and *$\neg p \neg q$* in a pack containing cards that were either yellow or red and had either a circle or a diamond printed on them (say, 1 yellow circle, 4 yellow diamonds, 16 red circles, and 16 red diamonds), and they were asked to evaluate the probability that a claim like "If the card is yellow then it has a circle printed on it" was true for a card drawn at random from the pack. According to Evans et al., a material implication account and, thus, the mental model theory would predict a probability $P(\text{MI}) = P(p q) + P(\neg p q) + P(\neg p \neg q)$ because the material conditional is true in each of these three cases (but see Schroyens & Schaeken, 2004). However, this kind of response was practically never observed, participants responding mainly by evaluating the probability of the conditional as the conditional probability $P(q/p) = P(p q) / [P(p q) + P(p \neg q)]$ or the conjunctive probability $P(p q)$. Oberauer and Wilhelm (2003)

basically observed the same phenomena in a similar but independent study.

According to Evans et al. (2005), these facts indicate that the psychological meaning of conditionals is not the material implication but is suppositional in nature. When evaluating the probability of a conditional, most individuals disregard the *not p* cases that are mainly judged as irrelevant to the truth of the conditional and direct their attention to those cases in which the antecedent is satisfied, using the Ramsey test. They hypothetically add *p* to their stock of knowledge and evaluate their degree of belief in *q* given *p*. This procedure leads, as Evans et al. (2003) observed, to evaluating the probability of the conditional as the conditional probability. According to Evans et al. (2005), these facts are inconsistent with the mental model theory, which would predict that *If p then q* conditionals should be evaluated as the *not p* or *q* disjunction and the *If not q then not p* conditional, which both have the same extension and refer to the same set of possibilities (i.e., *p q*, $\neg p q$, and $\neg p \neg q$).

We would like to point out that the two main sources from which evidence against the mental model theory comes according to Evans and Over (2004) and Evans et al. (2005) are tasks in which people are asked to evaluate the truth value of conditionals. What is considered as contradicting the mental model theory is that people do not consider *not p* cases as making the conditional true,² thus indicating a defective table in truth table tasks, and adopting in probability tasks either a suppositional or a conjunctive interpretation in which the sole case making the conditional true is *p q*. Accordingly, Evans et al. (2005) claimed that the evidence of the defective truth table is a serious problem for Johnson-Laird and Byrne's (2002) theory.

Mental Models and Truth Values

It can be noted that most of Evans and colleagues' (2005) criticisms rely on the assumption that, within the mental model theory, the psychological meaning of a basic conditional is the truth function of the material implication (see also Oberauer & Wilhelm, 2003). However, Johnson-Laird and Byrne (2002) ex-

² It is important to note that in some studies, participants considered the *not p not q* case as making the conditional true when this case was presented in a verbal form involving explicit negations (Evans, 1983; Evans, Clibbens, & Rood, 1996). However, we are interested here in deciphering the processes that underpin reasoning from sentences to state of affairs on the one hand (i.e., reasoning about possibilities) and from reasoning from state of affairs to sentences on the other (i.e., reasoning about truth values). Thus, we refer to studies in which participants were asked to evaluate the truth value of conditionals from states of affairs, not verbal descriptions with explicit negations. Within this paradigm, *not p* cases are rarely considered as making the conditional true. For example, in the following experiment using pictorial and not verbal material to represent possibilities, *not p not q* cases are judged as making the conditional true in less than 1% of the trials. It is this phenomenon that Evans and colleagues (2005) referred to and that we aim at accounting for within our revised mental model theory. Moreover, the "true" responses for *not p not q* cases with explicit negations do not call the psychological existence of the defective truth table into question because they do not extend to *not p q* cases that are very rarely considered as making the rule true. The increase in "true" responses for *not p not q* cases with explicit negations could result from biconditional interpretations. It has indeed been demonstrated that explicit rather than implicit negations induce a biconditional pattern of inference production (Barrouillet & Lecas, 1998).

implicitly rejected this idea, assuming that conditionals are not truth functional. These diametrically opposed readings at least suggest that there is something unclear in the standard theory concerning truth values. The disagreement in interpreting the mental model theory comes from two main points in Johnson-Laird and Byrne: the psychological reality of the defective truth table and the epistemic and psychological status of the paradoxes of the material implication (i.e., given *not p*, it validly follows for the ordinary conditional *If p then q*, and given *q*, it validly follows for the ordinary conditional *If p then q*).

As far as the paradoxes are concerned, Johnson-Laird and Byrne (2002) considered these inferences as valid but argued that their oddity has nothing to do with conditionals. According to Evans et al. (2005), this claim confirms the truth functional meaning of basic conditionals within the mental model theory because the paradoxes are valid for a conditional if and only if the conditional is truth functional.

The most important point concerns the defective truth table. Johnson-Laird and Byrne (2002) considered that the defective truth table account of conditionals defended by Quine (1952), by Wason and Johnson-Laird (1972), and more recently by Evans and Over (2004) is plausible at first sight but founders on the case of biconditionals. Indeed, according to Johnson-Laird and Byrne, the biconditional *If and only if p then q*, which is synonymous with the conjunction *If p then q, and if not p then not p*, has a complete truth table (true for *p q* and *not p not q* and false in the other cases). Thus, if the conditionals had a defective truth table, Johnson-Laird and Byrne wondered how the conjunction of two conditionals with defective truth tables could result in the complete truth table of the biconditional. Of course, they were aware of the fact that individuals often consider *not p* cases as being irrelevant to the truth of the conditional. They explained this result by the nature of the initial model: Because the possibilities in which the antecedent is false are not explicitly represented, the conditional is deemed neither true nor false for these cases that are thus considered as irrelevant.

