



Conditional reasoning by mental models: chronometric and developmental evidence

Pierre Barrouillet*, Nelly Grosset, Jean-François Lecas

*Université de Bourgogne, LEAD–CNRS, Faculté des Sciences Gabriel, 6 Bld Gabriel,
21000 Dijon, France*

Received 11 June 1999; received in revised form 14 September 1999; accepted 2 February 2000

Abstract

The aim of this article is to verify two predictions resulting from the mental models theory of conditional reasoning. First, the denial of antecedent (DA) and modus tollens (MT) inferences should take longer to verify than modus ponens (MP) and affirmation of consequent (AC) because the former require subjects to flesh out the initial model whereas the latter do not. This prediction was confirmed in two reaction time experiments in adults. In line with Evans' proposal (Evans, J. St. B. T. (1993). The mental model theory of conditional reasoning: critical appraisal and revision. *Cognition*, 48, 1–20), there was a strong directionality effect: inferences from antecedent to consequent (MP and DA) took less time to verify than the inferences in the opposite direction (AC and MT). Second, the development of conditional reasoning should result from the increasing capacity to construct and coordinate more and more models. As a consequence, the pattern of conditional inference production should evolve with age from a one-model conjunctive pattern (production of MP and AC more frequent than DA and MT) to a three-model conditional production pattern (higher production rate for MP and MT than for DA and AC). This prediction was confirmed using an inference production task in children, adolescents, and adults. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Conditional reasoning; Mental models; Cognitive development

1. Introduction

It has recently been suggested that it is not possible to come to an empirical choice between the two main theories of human reasoning (i.e. Johnson-Laird and Byrne's

* Corresponding author. Tel.: +33-3-80-39-57-88; fax: +33-3-80-39-57-67.

E-mail address: barouil@satie.u-bourgogne.fr (P. Barrouillet)

mental models and mental logic theories) because both of them lack sufficient precision (Evans & Over, 1997). This article presents three experiments which aimed at verifying predictions resulting simply from the mental models theory (Johnson-Laird & Byrne, 1991; Johnson-Laird, Byrne & Schaeken, 1992, 1994) for the processing of *if then* conditional statements. The first two experiments compared reaction times in order to endorse the conclusions drawn from different conditional syllogisms in adults. The third experiment adopted a developmental approach and studied age-related development in the frequency of conditional inference production from childhood to adulthood.

The mental models theory postulates that deductive reasoning is a three-step process. The first step is thought to consist of a comprehension process which leads to the construction of a representation (a mental model) of the state of affairs the premises describe. According to the theory, this mental model would have the same logical structure as the described state of affairs. When there are several premises to be considered, and thus several mental models to be coordinated, the theory postulates coordination heuristics which preserve the information provided by the premises. The second step consists of the production of a putative conclusion from the resulting model. A given item of information would be considered to be a conclusion by the subject if it is both true in all the models constructed and not explicit in the premises. These first two steps are simply a matter of comprehension and do not embody any reasoning process per se. In contrast, in the third step, the subject would try to falsify the reached conclusion by constructing alternative models of the premises. If there is no alternative model in which this conclusion does not hold, then it is necessarily true, but if there is such a model, the conclusion must be judged to be uncertain.

The main prediction of the theory is that the greater the number of models a given reasoning requires, the harder it should be (i.e. prone to error, experienced as difficult by the subject, and taking a long time), because both the construction and manipulation of mental models would be carried out in a capacity-limited working memory. Thus, subjects would tend to reduce the cognitive load resulting from constructing and coordinating models by making some part of the information to be represented implicit. They would then construct simplified models, called initial models. If required by the current situation, the initial model could be fleshed out to provide a complete representation of the situation described.

These proposals can be illustrated in the field of the conditional reasoning and the deductive inferences permitted by *if then* statements. The mental models theory proposes that an *if p then q* premise would be represented by the following initial model

$$p \quad \dots \quad q \quad (1)$$

where each line represents a distinct model. The first line in the diagram denotes an explicit model of the situation in which the antecedent and the consequent are true. The three dots denote an implicit model standing for other possibilities that make the conditional true. This implicit model would contain a mental footnote indicating that it refers to possibilities in which the antecedent of the conditional is false. Mental

footnotes can be ephemeral and soon forgotten (Johnson-Laird, Legrenzi, Girotto, Legrenzi & Caverni, 1999). If the footnote is retained, then reasoners can, in principle, flesh out the initial model, leading either to a biconditional

$$\begin{array}{l} p \quad q \\ \neg p \quad \neg q \end{array} \quad (2)$$

or a conditional interpretation

$$\begin{array}{l} p \quad q \\ \neg p \quad \neg q \\ \neg p \quad q \end{array} \quad (3)$$

where \neg is a propositional-like tag which denotes negation. Thus, the set of mental models for *if p then q* represents all the possible situations in which the statement is true.

In a conditional syllogism task, the statement *if p then q* constitutes the major premise, with which four minor premises resulting either from the affirmation or the negation of the antecedent or the consequent can be coordinated. In formal logic, only the affirmation of antecedent (*p*), known as modus ponens (MP), and the negation of consequent (*not q*), known as modus tollens (MT), lead to certain conclusions which are *q* and *not p*, respectively. On the other hand, neither the affirmation of consequent (*q*, AC) nor the denial of antecedent (*not p*, DA) lead to a firm conclusion, although the subjects frequently endorse or produce the conclusions *p* from the minor premise *q* (for AC) and *not q* from *not p* (for DA, see Evans, Newstead & Byrne, 1993, for a review).

Thus, the endorsement or the production of both MP (deduce *q* from *p*) and AC (deduce *p* from *q*) would be possible from the initial model of *if p then q*, because it contains explicit representations of both *p* and *q*. The subject would then consider the information this model links with the information given by the minor premise to be the conclusion (i.e. *q* for MP, and *p* for AC). Note that *q* is a correct conclusion for MP because even within a complete (i.e. fleshed out) representation, there is only one model representing *p*, and this model links *q* with *p*, whereas the conclusion *p* for AC should be rejected because it does not hold in a complete three-model representation (i.e. there is a model linking *q* with *not p*).

According to Johnson-Laird (1993), the subjects endorse conclusions that hold in only some models because they fail to construct all the possible models. They are deductive satisficers, and when they come up with one conclusion that fits their available beliefs, they would not tend to search for others. The denial inferences DA (deduce *not q* from *not p*) and MT (deduce *not p* from *not q*) require subjects to flesh out the initial model because it does not contain explicit representations of *not p* and *not q*. Whether the resulting interpretation is biconditional (two models) or conditional (three models), the conclusion *not p* necessarily follows from *not q* (MT). If only two models are constructed (biconditional interpretation), the minor premise

not p brings about the conclusion *not q* (DA), but this conclusion does not hold in a complete conditional representation.

Thus, the endorsement or the production of MP and AC would require only one model whereas MT and DA would require two models. The mental models theory therefore predicts that among the logical inferences MP and MT, MP should be easier to produce than MT because the latter requires a fleshing out process. Of the fallacious inferences AC and DA, AC should be produced more often than DA because the latter requires two models whereas the former is supported by the initial model of the conditional. However, empirical data contradict the prediction of a more frequent endorsement and production of AC than DA. When a wide range of experiments are considered, it appears that AC and DA are endorsed and produced with equal frequencies (Evans, 1993; Evans et al., 1993). O'Brien, Braine and Yang (1994) have used this fact to argue against the mental models theory and have claimed that the mental logic has no problem in accounting for it.