We agree that this account could hold for the cases in which individuals construct only the initial model, but what would happen if people flesh out the initial model and make explicit the possibilities in which the antecedent is false? It seems that the answer given by the standard mental model theory is that individuals would consider that *not p* possibilities, once explicitly represented, make the conditional true. This can be inferred from Girotto and Johnson-Laird (2004), who studied the way individuals evaluate the probability of conditionals. Apart from strategies based on the sole initial model, Girotto and Johnson-Laird described a complete strategy in which participants use the fully explicit models of the conditional (i.e., the three models). Within this strategy, reasoners would “compute the complete partition and treat a conditional as true in any case compatible with what is possible according to the conditional” (Girotto & Johnson-Laird, 2004, p. 211). These authors presented the complete strategy as difficult because more models have to be considered, but they did not reject it as psychologically implausible, and labeling it as complete suggests that, in their view, it is the most elaborated. In other words, according to the standard mental model theory, people who grasp that *not p* cases are compatible with the conditional should consider these possibilities as making the conditional true.

In summary, there is some contradiction in the way the standard mental model theory accounts for reasoning about possibilities and

reasoning about truth values. When considering reasoning about possibilities, Johnson-Laird and Byrne (2002) claimed that conditionals are not truth functional, but when considering reasoning about truth value, the standard theory denies the psychological reality of the defective truth table and assumes that reasoners should conform to the material implication meaning. Of course, the former proposal can be understood because the meaning of the conditional is captured by a set of mental models that represent possibilities and not truth values. However, critics are not wrong when claiming that, within the mental model theory, the psychological meaning of a basic conditional is the truth function of the material implication because, when considering judgments about truth and falsity, this theory assumes that reasoners who rely on fully explicit models would judge the conditional as true for *p q*, *not p q*, and *not p not q* cases and false for the *p not q* case, which is exactly the truth table of the material implication. The following theoretical proposal resolves this contradiction.

A Theoretical Proposal

We must admit that, like Evans et al. (2005), we are not convinced by Johnson-Laird and Byrne’s (2002) account of the defective table, which would result from an initial representation in which possibilities with a false antecedent are not explicitly represented. As noted by Evans et al., this initial model can be fleshed out with *not p* models, and this fleshing out does not appear so difficult if we consider the high production rate of modus tollens usually observed, this inference necessitating the construction of the $\neg p \neg q$ in the model theory (see, e.g., Barrouillet et al., 2000, for production rates of modus tollens higher than 80% in adolescents). Moreover, we have shown in many experiments that when asked to list the possibilities compatible with a basic conditional, older adolescents and adults have no difficulties evoking *not p* cases, especially *not p not q* (Barrouillet & Lecas, 1998). This evocation is even easier when cases are not to be constructed but are presented to participants who have just to evaluate their compatibility with a given conditional. In this case, even adolescents massively recognize that *not p not q* cases are compatible with basic conditionals (Barrouillet & Lecas, 1998; Lecas & Barrouillet, 1999). It should be noted that the tasks used by Barrouillet and Lecas are very akin to truth-testing tasks, the sole difference being that instead of evaluating the truth value of the conditional rule, participants are asked to identify those cases that are incompatible with it. The *not p not q* possibility is so easily represented by reasoners in these tasks and so massively considered as compatible with the conditional that it becomes impossible to understand why it is evaluated as irrelevant in truth table tasks. Finally, if the indeterminate responses and the defective table were due to a failure in fleshing out the initial model, these responses should be more frequent in individuals with lower cognitive capacities. However, it has recently been demonstrated that adults with higher and not lower cognitive ability are more likely to give defective truth table responses (Evans, Handley, Neilens, & Over, 2007). Thus, if we want to preserve the mental models as a theoretical framework for conditional reasoning, some modifications of the standard theory are needed to account for reasoning about truth values without leaving the capacity of the theory to account for reasoning about possibilities behind.

Recall that the defective table consists in evaluating the conditional as true for the $p \rightarrow q$ case, as false for the $p \rightarrow \neg q$ case, and as indeterminate for the *not p* cases that are considered as irrelevant. Thus, among the three possibilities that constitute the core meaning of the basic conditional according to the standard theory, individuals consider only one of them as making the conditional true. It is worth noting that this is not a priori incompatible with the mental model approach because the three possibilities are not psychologically equivalent, one being explicitly represented within the initial model while the others remain initially implicit and can only be represented explicitly through the fleshing-out process (Johnson-Laird & Byrne, 1991; Johnson-Laird et al., 1992). Accordingly, in their account of the defective table presented above, Johnson-Laird and Byrne (2002) hinted that individuals would consider the conditional as true only for the possibilities explicitly represented. However, we have seen that this account does not work because, when fleshed out, the mental models would no longer lead to the defective truth table but to the truth function of the material implication.

Nonetheless, Johnson-Laird and Byrne's (2002) proposal introduces a new and interesting dichotomy between initial and fleshed-out models that goes far beyond the explanation in terms of cognitive economy on which the notion of the initial model was based (Johnson-Laird & Byrne, 1991). Our proposal is that the distinction between initial and fleshed-out models is epistemic in nature: the initial model would be restricted to those possibilities that make the assertion true, the other compatible possibilities necessitating a fleshing-out process to be made explicit. Introducing this new distinction necessitates only a slight modification of the principle of truth. Recall that this principle states that "each mental model of a set of assertions represents a possibility given the truth of the assertions, and each mental model represents a clause in these assertions only when it is true in that possibility" (Johnson-Laird & Byrne, 2002, p. 653). We suggest adding the following: The possibilities that make the assertions true give rise to explicit mental models, whereas the others lead to models that remain initially implicit.

Following the principle of implicit models, "basic conditionals have mental models representing the possibilities in which their antecedents are satisfied, but only implicit mental models for the possibilities in which their antecedents are not satisfied" (Johnson-Laird & Byrne, 2002, p. 654). It follows from the modification introduced above that basic conditionals will be evaluated as true only for those possibilities represented within the initial model in which the antecedent is satisfied (i.e., $p \rightarrow q$). Possibilities in which the antecedent is not satisfied (i.e., the *not p* cases) are not part of the initial model and thus remain initially implicit because they do not make the conditional true, but the conditional cannot either be false in these cases because they can occur given its truth and are consistent with it. As a consequence, the truth value of the conditional should be considered as indeterminate for these possibilities. By contrast, those possibilities incompatible with the conditional, that is, the complement of the set of explicit and implicit models ($p \rightarrow \neg q$), falsify the conditional.

It is worth noting here that the main distinction we introduce is not between implicit and explicit models but between initial and fleshed-out models. Our proposal substantiates the principle of implicit models by providing a psychological basis for the distinction between models represented explicitly right from the start and

those that remain initially implicit. Among the possibilities compatible with a sentence, those that make it true are represented within the initial model, while the others require the fleshing-out process to be fully represented. However, even when explicitly represented, they are not considered as making the sentence true because they were not part of the initial model. In other words, the fleshing-out process can be understood as a process by which people extend the initial meaning of an assertion beyond the states of affairs that make it true to those states of affairs that are only compatible with it.³

It is also important to note that the additional assumption that we propose does not lead to any contradiction. In particular, assuming that some possibilities do not make the conditional true when they occur does not contradict the principle of truth stating that mental models represent those states of affairs that are possible when the assertion is true. It does not necessarily follow from this principle that each of these possibilities makes the assertion true when it occurs, because a given state of affairs can be compatible with an assertion, and then possible, without making this assertion true, whereas another state of affairs can make the assertion true. For example, an *A* that is not a *B* is compatible with the statement *Some As are Bs* and is a part of its complete representation, but it does not make this statement true, whereas an *A* that is also a *B* does. In the same principle of truth, Johnson-Laird and Byrne (2002) went further, stating that each mental model represents a clause in these assertions only when it is true in that possibility. Does this proposal indicate that each possibility should make the conditional true when it occurs? We do not think so. This proposal indicates only that the content of a given model represents what is true within this possibility. Within the initial representation of the conditional *If there is an A then there is a B*,

A *B*

the model represents the atomic clauses *There is an A* and *There is a B* because they are both true in that possibility. Thus, it does not follow from the principle of truth that people should evaluate a conditional as true for each possibility by which they represent the conditional in a complete representation.

Our theory accounts perfectly for the phenomena observed in the truth-testing task (i.e. the defective truth table) and also in the probability task. When asked to evaluate the probability of a conditional, people concentrate on those cases for which the conditional can be readily evaluated, that is, p cases. Thus, they give the conditional probability if they consider the ratio between p and q

³ This proposal seems to extend beyond basic conditionals and may be a general law within the mental models theory. For example, it can be observed that the initial model of the disjunction $p \vee q$ is restricted to the situations that make it true when they occur:

p

q

The fleshing-out process adds information that is compatible with the disjunction but that does not make it true when occurring in isolation (i.e., *not p* within the first model or *not q* within the second). It can also be noted that the model $p \vee q$ is not part of the initial representation because all the disjunctions are not true in this case.

cases and p cases or the conjunctive probability if they consider simply the proportion of cases in which the conditional is true among all the possible cases (Evans et al., 2003; Girotto & Johnson-Laird, 2004). This explains also why, as Evans et al. (2005) noted, the probability of a conditional is never judged to be equal to *not p or q*. Indeed, according to Johnson-Laird et al. (1992), the initial model of the disjunction *not p or q* contains two models:

$$\begin{array}{l} \neg p \\ q \end{array}$$

Thus, according to our hypothesis, the conditional is true only for the p q case, whereas the disjunction is made true by the occurrence of *not p* and the occurrence of q . Thus, there is no reason to expect the same judgment of probability for the two assertions. For the same reason, *If p then q* does not have, psychologically, the same probability as *If not q then not p*, as Evans et al. (2003) demonstrated. Their initial models differ (i.e., p q for the former and $\neg q \neg p$ for the latter), and as a consequence, these two conditionals are psychologically true for different possibilities even if they are in the end compatible with the same possibilities.

An Empirical Prediction

Our account permits a striking and new prediction that can be easily tested. We assume that the possibility represented by the initial model of a conditional makes this conditional true, those captured by the models constructed through fleshing out leave the truth value of the conditional indeterminate, and the complement of the previous possibilities makes the conditional false. Thus, the evaluation of basic conditionals should depend on the way the initial model is fleshed out and individuals construct mental models. Among the various factors that can affect this construction, a simple way to test our hypothesis is to take advantage of the age-related changes in the fleshing-out process. As we have described above, the set of models constructed to represent basic conditionals varies with age; if our hypothesis is correct, these variations should affect the evaluation of conditionals in a precise way. For example, when listing cases compatible with the conditional, children often construct only the initial model p q but are unable to think about additional possibilities. Thus, according to our hypothesis, when evaluating a basic conditional, they would consider it as true for the p q possibility but false in all the other cases, including *not p* cases. By contrast, when reasoning about possibilities, adolescents tend to favor a biconditional interpretation by fleshing out this initial model with an additional $\neg p \neg q$ model. They should thus evaluate the conditional as true for the p q case, as indeterminate for the $\neg p \neg q$ case constructed through fleshing out, and as false in the other cases (i.e., $p \neg q$ but also $\neg p q$). Finally, most of the oldest adolescents and adults are able to construct the three-model representation, and they should produce the defective truth table described above when evaluating conditionals. Thus, the evaluation of basic conditionals should vary with the set of mental models children, adolescents, and adults are able to construct. In this respect, the processing of *not p* cases should be critical. Both *not p* cases should initially be considered as making the conditional false, and the evaluation should progressively turn into the response of indeterminacy. However, this developmental

change should occur sooner in development for the $\neg p \neg q$ than for the $\neg p q$ possibility. We tested these hypotheses in the following experiment in which third, sixth, and ninth graders and adults were presented with a truth-testing task.