The theories relating to the mental logic (Braine, 1990; Braine & O'Brien, 1991; Braine & Romain, 1983; Rips, 1994) propose that deductions are permitted by a set of inference schemas (formal rules of inference) stored in long term memory. These rules would apply to propositional representations resulting from a comprehension process of the premises. As far as *if* is concerned, these theories (e.g. Braine & O'Brien, 1991) presuppose that two rules are available. The first would correspond to MP: when two propositions of the form *if p then q* and *p*, respectively, are available, the rule derives a conclusion of the form *q*. The second is aimed to introduce *if* by making it possible to state that *if p then q* for any proposition *q* which can be derived from a supposition *p*. Such a model predicts, as does the mental models theory, that MT should be more difficult than MP. The latter would be directly achieved by applying an available schema while the former requires a reasoning strategy by *reductio ad absurdum*. As far as DA and AC are concerned, these inferences would result from pragmatic principles or invited inferences. There is no reason to suppose one of them to be more difficult than the other. That is probably the reason why O'Brien et al. (1994) have claimed that mental logic has no difficulty in accounting for their equal frequency of production.

Evans (1993) has suggested modifying the mental models theory in order to make it compatible with this phenomenon and has suggested two explanatory factors. The first would be a negative conclusion bias leading to a more frequent acceptance of negative than affirmative conclusions. Now, this bias favors DA acceptance (negative conclusion *not q*) and impedes AC (affirmative conclusion *p*) when both the antecedent and the consequent of the major premise *if p then q* are affirmative propositions. The second factor would be that the models of the conditional are directional in nature and are oriented from the antecedent to the consequent. According to this proposal, MP and DA (minor premises *p* and *not p*, respectively) would be forward inferences and hence easier to endorse than the corresponding backward inferences AC (for the first model, minor premise *q*) and MT (for the second model, minor premise *not q*, see Ormerod, Manktelow & Jones, 1993, for an account of the directionality of models resulting from *if p then q* conditionals). Thus, both the negative conclusion bias and the directionality of models would favor DA (which

is a forward inference leading to a negative conclusion), and impede AC (which is backward inference with a positive conclusion), thus leading to equal frequencies of endorsement although the former requires one more model than the latter.

Two comments could be added to the Evans (1993) proposals. First of all, the mental models theory permits predictions not only of the frequency of acceptance of the different inferences as a function of the number of models they require, but also of the time needed for these acceptances. Thus, according to the theory, the inferences which require one model only (MP and AC) should be faster than those which require two models (DA and MT), because it should take longer to construct two models by means of the fleshing out process compared to one model provided by the initial model of the conditional. Marcus and Rips (1979) have already studied conditional syllogisms using RT measures. However, they used only three different conditional premises which elicited major differences in the rates of endorsement of inferences. Second, it is possible that the number of models required to produce a given inference has a strong effect on its frequency of production only if the construction and maintenance of the models lead to a cognitive overload. Now, it is possible that the number of models required to endorse DA (i.e. two models) is not high enough to exhaust the cognitive resources of adult subjects. Thus, the effect of the higher cognitive load induced by DA compared to AC (two models rather than one) might be superseded by the opposite effects of both the negative conclusion bias and the directionality of models.

However, the cognitive load effect should be all the stronger the lower the subject's cognitive resources. For example, many authors have suggested that children have lower working memory capacity than adults (Case, 1985, 1992; Halford, 1993; Pascual Leone, 1988). If this is the case, young children should produce MP and AC more often than DA and MT because of their limited working memory capacity.

The general hypothesis that the process of inference production is constrained by the number of models required has been tested in three experiments. The first two experiments used a reaction times measure paradigm in which adult subjects were asked to evaluate conditional syllogisms. We hypothesized that the endorsement of a given inference should take all the longer the higher the number of models it requires. The third experiment used a developmental approach and the fact the number of models the children can construct and use for reasoning increases with age (Barrouillet, 1997; Barrouillet & Lecas, 1998, 1999; Lecas & Barrouillet, 1999). We hypothesized that the production rates of the four conditional syllogisms should evolve with age according to the number of models they require.

In Experiment 1, we asked adult subjects to judge conditional syllogisms presented on screen. For each syllogism, a first screen presented a conditional premise of the form *if p then q* preceded by a short text. According to the mental models theory, the interpretation of this conditional premise would lead to the construction of an initial model as in Eq. (1). A second screen presented two propositions (a minor premise and a conclusion) of the form *p* and *q*, *q* and *p*, *not p* and *not q*, and *not q* and *not p* for MP, AC, DA, and MT, respectively. The participants had to judge the validity of the conclusion in light of both the condi-

tional and the minor premises. Reaction times (RTs) for the acceptance of each type of inference were recorded. According to the mental models theory, RTs for MP and AC endorsement, which require one model, should be shorter than those for DA and MT, which make it necessary to flesh out the initial model. The Evans (1993) proposed directionality of models predicts that, among the initial-model inferences, MP should be faster than AC, and that among the second-model inferences, DA should be faster than MT, because MP and DA are forward inferences from antecedent to consequent whereas AC and MT are backward inferences from consequent to antecedent.

The Braine and O'Brien (1991) model does not predict this pattern of results. Although this model predicts longer RTs for MT than for MP, it does not predict longer RTs for DA than for AC because both inferences result from pragmatic principles and there is no reason to expect any difference in RT or frequency between them. As O'Brien et al. (1994) have claimed that mental logic has no difficulty in accounting for the equal frequencies of AC and DA production, we can suppose that mental logic predicts equal RTs in the same way.

The aim of Experiment 2 was to discount the alternative hypothesis that MP and AC took shorter than DA and MT because the latter require the processing of negative sentences. The same design as in Experiment 1 was used but both the minor premise and the conclusion were expressed using either synonyms of the antecedent and the consequent of the conditional premise for MP and AC, or antonyms for DA and MT. As for Experiment 1, we predicted the same effect of the number of models to be constructed and the same directionality effect of models.

The aim of Experiment 3 was to verify the effect of the number of models to be constructed on the frequency of inference production in childhood and adolescence. Several experiments conducted in our laboratory suggest an age-related evolution of the interpretation of *if p then q* statements, from a conjunctive-like interpretation resulting from the construction of a single model of the form $p \cdot q$, to a biconditional interpretation based on a two-model representation ($p \cdot q$ and $\neg p \cdot \neg q$), followed by an interpretation based on the complete three-model representation described by Johnson-Laird and Byrne (1991) (Barrouillet, 1997; Barrouillet & Lecas, 1998, 1999; Lecas & Barrouillet, 1999). In line with the proposals of Johnson-Laird and Byrne (1991), the growth in the number of models that children can construct and maintain depends on a related increase in working memory capacity (Barrouillet & Lecas, 1999). Accordingly, the rate of production of each type of inference should evolve as a function of the number of models the children can construct.

Modus ponens, which is based on the initial model, should be frequently and constantly produced across the three levels of interpretation previously described. Modus tollens, which is based on the second model, should be rare in children who construct one model only, and then more frequent when children are able to construct two or three models. In contrast, the frequency of production of AC should be high in children who construct one or two models and then decline when three models can be constructed, because the third model would represent instances of the form $\neg p \cdot q$ which prevent subjects from concluding p from q . Finally, the frequency

of production of DA which requires subjects to flesh out the initial model should be low in children who construct one model only, and should then increase when children are able to construct two models, and finally decline when the third model can be constructed (Table 1).

Thus, the mental models theory predicts a distinct developmental pattern for each of the four inferences: MP frequency should remain high and stable across development, MT frequency should increase with age, AC frequency should decrease, and DA frequency should first increase and then decrease. Therefore, there should be a developmental level where MP and AC will be produced more frequently than DA and MT. This gap between the one-model and the two-model inferences should disappear when children become able to construct two models. Finally, a third developmental level should correspond to the capacity to construct a complete three-model representation, where MP and MT should be more frequent than AC and DA.