Experiment

Method

Forty-seven third graders (mean age = 8.6 years), 38 sixth graders (mean age = 11.5 years), 39 ninth graders (mean age = 15.6 years), and 47 undergraduate students (mean age = 28.5 years) in education at the Université de Bourgogne, Dijon, France, participated as volunteers. A truth-testing task was administered to groups of about 20 participants using a video projector connected to a computer. In each trial, we displayed a conditional sentence on the top of the screen, and we invited participants to read it carefully and to try to understand it. All the sentences described the contents of a box in which there was a circle on the left and a star on the right and were of the form *If the circle is red then the star is black*. Sixteen different pairs of colors were used to form 16 different conditional statements. After a delay of 7 s, the box (a rectangle), the circle on its left part, and the star on its right part appeared in this order on the middle of the screen at the rate of one object per second. We adopted this procedure to facilitate the understanding of the sentence without any interference with a particular case. The objects were progressively displayed on screen to favor the analysis of the case under study. The match between the colors of the circle and the star displayed in the box and those involved in the conditional sentence was manipulated to obtain four trials in each of the four possible logical cases p q , $\neg p q$, $p \neg q$, and $\neg p \neg q$. For example, with the rule *If the circle is red then the star is black*, a $\neg p \neg q$ case was created by presenting a blue circle and a green star. Then, three possible responses were displayed at the bottom of the screen (“true,” “one cannot know,” and “false”). Each participant had the same response array printed on a response sheet and was asked to say if the contents of the box showed that what was announced by the sentence was either true or false, or if the content of the box did not permit one to know (i.e. the “one cannot know” response) by ticking the appropriate response. When all the participants had selected their response on the sheet, the following sentence was displayed on screen followed by the contents of the box and so on. The 16 trials were presented in two different random orders, half of the participants in each age being assigned to one of them.

Results

As our hypothesis predicted, we observed in each group a high rate of “true” responses for p q cases (99%, 99%, 97%, and 79% in third, sixth, and ninth graders and adults, respectively) and “false” responses for $p \neg q$ cases (75%, 73%, 90%, and 95%, respectively; see Table 3). Nonetheless, the age-related decrease in the former case and increase in the latter were statistically significant, $F(3, 167) = 9.23$, $p < .01$, and $F(3, 167) = 6.33$, $p < .01$, respectively, reflecting interesting developmental changes that are addressed below when analyzing response patterns. Our main prediction concerned the rate of indeterminate responses (“one cannot know”) on the *not p* cases. The developmental trend con-

Table 3
Distribution of Responses in Percent as a Function of Age Group and Type of Case in the Truth-Testing Task

		Age group											
		Grade 3			Grade 6			Grade 9			Adults		
Cases		T	I	F	T	I	F	T	I	F	T	I	F
p	q	99	0	1	99	1	0	97	3	0	79	21	0
$\neg p$	$\neg q$	0	14	86	0	34	66	1	70	29	1	90	10
$\neg p$	q	1	30	70	1	32	68	1	38	61	0	82	18
p	$\neg q$	0	25	75	1	26	73	1	8	90	0	5	95

Note. T = true; I = indeterminate; F = false.

formed to our hypothesis. In the first levels (third and sixth grades), $\neg p \neg q$ as well as $\neg p q$ cases mainly elicited "false" responses (86% and 70%, respectively, in third grade, 66% and 68%, respectively, in ninth grade). The difference was significant in third graders, with a higher number of "false" responses in $\neg p \neg q$ cases, $t(46) = 2.07, p < .05$, and disappeared in sixth graders, $t(37) = 0.12, p = .91$. The following pattern analyses shed light on the difference observed in third graders. In a second phase (ninth grade), the $\neg p q$ case was still considered as making the conditional false, with only 38% of indeterminate but 61% of "false" responses, whereas the $\neg p \neg q$ case predominantly induced the predicted indeterminate response (70%), $t(38) = 3.78, p < .001$. Finally, adults exhibited a high rate of indeterminate responses for both types of *not p* cases (90% and 82% for $\neg p \neg q$ and $\neg p q$, respectively), though the difference remained significant, $t(46) = 2.09, p < .05$. An analysis of variance with the type of *not p* case as a within-subject and the age group as a between-subject factor confirmed the effects of age, $F(3, 167) = 61.66, p < .001$, and the interaction between ages and type of *not p* cases on the number of indeterminate responses, $F(3, 167) = 5.95, p < .01$, whereas there was no significant effect of the type of case, $F(1, 167) = 2.51, p = .11$. Accordingly, the rate of "false" responses on *not p* cases decreased with age (78%, 67%, 45%, and 14% in third, sixth, and ninth graders and adults, respectively), following the predicted linear trend, $F(1, 167) = 175.36, p < .001$.

We analyzed the individual response patterns by assuming that a given participant was consistent when he or she gave at least three out of four identical responses for each of the four types of cases. We distinguished between expected conjunctive ("true," "false," "false," and "false" coherent responses to $p q, \neg p \neg q, \neg p q$, and $p \neg q$ cases, respectively), biconditional ("true," indeterminate, "false," "false"), and conditional ("true," indeterminate, indeterminate, "false") patterns. Two additional patterns were frequently produced. The first, frequent in young participants, consisted in responding "true" for $p q$, "false" for $\neg p \neg q$, and indeterminate for the other cases. We suggest that this pattern reflects a matching strategy in which the conditional is judged as true when the contents of the box match the two colors evoked in the sentence, indeterminate when only one of the expected colors appears, and false when there is no match at all. The second pattern, unique to adults, consisted in responding "false" for $p \neg q$ and indeterminate for all the other cases. This pattern could reflect an elaborated understanding of conditional in which some adults

realize that positive instances cannot make the conditional definitely true whereas contradictory cases make it false. Our procedure classified 91%, 89%, 77%, and 83% of the third, sixth, and ninth graders and adults, respectively, as consistent in one of the five patterns.

As we predicted, conjunctive patterns reflecting the failure to flesh out the initial model $p q$ were predominant in children and tended to disappear progressively with age, whereas the defective truth table response, reflecting the three-model construction, never appeared in the youngest participants but was the predominant response in adults (see Figure 2). In line with our previous observations when studying reasoning about possibilities, the biconditional interpretation constitutes an intermediary developmental level favored by adolescents, its rate increasing and then decreasing with age. The developmental trend concerning these three interpretations in the truth-testing task perfectly matches that previously observed in reasoning about possibilities (see Figure 1). Whereas the matching pattern was mainly observed in the youngest participants, the elaborated interpretation was adopted by a small but not negligible proportion of adults who could correspond to those choosing the logical p and *not q* responses in the abstract version of Wason's (1966) selection task. The matching patterns in third graders explain why they produced more "false" responses on $\neg p \neg q$ than on $\neg p q$ cases, these latter cases eliciting the indeterminate response within the matching pattern. In the same way, the progressive disappearance of these patterns explains why the rate of "false" responses on $p \neg q$ cases increased with age, whereas the emergence of elaborated patterns sheds light on the surprising decline of "true" responses to $p q$ cases in adults. It should also be noted that the increasing rate of "false" responses on $p \neg q$ cases with age rules out the alternative explanation of our results according to which the developmental pattern we observed could result from a greater propensity in young children to respond "false."

In summary, the results confirm our predictions and make clear that the evaluation of conditionals varies in accordance with the set of models individuals can construct. The similarity in developmental trends between the identification of compatible and incompatible cases on the one hand and the evaluation of conditionals on the other supports the hypothesis that this evaluation can be based on the same representations individuals use to reason about possibilities. The response of indeterminacy on *not p* cases and the ensuing defective truth table are not ubiquitous phenomena but a developmental achievement depending on the nature of mental models constructed by children, adolescents, and then adults. This developmental trend explains in turn why adults with higher abilities are more prone to produce defective tables in truth-testing tasks and suppositional responses on probability evaluation tasks as Evans et al. (2007) observed.

Mental Models and the Suppositional Theory

We have demonstrated that a modified mental model theory is able to account for the key findings that have been considered as inconsistent with the standard theory. It is worth noting that this modification is not an ad hoc addendum for coping with new and embarrassing data (the phenomena related to the defective table have been known for a long time; see Johnson-Laird & Tagart, 1969), but an extension of existing theoretical proposals such as

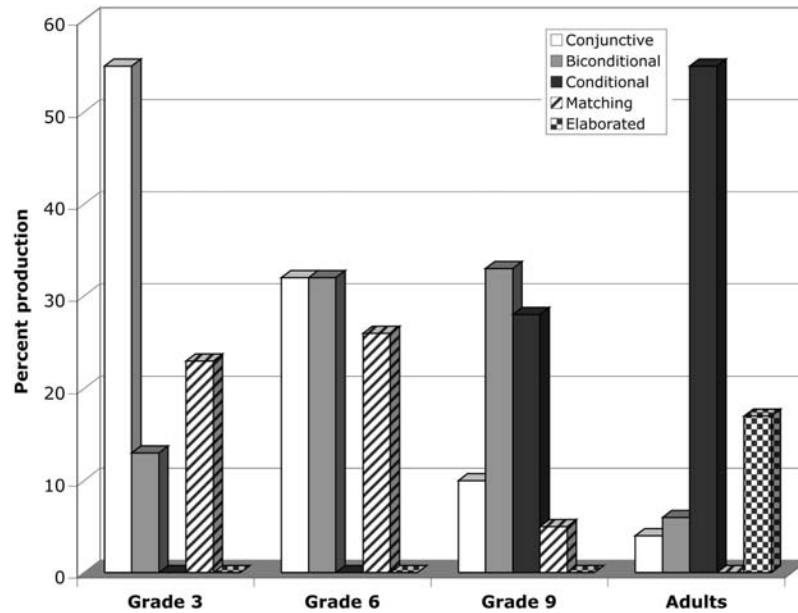


Figure 2. Distribution of response (T: true; I: indeterminate; F: false) pattern in the truth-testing task as a function of groups. For the $p \rightarrow q$, $\neg p \rightarrow \neg q$, $\neg p \rightarrow q$, and $p \rightarrow \neg q$ cases, respectively, the conjunctive pattern refers to consistent T F F F, the biconditional pattern to T I F F, and the conditional pattern to T I I F responses. The matching pattern corresponds to T F I I and the elaborated pattern to I I I F responses.

the distinction between initial and fleshed-out models. Most importantly, this modified theory permits new predictions that have been verified. Though we only introduced mere addenda to the principles of truth and implicit models of the standard theory, these changes have considerable consequences for a mental model theory of conditional reasoning.

By dissociating possibilities and truth values, assuming that a complete representation of the conditional involves three possibilities no longer implies rejecting the defective truth table and endorsing the paradoxes of the material implication as valid inferences, as Johnson-Laird and Byrne (2002) did. The conditional we propose is true for $p \rightarrow q$, false for $p \rightarrow \neg q$, but has no truth value for *not* p cases (at least in most of the adults) and presents thus a truth value gap as in the T3 theory family described by Evans and Over (2004). As a consequence, it is akin to the conditional described by Evans and Over in their suppositional account. Indeed, the meaning of a basic conditional would no longer be captured by the three possibilities it allows, as Johnson-Laird and Byrne assumed. In our account, this meaning would also incorporate the epistemic status of each of these possibilities. Recall that the mental model theory distinguishes between meaning and models. We assume that this epistemic status could be coded at the level of the semantics of the connective (i.e., the procedures that construct models, making a distinction between initial models that make the sentence true and additional models that are only compatible with it). It is this epistemic status that would differentiate *If p then q* from its contrapositive *If not q then not p* or the disjunction *not p or q*, even if the three sentences allow the same set of models.