Once again, this set of predictions does not follow from mental logic theories. Of course, these theories predict a high and stable rate of production for MP across development, because this inference corresponds to a schema stored in long term memory. They also predict an age-related increase in MT production because this inference would require a reasoning strategy based on hypothesis production, and one can suppose that the older the children, the smarter the implementation of this strategy would be. As we have already seen, AC and DA would result from pragmatic principles. According to Braine and O’Brien (1991), the development of the understanding of conditional sentence would be based on an increasing age-related capacity to set aside the pragmatic principles in situations which require logical treatment (e.g. laboratory tasks). If this is the case, we should therefore expect a frequent production of these inferences in young children and then a progressive decrease in the production rate of both DA and AC. The mental models theory predicts such a developmental trend for AC but it differs with regard to DA, for which it predicts an inverse U-shaped production rate curve (see above).

Table 1
Predictions about the type of inferences which should be produced (+) or not produced (–) as a function of the level of interpretation of the *if p then q* conditional premise

Minor premise	Conclusion	Inference	Level of interpretation of <i>if p then q</i> and models produced		
			Conjunctive	Biconditional	Conditional
			$p \quad q$	$p \quad q$ $\neg p \quad \neg q$	$p \quad q$ $\neg p \quad \neg q$
p	q	MP	+	+	+
q	p	AC	+	+	–
$not\ p$	$not\ q$	DA	–	+	–
$not\ q$	$not\ p$	MT	–	+	+

2. Experiment 1

The aim of his experiment was to verify that the endorsement of inferences based on the initial model of the conditional (i.e. MP and AC) is faster than that of inferences which require subjects to flesh out this initial model (i.e. DA and MT). For all the inferences, a major premise of the form *if p then q* was presented in a first screen. In order to control for a possible negative-conclusion-bias effect, all the conditional premises we used had an affirmative antecedent and consequent (e.g. if the musician is Polish, then he plays bassoon). Affirmative antecedents and consequents lead to affirmative conclusions for MP and AC (*q* and *p*, respectively) but to negative conclusions for DA and MT (*not q* and *not p*, respectively). Consequently, a possible negative conclusion bias should work against our main hypotheses by slowing down the endorsement of MP and AC but speeding up that of DA and MT. However, this control led us to use both a negative minor premise and a negative conclusion for DA and MT. It could be that processing a linguistic negation in the minor premise and conclusion led to extra time which might account for the expected difference in RTs between MP and AC on the one hand and DA and MT on the other. In order to control for this possible effect, we used either explicit or implicit negations. For example, *he is Italian* and *he is not Polish* are implicit and explicit negations of *he is Polish*, respectively. For the denial inferences DA and MT, conclusions were always expressed by an explicit negation, because MT (and maybe DA) does not hold with an implicit negative conclusion: from *if the musician is Polish, then he plays bassoon* and *he does not play bassoon*, it follows that *he is not Polish* but it does not follow that, for example, *he is Italian*. As a consequence, the two kinds of negation were used for the minor premise only, whereas the conclusion was always expressed using an explicit negation. If the expected difference in RTs between the one-model (MP and AC) and the two-model (DA and MT) inferences is due, at least in part, to the processing of an additional linguistic negation in the latter, then this difference should be greater when an explicit negation is used in the minor premise of DA and MT (DAE and MTE) than when an implicit negation is used (DAI and MTI). Furthermore, specific hypotheses could be derived from the mental models theory with regard to the type of negation embedded in minor premises.

The mental models theory claims that fleshing out is a process of constructing alternative models to be added to the initial model. Several studies have shown that the retrieval of alternative instances in long term memory has an impact on processing of the conditional (Barrouillet & Lecas, 1998; Markovits, Fleury, Quinn & Venet, 1998). Suppose that the antecedent or the consequent of a given conditional premise offers one alternative only, as is the case when the predicate of a proposition has an antonym (e.g. the man is bachelor).¹ It could be supposed that alternative instances of *bachelor* would not be represented as \neg *bachelor* but straightforwardly as *married*, because searching for alternatives to *bachelor* in long term memory

¹ This kind of conditional premise is referred to as binary in Legrenzi (1970) and Barrouillet and Lecas (1998).

would lead to the retrieval of *married*. Thus, when both the antecedent and the consequent of the conditional premise have an antonym (referred to as A conditionals for antonym), the fleshing out of the initial model would result in a first alternative model which does not contain a negation tag (\neg) but instead positive values for both antecedent and consequent (i.e. the antonyms). As a consequence, an implicit negation in the minor premise would match the content of such an alternative model, while explicit negation would not. On the other hand, when neither the antecedent nor the consequent have antonyms (referred to as N conditionals), the first alternative model would contain negated values (for antecedent and consequent) through the use of the \neg negation tag, as suggested by Johnson-Laird and Byrne (1991). Here, the explicit negation matches the content of such a model but implicit negation does not. It should be remembered that, for the denial inferences, the conclusion was always expressed with an explicit negation. Thus, a minor premise with explicit negation coupled with a negative conclusion matches the alternative model of N conditionals but not of A conditionals. A minor premise with implicit negation coupled with a negative conclusion (always explicit) partially matches the content of alternative models of both N conditionals (matching on conclusion) and A conditionals (matching on minor premise).

According to the mental models theory, conclusions are derived from the model(s) which the minor premise matches. When the reached conclusion matches the conclusion to be judged, it will be endorsed. We hypothesized that the easier this matching is (on minor premise and conclusion), the faster the endorsement of the conclusion will be. Thus, the endorsement of denial inferences (DA and MT) with an explicitly negated minor premise should be all the faster for N conditionals, because there is a complete matching between, on the one hand, the values in the alternative model ($\neg p \cdot \neg q$) and, on the other hand, the minor premise and the conclusion, but all the slower for A conditionals because neither the minor premise nor the conclusion match the content of the alternative model. In the case of denial inferences elicited by a minor premise with implicit negation, there should be no difference between A and N conditionals because both yield only a partial match with the alternative model.

2.1. Method

2.1.1. Participants

Thirty-two undergraduate students from the Université de Bourgogne took part in this experiment.

2.1.2. Material

Sixteen *if p then q* conditional premises with affirmative antecedent and consequent were used (Appendix A). Each premise presented an artificial relation between the antecedent and the consequent in order to prevent the construction of alternative models by retrieving experienced contingencies from long term memory. As far as it was possible, we avoided conditional relations known to be processed in particular ways, such as causal (Lecas & Barrouillet, 1999; Marcus & Rips, 1979;

Markovits et al., 1998) or deontic conditionals (Cheng & Holyoak, 1985). For half of the conditional premises, both the antecedent and the consequent involved a predicate which had an antonym. Each premise was shown with six different pairs of statements, one of the form p and q for the minor premise and the conclusion for MP, respectively, one of the form q and p for AC, two of the form $not\ p$ and $not\ q$ for DA, and two of the form $not\ q$ and $not\ p$ for MT. As far as the denial inferences DA and MT were concerned, the conclusion was always a sentence with an explicit negation (e.g. The musician does not play bassoon), whereas the minor premise contained either an explicit (The musician is not Polish) or an implicit (The musician is Russian) negation. Thus, for each of the 16 conditional premises, participants had to evaluate three valid inferences (one MP and two MT) and three invalid inferences (one AC and two DA), resulting in 96 inferences to be evaluated.

2.1.3. Procedure

The experiment was presented on screen and was monitored by Psycop software for Macintosh. Each trial began with a ready-signal (a striped rectangle) that participants removed by pressing any key on the keyboard. Then they were presented with a first screen containing a short text introducing an *if p then q* statement (conditional premise) to be memorized. When they pressed a key, the participants were immediately presented with a second screen containing two statements: a minor premise and a conclusion (the latter in bold type). They had to evaluate whether this conclusion logically and necessarily followed from the preceding conditional premise and the minor premise. Both the minor premise and the conclusion remained on the screen until the participants gave their response by pressing one of two keyboard keys identified as 'Yes' and 'No'. Both the reaction time and the type of response were registered and another trial was started (ready-signal).

The 96 experimental trials were presented in random order without any feedback on performance. This experimental session was preceded by a training session of three trials (one DA and one MT inference with explicit negation on the minor premise, and one MP inference). The experiment lasted between 25 and 40 min.

2.2. Results

The data obtained for two participants were discarded. One of them did not endorse any DA inference, and the other had mean RTs that differed from the mean RTs of the sample by more than two standard deviations on four types of inference (out of six). All the analyses were performed using both participants and conditional premises as random factors ($F1$ and $F2$ analyses, respectively).

2.2.1. Rate of endorsement

The rate of endorsement of each type of inference was high, probably due to the experimental design (Table 2). A 2 (type of conditional: A, with antonym, versus N, without antonym) \times 6 (type of inference: MP, AC, DA with explicit negation, DA with implicit negation, MT with explicit negation, MT with implicit negation) ANOVA with the two variables (for $F1$) or the latter only (for $F2$) as within subject

Table 2
 Production rates and mean reaction times as a function of the type of conditional premise (A, with, and N, without antonyms) in the 25 participants who endorsed at least one inference in each of the six experimental conditions, and mean reaction times (A and N conditionals confounded) in all the 30 participants in Experiment 1

Inferences						
	MP	AC	DAE ^a	DAI	MTE	MTI
A conditionals	2018 (0.958)	2355 (0.767)	3079 (0.650)	3045 (0.621)	3345 (0.708)	3117 (0.725)
N conditionals	1900 (0.963)	2214 (0.725)	2914 (0.533)	2979 (0.521)	2923 (0.588)	3108 (0.604)
Overall (all the participants)	1974	2340	2988	3021	3280	3274

^a DAE refers to denial of antecedent inference with an explicitly negated minor premise, whereas DAI refers to DA with an implicitly negated minor premise (see text).

factors was conducted on the rates of endorsement. There was a main effect of the type of inference ($F(5, 145) = 29.00$, $MSe = 0.042$, $P < 0.001$; $F(5, 70) = 34.13$, $MSe = 0.010$, $P < 0.001$), and a main effect of the type of conditional which was significant only when the participants were used as the random factor ($F(1, 29) = 13.28$, $MSe = 0.046$, $P < 0.005$); it just failed to reach significance with the conditional premises as the random factor ($F(1, 14) = 4.24$, $MSe = 0.038$, $P = 0.06$). The A conditionals elicited a higher rate of endorsement than the N conditionals (0.738 versus 0.656), mainly on denial inferences (i.e. DA and MT, 0.562 for N and 0.676 for A conditionals), whereas the rates of endorsement were close for the initial-model inferences (i.e. MP and AC, 0.844 for N and 0.863 for A conditionals). The interaction between the type of conditionals and the number of models required for the deduction (one model for MP and AC versus two models for DA and MT) was significant ($F(1, 29) = 12.51$, $MSe = 0.015$, $P < 0.005$; $F(1, 14) = 10.94$, $MSe = 0.005$, $P < 0.005$). Finally, AC was more frequently endorsed than DA (0.746 and 0.582, respectively) ($F(1, 29) = 27.24$, $MSe = 0.040$, $P < 0.001$; $F(1, 14) = 58.13$, $MSe = 0.005$, $P < 0.001$).

2.2.2. Reaction times

Some of the 30 participants did not produce at least one RT for each of the 12 experimental conditions (i.e. 6 inferences \times 2 types of conditional). As a consequence, a first analysis was performed on RTs whatever the type of conditional considered (A or N). A second analysis was performed with the type of conditional and the type of inference as independent variables on the data from the participants who had at least one RT in each of the 12 experimental conditions ($n = 25$).

A first one-way ANOVA was performed on the mean RTs for endorsement with the type of inference as the within-subject factor. For each participant, the RTs which differed from his or her own mean RT on each of the six inferences by more than two standard deviations were discarded. This procedure led to the discarding of less than 5% of the RTs registered. In accordance with our hypothesis, DA and MT (3005 and 3277 ms, respectively) took longer to verify than MP and AC (1974 and 2340 ms, respectively) ($F(1, 29) = 103.75$, $MSe = 373\ 000$, $P < 0.001$; $F(1, 14) = 272.37$, $MSe = 74\ 600$, $P < 0.001$). Thus, the higher the number of models the inference required (two models for DA and MT versus one model only for MP and AC), the longer the time needed for its endorsement. The effect of directionality was also significant. Among the one-model inferences, the backward inference AC took longer to endorse than the forward inference MP ($F(1, 29) = 23.56$, $MSe = 89\ 954$, $P < 0.001$; $F(1, 14) = 16.90$, $MSe = 25\ 489$, $P = 0.001$). Among the two-model inferences, the backward inference MT took longer to endorse than the forward inference DA ($F(1, 29) = 5.68$, $MSe = 390\ 038$, $P = 0.02$; $F(1, 14) = 5.19$, $MSe = 182\ 231$, $P < 0.05$). The number of models required and the directionality did not interact ($F(1, 29) < 1$, $F(1, 14) < 1$). As the mental models theory predicts, AC (2340 ms) took less time to endorse than DA, whether the negation used for the minor premise was explicit (2988 ms) ($F(1, 29) = 42.94$, $MSe = 146\ 922$, $P < 0.001$; $F(1, 14) = 161.50$,

$MSe = 26\,832$, $P < 0.001$), or implicit (3021 ms) ($F(1, 29) = 30.12$, $MSe = 231\,329$, $P < 0.001$; $F(1, 14) = 73.20$, $MSe = 59\,872$, $P < 0.001$).

This first analysis totally confirmed the predictions derived from the mental models theory, irrespective of whether the participants ($F1$) or the conditionals ($F2$) were used as the random factor. In this latter case, none of the observed effects interacted with the type of conditional (A or N). We replicated the results of Marcus and Rips (1979), who observed that AC was endorsed more rapidly than DA (difference of 599 ms compared to 665 ms in our experiment), though they used only three different conditional premises in their experiment. However, they observed shorter RTs for AC than MP, a phenomenon they could not explain, whereas we observed the reverse effect which was predicted by the directionality of models hypothesis (Evans, 1993).

A second ANOVA was performed to evaluate the effects of the type of conditional (A versus N) and the type of negation used in the minor premise (implicit versus explicit) for denial inferences (i.e. DA and MT), with the type of conditional and the type of inference as within-subject factors. This analysis, conducted on the data for only 25 participants, confirmed both the effect of the number of models required and the directionality effect previously observed ($P < 0.001$ in each case). Furthermore, A conditionals elicited longer endorsements (2826 ms) than N conditionals (2673 ms) ($F(1, 24) = 10.43$, $MSe = 169\,225$, $P < 0.01$; $F(1, 14) = 9.78$, $MSe = 143\,862$, $P < 0.01$). This effect did not interact either with the type of inference ($F(1, 24) = 1.44$, $P = 0.22$, $F(1, 14) = 1.01$), or with the number of models required ($F1$ and $F2 < 1$), contrary to what was observed for the frequencies of endorsement. Thus, although the A conditionals elicited higher frequency of endorsement than the N conditionals (mainly for the denial inferences), these endorsements took longer.