This account predicts, as we observed, the defective truth table response in truth-testing tasks and the conditional and conjunctive probability responses when reasoners evaluate the probability of the conditional. It can be also assumed that, when evaluating the

conditional, individuals would use the Ramsey test as assumed by the suppositional account of the conditional. According to Evans and Over (2004), when evaluating a conditional *If p then q*, people hypothetically suppose p and run a mental simulation in which they evaluate q by making subjective probability judgments about q given p . This is not so different from Johnson-Laird and Byrne (2002, p. 649), who assumed that “the antecedent refers to a possibility, and the consequent is interpreted in that context.” However, assuming that $p \rightarrow q$ is the sole model making the conditional true permits us to understand why individuals concentrate on a context that satisfies the antecedent. The Ramsey test within our account would not differ from Evans and Over’s conception, that is, people trying to evaluate the relative likelihood of states of affairs making the conditional true (i.e., $p \rightarrow q$) or false (i.e., $p \rightarrow \neg q$). Thus, there is no contradiction in assuming that the basic conditionals are suppositional in nature but represented by a set of three models when people think about what they refer to.

Although our modified mental model theory constitutes an important rapprochement between the mental model and the suppositional accounts of conditional, there are still important differences with Evans and Over’s (2004) theory that we would like to comment on. Though the suppositional theory mainly focuses on reasoning about truth values, it also addresses the problem of conditional inference. Concerning this last point, Evans and Over did not adopt the view of an innate mental logic but endorsed a conception previously developed by Braine and O’Brien (1991) and Rips (1994) in which modus tollens (from *If p then q* and *not q* concluding *not p*) is underpinned by a suppositional strategy through reductio ad absurdum, whereas the fallacies affirmation of the consequent (from q concluding p) and denial of the antecedent (from *not p* concluding *not q*) result from implicatures and pragmatic inferences. However, we have elsewhere demonstrated that

suppositional theories such as those of Marcus and Rips (1979) and Braine and O'Brien make predictions about the development of conditional reasoning and the time course of conditional inferences in adults that are contradicted by empirical evidence (Barrouillet et al., 2000; Barrouillet & Lecas, 1998). By contrast, one interest of our account is to retain the strength of the mental model approach in accounting for reasoning about possibilities. Indeed, the mechanism of inference production within the standard mental model theory is not committed to any assumption concerning truth values, a characteristic inherited by our account. For example, a reasoner who has constructed the complete three-model representation and is given the minor premise of modus tollens (i.e., *not q*) will eliminate contradictory possibilities (*p q* and *not p q*). The remaining model $\neg p \neg q$ is the only possibility compatible with the conditional in which the minor premise is true. Thus, even if *not p not q* does not make the conditional true when it occurs, the conclusion *not p* can be drawn with certainty. The same line of reasoning applies to the fallacies that are endorsed when people fail to construct a complete representation.

Objections and Replies

Contrary to our account of modus tollens, it could be objected that if false-antecedent possibilities do not make a conditional true, modus tollens could no longer be drawn in the mental model theory because the inferential process leads to the remaining mental model $\neg p \neg q$ in which the minor premise is still true but the conditional premise is not. As a consequence, the inferential process described by the algorithmic level of the theory would not comply with truth preservation, which is the core of logical inferences, and the mental model theory could no longer be interpreted as performing inferences since it would not conform to a consistent theory of inference at the computational level.⁴

It is important to remember that, contrary to a recurrent claim, the mental model theory at the computational level does not characterize deductive competence as standard formal logic and that it computes not truth functions but semantic information. Within this framework, a valid conclusion is a conclusion (a) that preserves the semantic information conveyed by the truth of the premises without increasing or decreasing it, (b) that is more parsimonious than premises, and (c) that asserts something new (Johnson-Laird & Byrne, 1991). The mental model theory defines semantic information not in reference to truth preservation but only as the set of models of the possibilities compatible with a given proposition, a conclusion being valid only if it holds in all the models of the premises. Now, the changes we have introduced into the mental model theory leave these postulates unchanged. The meaning of *if* in the standard mental model theory is the set of procedures constructing the mental models. Adding, as we have suggested, information concerning the epistemic status of each model within this semantics does not affect the set of models required by a modus tollens inference. This inference is still valid at the computational level of the theory because deduction calls for a search for potentially falsifying models,⁵ and there is no such model compatible with the semantics of *if* in both the standard and our modified theory of mental models.

A second objection is that the distinction between reasoning about possibilities and reasoning about truth values is untenable because concluding *A* and concluding *A is true* are one and the

same thing. Thus, reasoning about possibilities could not be distinguished from reasoning about truth values.⁶

It is true that the conclusions *A* and *A is true* cannot be logically distinguished. However, the distinction that we propose is not at the logical but at the psychological level and stands on the distinction made by the mental model theory between meaning and models. By reasoning about possibilities, we refer to those forms of reasoning requiring a simple scanning of the content of the models maintained in working memory. This is the case for the conditional inference task in which the conclusion to be evaluated is always an atomic proposition most often represented by a single token within a unique model. By reasoning about the truth value of the conditional, we refer to those forms of reasoning in which the current content of the available representations has to be confronted with the meaning of the conditional connective. These two forms of reasoning involve different cognitive processes reflecting the difference between declarative and procedural knowledge as defined by Anderson (1993). Mental models involve declarative knowledge represented in working memory, whereas the meaning of the conditional can be understood as a set of procedures constructing models (Johnson-Laird & Byrne, 1991). What is directly accessible is the content of the models, whereas procedural knowledge is more difficult to represent in working memory (Anderson, 1993). Thus, scanning representations constituting the content of our consciousness to verify atomic propositions (i.e., reasoning about possibilities) is psychologically basic, whereas going back from these representations to their presumed constructive procedures (i.e., reasoning about truth values) is far more difficult, particularly when the meaning to be assessed involves a connective calling for several models, as is the case for the conditional.

A third objection to the suppositional theory of conditionals that could also be addressed to our proposal has been raised by Johnson-Laird (2005), who claimed that T3 theories have the extraordinary consequence of turning all true conditionals into conjunctions. Because *p q* is the sole case making a basic conditional true, if such a conditional is true, its antecedent and its consequent are true. Thus, considering the following conditional as true—*If the U.S. Constitution is amended to allow people born elsewhere to be president, then Arnold Schwarzenegger will be delighted*—should lead to the certain conclusion that the U.S. Constitution is amended and Arnold Schwarzenegger is delighted, which is, as Johnson-Laird noted, absurd.