As far as the type of negation for denial inferences was concerned, it varied as a function of the type of conditional. When A conditionals were used, explicit negations elicited longer endorsement times than implicit negations (3212 and 3081 ms, respectively), whereas the reverse was observed with N conditionals (2919 and 3044 ms, respectively). This interaction did not reach significance ($F(1, 24) = 3.51$, $MSe = 233\,500$, $P = 0.07$; $F(1, 14) = 3.56$, $MSe = 51\,923$, $P = 0.08$). Nonetheless, as we predicted, the denial inferences with explicit negation were more rapidly endorsed with N conditionals than with A conditionals (2919 and 3212 ms, respectively) ($F(1, 24) = 6.63$, $MSe = 324\,712$, $P = 0.02$; $F(1, 14) = 9.66$, $MSe = 106\,309$, $P < 0.01$), whereas there was no difference when the negation was implicit ($F1 < 1$, $F(1, 14) = 1.91$, $MSe = 86\,367$, $P = 0.19$).

Thus, the results of this experiment confirmed the main hypothesis derived from the mental models theory. The denial inferences DA and MT which make it necessary to flesh out the initial model took longer to endorse than MP and AC which do not require such a fleshing out. A directionality effect was observed, with forward inferences (MP and DA) being endorsed more rapidly than backward inferences (AC and MT, respectively). These effects were significant whatever the type of random factor used (i.e. subjects or items), and did not interact. Moreover, the time required to endorse the denial inferences DA and MT depended on the strength of the

matching between the wording of the minor premise and the conclusion and the nature of the representations involved in the alternative models we hypothesized. On the one hand, when the predicates involved in the conditional premise do not have antonyms (N conditionals), the theory predicts that the alternative model contains representations of these predicates which are cued by negation tags. These representations would be perfectly matched by explicit negative sentences. In fact, the shortest RTs on denial inferences were observed when both the minor premise and the conclusion were explicitly negated and applied to N conditionals (2919 ms), whereas the longest RTs were registered when explicit negations were applied to A conditionals (3212 ms).

Our design was a partial replication of the experiment by Marcus and Rips (1979). However, this design had a weakness because there was a confound between the number of models the inference required and the relation between the conditional premise, on the one hand, and both the minor premise and the conclusion on the other. In effect, for MP and AC, both the minor premise and the conclusion matched the terms of the conditional premises, whereas these propositions either differed from the terms of the conditional (implicit negation) or involved negations which required processing (explicit negation) for the denial inferences DA and MT. Such a matching could explain why AC and MP were endorsed more quickly than DA and MT. We therefore designed a second experiment to avoid this flaw in which we used minor premises and conclusions which differed from the wording of the conditional premises for all the inferences which were to be evaluated.

3. Experiment 2

In this experiment, both the minor premise and the conclusion included either synonyms of the terms involved in the conditional premise for AC and MP, or antonyms for DA and MT. For example, for a conditional premise like ‘if it is a high tree, then the mushrooms are toxic’, we proposed ‘it is a tall tree’ and ‘the mushrooms are harmful’ as the minor premise and conclusion, respectively for MP, the reverse for AC, and ‘it is a small tree’ and ‘the mushrooms are good’ for DA and MT. Thus, neither the initial-model (MP and AC) nor the two-model inferences (DA and MT) had a minor premise or conclusion which directly matched the content of the conditional premise. The predictions were the same as those of Experiment 1.

3.1. Method

3.1.1. Participants

Thirty undergraduate students from the Université de Bourgogne took part in this experiment.

3.1.2. Material and procedure

Twelve *if p then q* premises which permitted a negation by antonyms and an affirmation by synonyms of both the antecedent and the consequent were used (Appendix B). For MP and AC, both the minor premise and the conclusion

contained synonyms of the antecedent and the consequent, whereas we used antonyms for DA and MT. The participants had to evaluate six inferences for each of the 12 conditional premises: the four conditional syllogisms MP, AC, DA, and MT, and two ‘false’ inferences intended for the control of participants’ attention during the task. These two false inferences linked an explicit affirmation of the antecedent with an explicit negation of the consequent, either as minor premise and conclusion, respectively, or the reverse. A low rate of endorsement of these ‘false’ inferences should provide evidence that participants paid attention to the task. In all, there were 48 experimental trials and 24 control trials.

The procedure was the same as in Experiment 1, except that RTs for the control trials were not registered.

3.2. Results

The rate of endorsement of inferences was as high as in Experiment 1 (88, 79, 78 and 72% for MP, AC, DA and MT, respectively). The mean rate of endorsement for the 24 control trials was only 5% (max 5, min 0 out of 24). Thus, all the participants were considered as having paid sufficient attention to the task and were included in the analysis.

For each participant, the RTs for endorsement were registered and those which differed from the mean by more than two standard deviations were discarded. This procedure resulted in less than 5% of the RTs being rejected.

We conducted a first 2 (number of models required: 1 for MP and AC versus 2 for DA and MT) \times 2 (directionality: forward for MP and DA versus backward for AC and MT) ANOVA on the mean RTs for each of the four inferences with the participants as the random factor (see Table 3). As previously observed, the two-model inferences DA and MT (mean RT of 3050 ms) took longer to endorse than the initial-model inferences MP and AC (mean RT of 2806 ms) ($F(1, 29) = 7.23$, $MSe = 246\ 954$, $P = 0.01$), and the backward inferences took longer than the forward inferences (2999 and 2857 ms, respectively) ($F(1, 29) = 11.11$, $MSe = 54\ 341$, $P < 0.005$). These main effects accounted for 97% of the experimental variance and their interaction was not significant ($F < 1$).

A second ANOVA with the same design was performed on the mean RTs with the 12 conditional premises as the random factor. These means were computed for each conditional premise from the RTs of participants who endorsed all of the four inferences. Thus, the number of participants included varied from one conditional

Table 3
Mean reaction times for the endorsement of the four conditional syllogisms in Experiment 2

Inferences			
MP	AC	DA	MT
2709	2903	3006	3095

premise to another, from 11 to 22 (mean 15.6) out of 30. There was a significant effect of directionality (2768 and 2947 ms for forward and backward inferences, respectively) ($F(1, 11) = 7.30$, $MSe = 52\,945$, $P < 0.05$), but the effect of the number of models failed to reach significance ($F(1, 11) = 3.27$, $MSe = 69\,740$, $P < 0.10$). This latter result suggests some heterogeneity among the conditional premises used and the quality of the synonyms and antonyms presented. A close inspection of the results revealed that two conditional premises (i.e. 2 and 4 in Appendix B) differed strongly from the others on the number of MP inferences endorsed (22 and 21 out of 30 for conditionals 2 and 4, respectively, versus a mean of 27.3 (SD 1.34) for the other conditionals). This suggests that participants judged the minor premise or the conclusion as poor synonyms of the antecedent and the consequent of the conditional premise to be processed. As far as conditional 2 was concerned, it could be that some participants considered that ‘modern’ is not a synonym of ‘contemporary’ (in music) because ‘modern’ probably refers to pop-music whereas ‘contemporary’ refers to a sophisticated style of music (maybe judged as boring). We did not find any explanation for conditional 4. When these two conditionals were discarded, the main effect of the number of models was significant ($F(1, 9) = 15.39$, $MSe = 33\,200$, $P < 0.001$).

Thus, this experiment confirmed the two main results previously observed: (a) the higher the number of models required, the longer the inference takes to endorse; (b) the forward inferences were endorsed more rapidly than the backward inferences. It should also be noted that the effect of the number of models was stronger in Experiment 1 (983 ms) than in Experiment 2 (250 ms). This was mainly due to the increase in the RTs for the first-model inferences MP and AC (2157 ms in Experiment 1 versus 2806 ms in Experiment 2), whereas the RTs for DA and MT remained practically unchanged (3140 ms in Experiment 1 and 3050 ms in Experiment 2). Indeed, the rapid responses for MP and AC in Experiment 1 were probably due to a perfect matching between the antecedent and the consequent of the conditional premise on the one hand, and the minor premise and the conclusion to be evaluated on the other hand. The use of synonyms in this experiment slowed down this matching. However, it could be that the difference observed between the first- and two-model inferences in Experiment 2 underestimated the time required for the fleshing out process.