Though this criticism is clever, the distinction between reasoning about possibilities and about truth values permits us to solve this potential problem. To imagine what can occur from a true conditional, as Johnson-Laird (2005) proposed with the Schwarzenegger problem, is reasoning about possibilities, and we assume,

⁴ We would like to thank Mike Oaksford for making us aware of this potential problem.

⁵ The test of validity involves constructing a model in which the negation of the putative conclusion is true and assessing whether any of the previous premises are false in this view. If this is the case, the conclusion is valid. As far as modus tollens is concerned, the resulting model is $\neg p \neg q$ in which $\neg p$ is new information standing as a putative conclusion. Introducing its negation leads to the model *p* $\neg q$ that falsifies the conditional premise. The conclusion $\neg p$ is therefore valid.

⁶ We would like to thank an anonymous reviewer for this suggestion.

following the mental model theory, that a true conditional does not lead to one but to three possibilities. Thus, a true conditional involves not the truth but only the possibility of the conjunction of its antecedent and consequent, except for young children, who are not able to flesh out their initial model.

Finally, we have not addressed here the nature of the mental models on which the inferences are based. Evans and Over (2004), as well as Barrouillet and Lecas (1998) and Markovits and Barrouillet (2002), assumed that models are relational, have directionality, and bring to bear linguistic and pragmatic experience. When these characteristics are introduced, conditional reasoning by models would no longer be extensional. However, these problems, though interesting, are outside the scope of this note.

Conclusions

Different theorists privilege selected sets of data to support their claims. Mental modelers usually privilege tasks involving reasoning about possibilities, whereas recent advocates of suppositional accounts prefer tasks involving reasoning about truth value. Our proposal can be seen as an attempt to integrate the two realms in a comprehensive approach by producing a theory able to explain how individuals extend their beliefs through deductive reasoning from premises they consider as true and how they decide to believe assertions. Of course, as stated by Evans et al. (2005), how people decide to what extent to believe ordinary conditionals is a fundamental question. Another fundamental question of equal importance is what kind of world representation people construct when they have decided to believe an ordinary conditional. Our suggestion is that a psychological theory must address the two questions within the same theoretical framework and that our modified mental model theory meets this challenge. According to this theory, people understand conditionals by representing the state of affairs that can occur when these sentences are true. Because only those possibilities that make the sentence true when they occur are initially explicitly represented, the initial model of a basic conditional is restricted to the $p q$ model. The other possibilities in which the antecedent is not satisfied can be made explicit through fleshing out, but they leave the truth value of the conditional indeterminate. Thus, when evaluating a conditional, people focus on only those possibilities in which the antecedent is satisfied.

Several studies have indicated that adults and also older adolescents judge the possibilities $p q$, $\neg p q$, and $\neg p \neg q$ as compatible and the $p \neg q$ possibility as incompatible with basic conditionals. It is highly probable that this interpretation of *if* underpins the material implication account that some philosophical logicians have proposed for conditionals. When considering developmental studies on reasoning about possibilities and the developmental trend they reveal toward the three-possibility representation, it becomes obvious that the material implication grasps something of the way individuals understand conditionals, the unmodulated meaning of a conditional yielding models of possibilities corresponding to a material implication. This does not mean that these individuals are ready to endorse all the consequences that result from this logical formalism: Very few people would consider that *not p* cases make the conditional true. This is not so surprising: Because logical formalisms are formalisms, they can neither be totally at odds nor perfectly reflect psychological intuitions (Schroyens, Schaeken, & Dieussaert, 2008). This is clearly dem-

onstrated by studies in which people are no longer asked to reason from but to evaluate conditionals. In these tasks, adults base their judgment on p cases only, a tendency very akin to the procedure known as the Ramsey test (hypothetically adding p to our stock of knowledge and arguing on that basis about q). There is no doubt that this bias has led many logicians to reject the theory that ordinary conditionals of natural language are extensional or material implications. It is clear that the suppositional account of conditionals put forward by Evans and Over (2004) grasps something of the way people understand and evaluate conditionals. In our view, the conflicting coexistence of model-based and suppositional accounts of conditional reasoning results from the fact that reasoning about possibilities and reasoning about truth values involve different processes that lead to apparent contradictions (e.g., all the possibilities compatible with a basic conditional do not lead to the same evaluation of this conditional when they occur). Our analysis suggests that the two kinds of reasoning involve different processes but the same representations and that a theory based on mental models can account for both.

References

- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- Barrouillet, P. (1997). Modifying the representation of if . . . then sentences in adolescents by inducing a structure mapping strategy. *Current Psychology of Cognition*, 16, 609–637.
- Barrouillet, P., Grosset, N., & Lecas, J. F. (2000). Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition*, 75, 237–266.
- Barrouillet, P., & Lecas, J. F. (1998). How can mental models account for content effects in conditional reasoning? A developmental perspective. *Cognition*, 67, 209–253.
- Barrouillet, P., & Lecas, J. F. (1999). Mental models in conditional reasoning and working memory. *Thinking and Reasoning*, 5, 289–302.
- Barrouillet, P., & Lecas, J. F. (2002). Content and context effects in children's and adults' conditional reasoning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 55(A), 839–854.
- Braine, M. D. S., & O'Brien, D. P. (1991). A theory of *if*: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, 98, 182–203.
- Evans, J. St. B. T. (1972). Interpretation and matching bias in a reasoning task. *British Journal of Psychology*, 24, 193–199.
- Evans, J. St. B. T. (1983). Linguistic determinants of bias in conditional reasoning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 35(A), 635–644.
- Evans, J. St. B. T., Clibbens, J., & Rood, B. (1996). The role of implicit and explicit negation in conditional reasoning bias. *Journal of Memory and Language*, 35, 392–409.
- Evans, J. St. B. T., Handley, S. J., Neilens, H., & Over, D. E. (2007). Thinking about conditionals: A study of individual differences. *Memory & Cognition*, 35, 1785–1800.
- Evans, J. St. B. T., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 321–355.
- Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hillsdale, NJ: Erlbaum.
- Evans, J. St. B. T., & Over, D. E. (2004). *IF*. Oxford, England: Oxford University Press.
- Evans, J. St. B. T., Over, D. E., & Handley, S. J. (2005). Suppositions, extensionality, and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002). *Psychological Review*, 112, 1040–1052.