In effect, the first p - q mental model would contain representations of the terms the conditional premise involves. Thus, the endorsement of both MP and AC inevitably requires subjects to compute the correctness of the fit between these representations and the synonyms proposed. On the other hand, the first alternative model needed to endorse DA and MT would link opposite values of the antecedent and the consequent which are retrieved from LTM. Thus, it is possible that these opposite values corresponded exactly to the antonyms proposed as the minor premise and conclusion. As a consequence, the probability of a perfect matching leading to faster answers was necessarily higher for the two-model inferences DA and MT than for the first-model inferences MP and AC. In other words, the design we used could only weaken the effect we hypothesized.

3.3. Discussion

The results of both experiments confirmed all the predictions derived from the mental models theory. The higher the number of models required for a given inference, the longer the time needed to endorse it. Thus, MP and AC, which would require only the initial model, were endorsed more rapidly than DA and MT, which require an additional fleshing out process. This effect persisted even when none of the terms used in the minor premise and the conclusion replicated the wording of the conditional premise. As the directionality hypothesis of Evans (1993) predicted, backward inferences (i.e. AC and MT) took longer to endorse than forward inferences (MP and DA, respectively). The fact that these effects were significant whatever the random factor used (participants or items) testifies to their robustness. The effect of the type of negation involved in the minor premise conformed to the predictions derived from the mental models theory (Experiment 1). A denial inference is all the faster endorsed if the minor premise and the conclusion match the content of the alternative model that is constructed. The inferences from minor premises with an explicit negation were more rapidly endorsed when the conditional premise did not contain terms likely to be negated by antonyms (in such a case, the alternative models would contain negation tags) than when the conditional premise contained such terms (in that case, the alternative models would represent these antonyms rather than negation tags).

How could the Braine and O'Brien (1991) theory account for these results? As we have observed, their theory predicts longer RTs for MT than for MP, because the former requires a reasoning strategy by *reductio ad absurdum*, whereas the latter would be directly produced by a cognitive rule. However, as far as AC and DA are concerned, there is no clear prediction concerning (a) their relative speed and (b), more importantly, how these inferences should be integrated in an overall pattern of RTs for all the inferences. We suggest that at least three hypotheses should be put forward to account for our results.

First, it should be assumed that the invited inferences underlying AC and DA are prompted by minor premises of the form q and *not p*, respectively, rather than by the mere comprehension of the conditional premise *if p then q*. In effect, we may suppose that the invited inference *if q then p* is produced automatically when the conditional premise *if p then q* is read. The comprehension process of the conditional sentence should lead subjects to store two propositions of the form *if p then q* and *if q then p* in working memory. Each of these could then be processed by the same MP mental rule when propositions either of the form p (for MP) or of the form q (for AC) enter working memory. Thus, if this invited inference was produced automatically when the conditional premise was processed, then the time course of MP and AC should be equivalent. However, this was not the case. The same argument holds for DA and the invited inference *if not p then not q*.

Second, we might suppose that the invited inference *if not p then not q* takes longer to produce than the inference *if q then p*, because in Experiment 1 DA took longer to endorse than AC. This suggestion is highly plausible because the produc-

tion of the invited inference *if not p then not q* requires the computation of opposite values or negations which could lead to extra time.

Finally, we can hypothesize a rather complex relation between the time courses of different cognitive processes such as (a) the activation of a production rule for MP, (b) the production of additional interpretations from a given sentence (i.e. invited inferences) for AC and DA, and (c) planning a strategy of reasoning for MT. In effect, the differences in RTs between MP and AC, on the one hand, and between DA and MT on the other, were approximately equivalent in the two experiments because the directionality effect did not interact with the number of models required. It seems difficult to account for this pattern of result without evoking some ad hoc proposition. It might be suggested that, in many cases, MT does not follow from a reasoning strategy but from “participants accepting the invited inferences and thus assuming that *p* and *q* co-occur and that *not p* and *not q* co-occur”, as suggested by Romain, Connell and Braine (1983, p. 479). However, this latter proposal does not differ in any clear way from the mental models theory, because it is not clear in this case why we should need to hypothesize an MP inference schema stored in the lexical entry of *if*.

Thus, our results are not totally at odds with the Braine and O’Brien (1991) theory of *if*. Nevertheless, it needs many additional theoretical constructs to account for the time course of the different inferences. By contrast, these results were exactly what the mental models theory predicts. However, there is another model of conditional syllogisms in connection with the mental logic theory which permits precise predictions about the time course of the different inferences we studied, namely the Consistency Model (CM) proposed by Marcus and Rips (1979). The CM conceives the evaluation of conditional syllogisms as a four-stage process. The first stage comprises the preliminary encoding of the syllogism and the choice of an interpretation for the conditional (i.e. material conditional or material biconditional). Stage 2 compares the minor premise to the antecedent and the conclusion to the consequent of the conditional. When the minor premise matches the antecedent, a conclusion *q* is always accepted (this is MP), but a conclusion *not q* is always rejected. A complete mismatch at this stage takes the subject to Stage 3 which provides a decision regarding the consistency of the conclusion and the two premises. A decision that the propositions are consistent leads to Stage 4 whereas an inconsistent set of propositions leads subjects to reject the conclusion as always being false. Finally, Stage 4 comprises the negation of the conclusion and an examination of this negated conclusion for consistency, as in Stage 3. A decision of inconsistency leads subjects to endorse the conclusion. This process accounts for MT endorsement, because the negation of the MT conclusion *not p* leads to *p*, and the propositions *p* and *not q* (minor premise for MT) are inconsistent with a conditional *if p then q*. As far as AC and DA are concerned, the negation of the conclusion leads to a decision of consistency. In the former, the negation of the conclusion *p* is *not p* which is consistent with a minor premise *q* and the conditional *if p then q*. In the latter, the negated conclusion is *q*, which is consistent with the minor premise *not p*. These decisions of consistency would lead subjects to respond ‘sometimes true’, which is correct for both DA and AC.

According to CM, MP would result from Stage 2, whereas MT, DA and AC would require a complete four-stage process. However, the last three inferences would not have exactly the same time course because it would be more difficult to compute a double negation for both the MT (i.e. deny *not p*) and DA (i.e. deny *not q*) conclusions than to compute a simple negation for AC (deny the conclusion *p*). Thus, CM accounts for the faster endorsement of MP compared to AC, and of AC compared to both DA and MT. However, as far as RTs are concerned, AC should be closer to DA and MT than to MP, because DA and MT differ from AC only in the more complex processing of the negation, whereas AC requires two stages more than MP. The results of Experiment 1 where we used approximately the same type of syllogisms as Marcus and Rips (1979) contradicts this prediction (see Table 2). Moreover, CM does not distinguish DA from MT which require exactly the same processing stages and computations. Now, in our two experiments and as predicted by the Evans (1993) hypothesis, the difference between DA and MT was comparable to that observed between MP and AC. In fact, Marcus and Rips (1979) observed a difference of 242 ms between MT and DA (269 ms in our Experiment 1) which they could not explain.

Though the Consistency Model accounts for many phenomena, it does not account for the recurrent pattern of results we observed in which MP and AC differ from DA and MT, with the same difference occurring within each of these pairs. Thus, CM does not appear to account for the time course of inference evaluation better than the Braine and O'Brien (1991) theory of *if* does. By contrast, the mental models theory provides us with a simple explanation. The time needed to endorse an inference depends on the number of models required. Within a given model, an inference is endorsed all the more rapidly if its direction matches the directionality of the model.