- Giroto, V., & Johnson-Laird, P. N. (2004). The probability of conditionals. *Psychologia*, 47, 207–225.
- Giroto, V., Mazzocco, A., & Tasso, A. (1997). The effect of premise order in conditional reasoning: A test of the mental model theory. *Cognition*, 63, 1–28.
- Johnson-Laird, P. N. (2005). If bears eat in the woods . . . ? *Trends in Cognitive Sciences*, 9, 43–44.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.
- Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, 109, 646–678.
- Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review*, 3, 418–439.
- Johnson-Laird, P. N., & Savary, F. (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition*, 71, 191–229.
- Johnson-Laird, P. N., & Tagart, J. (1969). How implication is understood. *American Journal of Psychology*, 2, 367–373.
- Kuhn, D. (1977). Conditional reasoning in children. *Developmental Psychology*, 13, 342–353.
- Lecas, J. F., & Barrouillet, P. (1999). Understanding of conditional rules in childhood and adolescence: A mental models approach. *Current Psychology of Cognition*, 18, 363–396.
- Marcus, S. L., & Rips, L. J. (1979). Conditional reasoning. *Journal of Verbal Learning and Verbal Behavior*, 18, 199–223.
- Markovits, H., & Barrouillet, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review*, 22, 5–36.
- Moshman, D. (1990). The development of metalogical understanding. In W. F. Overton (Ed.), *Reasoning, necessity, and logic: Developmental perspectives* (pp. 205–226). Hillsdale, NJ: Erlbaum.
- Oberauer, K., & Wilhelm, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models, and personal utilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 680–693.
- Paris, S. G. (1973). Comprehension of language connectives and propositional logical relationships. *Journal of Experimental Child Psychology*, 16, 278–291.
- Quine, W. V. O. (1952). *Methods of logic*. London: Routledge.
- Rips, L. J. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.
- Schroyens, W., & Schaeken, W. (2003). A critique of Oaksford, Chater, and Larkin's (2000) conditional probability model of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 140–149.
- Schroyens, W., & Schaeken, W. (2004). Guilt by association: On iffy propositions and the proper treatment of mental-models theory. *Current Psychology Letters: Behaviour, Brain & Cognition*, 12(1). Retrieved from <http://cpl.revues.org/document411.html>
- Schroyens, W., Schaeken, W., & Dieussaert, K. (2008). "The" interpretation(s) of conditionals. *Experimental Psychology*, 55, 173–181.
- Schroyens, W., Schaeken, W., & Handley, S. (2003). In search of counterexamples: Deductive rationality in human reasoning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 56(A), 1129–1145.
- Sevenants, A., Schroyens, W., Dieussaert, K., & Schaeken, W. (2006). The truth about possibilities-based and truth-based truth table tasks. In *Proceedings of 28th Annual Conference of the Cognitive Science Society* (pp. 2141–2145). Mahwah, NJ: Erlbaum.
- Taplin, J. E., Staudenmayer, H., & Taddonio, J. L. (1974). Developmental changes in conditional reasoning: Linguistic or logical? *Journal of Experimental Child Psychology*, 17, 360–373.
- Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (pp. 106–137). Harmondsworth, England: Penguin.
- Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.

Received July 28, 2006

Revision received November 16, 2007

Accepted November 24, 2007 ■

Postscript: A Good Psychological Theory of Reasoning Must Predict Behavior and Explain the Data

Pierre Barrouillet and Caroline Gauffroy
Université de Genève

Jean-François Lecas
Université de Bourgogne

In their stimulating comment on Barrouillet, Gauffroy, and Lecas (2008), Oberauer and Oaksford (2008) argued that we gave no rationale for the assumption that the mental models of false-antecedent cases have a different epistemic status from the $p \rightarrow q$ model and that no new computational level account of the conditional was provided to justify this move. The resulting lack of coherent principles would undermine the heuristic power of the mental model theory and lead to unacceptable consequences. Though our theory probably presents many weaknesses, it is false to assume that we did not give any rationale for the truth value gap of conditionals. Our proposal elaborates the distinction between initial and fleshed-out models that is at the core of the theory from its very beginning (Johnson-Laird & Byrne, 1991), and assumes that people consider those cases that match the initial model as making the conditional true whereas the cases that match the models added through fleshing out are judged as leaving its truth value indeterminate. We argue that it is highly rational to behave in this way. The initial model is what comes

immediately to mind when understanding a sentence. It is produced by relatively automatic and effortless processes that correspond to the heuristic system in Evans's (2006, 2008) heuristic-analytic theory. According to Evans, these heuristic processes deliver an initial model that is the most plausible and relevant. Thus, it is rational that people regard this model as the core meaning of the conditional and consider those cases that match it as making the sentence true. By contrast, additional models necessitate a fleshing-out process that relies on an analytic system, which is optional, demanding, and therefore error prone. It is also rational to consider that these models are compatible with the meaning of the sentence but do not make it true because they do not match the representation that immediately comes to mind. Once more, this proposal does not necessitate any modification of the computational level of the mental models theory. Truth values are not represented at the algorithmic level nor are truth functions computed at the computational level.

The supposed first undesirable consequence of this view would concern the negated conditionals. According to the probabilistic interpretation of conditionals, the negation of *If p then q* is *If p then not q*, and the probability of the negated conditional is $P(\neg q/p) = 1 - P(q/p)$. Handley, Evans, and Thompson (2006) added that what is important is that our degree of belief in the former would be complementary to our degree of belief in the latter. Oberauer and Oaksford (2008) noted that our theory has no basis for treating the negation of conditional in