As we suggested in Section 1, the developmental approach provides another way of testing the mental models theory. Barrouillet and Lecas (1998, 1999) and Lecas and Barrouillet (1999) have recently proposed a developmental model for conditional reasoning in childhood and adolescence. Following the Johnson-Laird and Byrne (1991) theory, this model proposes that the understanding of *if...then* statements depends on the number of models the subjects are able to construct and maintain in working memory. A first developmental level of interpretation involves the construction of one model only in which the representations of *p* and *q* co-occur without exhaustion tags on *p* nor implicit model. This first level would lead children to produce MP and AC more frequently than DA and MT. A second level would be reached by adding a second *not p-not q* model, which results in a biconditional interpretation which leads to the frequent production of all the MP, AC, DA, and MT inferences. Finally, subjects should reach a conditional level of understanding underpinned by the construction of the three models described by Johnson-Laird and Byrne (1991). This latter level should result in the more frequent production of MP and MT than of AC and DA (see Table 1).

The predictions resulting from this model can be summarized using two indexes. The first, which we have called the Conjunctive Index, is the difference between the production rates of MP and AC, on the one hand, and AC and DA on the other. If the

first interpretational level is conjunctive in nature, then the Conjunctive Index should be high in young children and then decrease with age, because both the biconditional and the conditional interpretations should result in a Conjunctive Index near to zero (see Table 1). The second, which we term the Logical Index, is the difference between the production rates of MP and MT, on the one hand, and AC and DA on the other. According to our model, this index should increase with age because both the conjunctive and biconditional interpretations should result in low Logical Index values whereas the conditional interpretation should result in high values. These predictions were tested in Experiment 3.

4. Experiment 3

In this experiment, third, sixth, and ninth graders and adults had to produce conclusions from the four types of conditional syllogism.

4.1. Method

4.1.1. Participants

Thirty-six children from the third (mean age 8.4 years; SD 0.3), sixth (mean age 11.3 years; SD 0.4), and ninth grade (mean age 15.6 years; SD 0.7), and 36 undergraduate students of the Université de Bourgogne took part in this experiment.

4.1.2. Material and procedure

Twelve *if p then q* conditional premises which presented an artificial relation between the antecedent and the consequent were used. Each of these 12 conditionals could be associated with one of six possible minor premises either of the form *p* (MP), *q* (AC), *not p* (DA, either with an implicit or an explicit negation, DAI and DAE, respectively), or *not q* (MT, either with an implicit or an explicit negation, MTI and MTE, respectively). This resulted in 72 different conditional syllogisms.

The participants were presented with 24 of these 72 syllogisms (experimental problems) and six additional problems used as distractors (see below). Each participant saw each of the 12 conditional premises twice, once with an affirmative minor premise (either MP or AC) and once with a negative minor premise (either DA or MT). The minor premises were permuted so that participants saw six MP problems, six AC, six DA (three with an implicit negation and three with an explicit negation), and six MT (three with an implicit negation and three with an explicit negation). The six additional problems were disjunctive syllogisms.

The task was administered to groups of 18 participants each. Each participant received a 30-page notebook with one problem presented on each page. Each experimental problem consisted of a short scenario introducing a proposition of the form *if p then q* followed by a minor premise of the form *p*, *q*, *not p*, or *not q* (with the two latter premises being negated either semantically or syntactically). At the bottom of the page the subjects had to select one of the two options ‘nothing follows’ or ‘something follows’ and, in the latter case, were asked to write down this conclusion. The experimenter had already explained, by means of disjunctive syllo-

gisms, that the conclusion was a new item of information which could be inferred with certainty from the information provided by the premises. The participants were not subject to any time constraints.

For the MP and AC syllogisms, conclusions of the form q and p , respectively, were considered as MP and AC productions. For the DA and MT syllogisms, conclusions of the form *not* q and *not* p , respectively, were considered as DA and MT productions, even when these negative conclusions took the form of an implicit negation.

4.2. Results

The production rates of each of the four inferences are displayed in Table 4. First, a 4 (level: third, sixth, ninth grade, and adults) \times 2 (type of inference: DA versus MT) \times 2 (type of negation: explicit versus implicit) ANOVA with type of inference and type of negation as within-subjects factors was performed on the number of denial inferences produced. There was a main effect of level, the production rate of denial inferences increasing from third (1.19 out of 3) to sixth grade (2.52) and then decreasing from ninth grade (2.40) to adulthood (1.06) ($F(3, 140) = 33.00$, $MSe = 2.61$, $P < 0.001$). This quadratic trend was exactly what the model summarized in Table 1 predicted. Among the denial inferences, MT was produced more often than DA (1.93 and 1.65 out of 3, respectively) ($F(1, 140) = 14.75$, $MSe = 0.79$, $P < 0.001$), but the effect of the type of inference interacted with the level. As our model predicted, the difference between MT and DA increased with age: virtually nil at third (1.17 and 1.21 for MT and DA, respectively) ($F < 1$) and sixth grade (2.53 and 2.51) ($F < 1$), this difference increased slightly in ninth grade (2.50 and 2.29) ($F = 1.97$, $P = 0.16$) and was highly significant in adults (1.54 and 0.58 for MT and DA, respectively) ($F = 41.77$, $P < 0.001$). As already observed (Barrouillet & Lecas, 1998), explicit negations elicited more denial inferences than implicit negations (1.93 and 1.66 out of 3, respectively) ($F(1, 140) = 30.75$, $MSe = 0.34$, $P < 0.001$), but this effect did not interact either with the level or with the type of inference. In consequence, this factor was not taken into account in the following analysis.

A one-way ANOVA with the four levels as between-subjects factor was conducted on the Conjunctive Index which was computed for each participant by subtracting the total number of DA and MT from the total number of MP and AC

Table 4
Production rates for the four inferences as a function of level in Experiment 3

Inferences	Level			
	Grade 3	Grade 6	Grade 9	Adults
MP	81	93	86	95
AC	75	92	81	46
DA	40	84	76	19
MT	39	84	83	52

produced (possible values from -12 to $+12$). This mean Conjunctive Index decreased from childhood to adolescence (4.58, 0.97, and 0.44 for third, sixth, and ninth graders, respectively), but increased from ninth grade to adulthood (4.19) ($F(3, 140) = 11.54$, $MSe = 14.32$, $P < 0.001$). As we predicted, the younger participants (third graders) manifested a conjunctive pattern of results: MP and AC (4.83 and 4.50 out of 6, respectively) were more often produced than DA and MT (2.42 and 2.33) ($F(1, 35) = 49.13$, $MSe = 3.85$, $P < 0.001$), and this comparison accounted for 99% of the experimental variance in this group. Sixth and ninth graders' productions conformed to a biconditional pattern with high production rates on all of the four inferences (Table 4). However, adult subjects, like third graders, produced more MP and AC (5.69 and 2.78, respectively) than DA and MT (1.17 and 3.11) ($F(1, 35) = 29.18$, $MSe = 5.43$, $P < 0.001$), this accounting for their high mean Conjunctive Index (4.19).

Another one-way ANOVA with the same design as the previous one was conducted on the Logical Index which was computed by subtracting the total number of AC and DA from the total number of MP and MT produced. As we predicted, this Logical Index increased with age (0.25, 0.08, 0.72, 4.86 for third, sixth, ninth graders, and adults, respectively) ($F(3, 140) = 21.53$, $MSe = 6.62$, $P < 0.001$). In adults only, MP and MT (5.69 and 3.11 out of 6, respectively) were produced more often than AC and DA (2.78 and 1.17, respectively) ($F(1, 35) = 47.85$, $MSe = 4.45$, $P < 0.001$), whereas this comparison did not reach significance in any other group.

To summarize, as we predicted, a first conjunctive interpretation (third graders) preceded a biconditional interpretation (sixth and ninth graders). Adult participants exhibited a pattern of results which was a mixture between conjunctive and conditional interpretations, with both of the indexes we used being significant in this group. Evidence for these tendencies was provided by an analysis of the individual patterns of responses.

4.2.1. Individual analysis

Each of the individual patterns (i.e. the number of productions for each of the four inferences MP, AC, DA, and MT) was compared to what would have been observed if the interpretation of the conditional had been entirely conjunctive, biconditional or conditional. These comparisons resulted in three distances from the three main conditional interpretations. Let us take the example of a participant producing six, three, two and five expected conclusions for MP, AC, DA and MT, respectively. Considering that the scores should be 6-6-0-0, 6-6-6-6, and 6-0-0-6 for perfect conjunctive, biconditional and conditional interpretations, respectively, the three distances would be 10, 8, and 6, respectively. A given participant was considered as exhibiting a particular interpretation when his or her pattern did not differ from this interpretation by more than 6, provided that there were no two equal distances. This procedure allowed us to classify 26 third graders (out of 36), 34 sixth graders, 30 ninth graders, and 28 adults (Table 5).

These individual analyses revealed that more than half of the 26 classified third graders exhibited a conjunctive interpretation, and the others a biconditional inter-

Table 5
Distribution of the participants in each level (out of 36) across the three main response patterns in Experiment 3

Pattern	Level			
	Grade 3	Grade 6	Grade 9	Adults
Conjunctive	14	4	2	13
Biconditional	12	30	25	1
Conditional	0	0	3	14

pretation. The 14 ‘conjunctive’ participants produced 85 and 79% of MP and AC, respectively, but only 10% of both DA and MT. It should be noted that the 10 participants who were not classified did not differ greatly from the others, except in a lower rate of inference production (89, 85, 46, and 40% for MP, AC, DA, and MT for the classified participants compared to 58, 48, 27, and 35% for the non-classified participants, respectively). Sixth and ninth graders mainly exhibited biconditional interpretations, with conjunctive and conditional patterns being very rare. These results provided strong confirmation of our model, and particularly of the existence of a conjunctive-like interpretation resulting in the production of both MP and AC but not of DA or MT, which require the construction of an alternative model.

As far as adults are concerned, the main interpretation was, as we predicted, a material conditional. However, the other dominant interpretation was not biconditional but conjunctive in nature. These participants produced 99, 82, 17, and 17% of MP, AC, DA, and MT inferences, respectively. These rates were very close to those observed in third graders. In fact, these results were in line with the Johnson-Laird and Byrne (1991) proposals. In effect, Johnson-Laird and Byrne (1991) suggested that adults would represent *if p then q* statements in an initial model containing both an explicit representation of the *p-q* co-occurrence and an implicit model which could be fleshed out if needed. However, let us suppose that the subject does not flesh out this implicit model. In this case, the mental models theory predicts that the subject should produce MP and AC inferences supported by the initial model but that he or she should judge that nothing follows from the two minor premises *not p* (DA) and *not q* (MT). In other words, he or she should produce a conjunctive pattern of inferences.

Another possibility could be that conjunctive patterns in adults result from a failure to read out conclusions from complex fleshed out mental models (i.e. at least three different types of models for the conditional). In this latter case, some participants could adopt a cautiousness strategy, drawing inferences only from minor premises which match the content of the conditional premise (i.e. *p* and *q*), and avoiding drawing conclusions from an uncertain and complex set of fleshed out mental models. This strategy would lead to the same conjunctive pattern.

On the other hand, our results suggest that when adults succeed in fleshing out the initial model, they achieve a complete conditional interpretation, because there was

only one participant who exhibited a biconditional interpretation whereas there were 14 participants who produced a conditional pattern of responses.

4.3. Discussion

The results of this experiment lend strong support to our developmental model which is derived from the Johnson-Laird and Byrne (1991) theory. The development of the understanding of conditional sentences evolves from a conjunctive interpretation, which is underpinned by a simple p - q model, to a biconditional interpretation which requires subjects to add an alternative *not p-not q* model to the former, and finally to a conditional interpretation permitted by the complete three-model representation Johnson-Laird has proposed. Contrary to what we observed with instance production or instance evaluation tasks (Barrouillet & Lecas, 1998, 1999; Lecas & Barrouillet, 1999), the inference production task has revealed a renewal of conjunctive interpretations in adulthood whereas such interpretations had disappeared during adolescence.

This developmental trend suggests that the representation of the conditional does not contain implicit models until adulthood. Indeed, implicit models would require a fleshing out process for denial inferences, and we have interpreted the frequent conjunctive patterns in adults as a failure to flesh out the implicit model. Now, there is no reason to suppose that adolescents are better able than adults to achieve this fleshing out. As a consequence, the rare conjunctive patterns in adolescents (only 6 out of 72 sixth and ninth graders) suggest that their initial model of the conditional contains both a p - q and a *not p-not q* model. Note that this result coincides with the Markovits (1993) developmental theory which proposes such a double initial model for *if* in children and adolescents. However, our previous works and the present results suggest that there is an earlier developmental level where conditional statements are represented by one model only. The use of the implicit model that the Johnson-Laird and Byrne (1991) theory proposes would result from the increase in the number of models to be represented. It seems that it is only in adulthood that the representation reaches such a level of complexity that an implicit model is needed. In effect, adults manifest either a conditional pattern when they succeed in fleshing out the implicit model, or a conjunctive pattern when they fail to do it.

It could be noted that these results differ largely from those we observed in Experiments 1 and 2, where adults very frequently endorsed AC, DA, and MT inferences. However, it must be stressed that these experiments differed in their design in two ways. First, in Experiments 1 and 2, participants did not have to produce but to evaluate inferences. It could be that the simultaneous presentation of a minor premise and a conclusion in Experiments 1 and 2 triggered the construction of an alternative model for DA and MT in a more effective way than the mere presentation of the minor premise in Experiment 3. Moreover, the need for a quick response was stressed in Experiments 1 and 2. As suggested by Johnson-Laird (1993), subjects would be deductive satisficers, and the need for a quick response

would incline them to endorse an inference as soon as they find the convenient model, without looking for counter-examples.

Once more, these results did not fit very well with the mental logic theory. In effect, AC and DA, which are both supposed to result from invited inferences, differed in their production rates. As predicted by the mental models theory, AC was more frequent than DA in third graders and in adults. Moreover, conjunctive patterns were frequent in children but disappeared in adolescence. This developmental trend suggests that the conjunctive-like response pattern is the lowest developmental level. This runs counter to the claims of O'Brien, Dias and Roazzi (1998) who believe the lowest level to be the biconditional.

These authors noted that virtually no experiment reported the conjunctive pattern predicted by the mental models theory in young children and concluded that developmental research contradicts the mental models theory. However, some of us have already observed this pattern in an inference production task (Barrouillet & Lecas, 1998), and conjunctive interpretations have already been observed with various paradigms (Barrouillet & Lecas, 1998; Lecas & Barrouillet, 1999; Paris, 1973; Taplin, Staudenmayer & Tadonio, 1974). Moreover, it could be that the conjunctive pattern only appears when children have to produce inferences from unfamiliar relations between antecedent and consequent as we used in this experiment. Indeed, unfamiliar relations do not allow the use of factual knowledge which could underlie inference production. When young children conclude from the premises 'if it rains then the grass is wet' and 'the grass is dry' that it has not rained, it is not clear whether they use a genuine reasoning process or simply retrieve from memory available knowledge about contingencies in the world. Our results suggest that although young children are inclined to endorse and produce denial inferences in many experiments, the predicted conjunctive pattern appears when they have to produce conclusions from unfamiliar relations.

We have seen that the results of the first two experiments could be accounted for by the mental logic approach provided that AC is accepted as being easier to produce than DA because the latter involves negations. Accordingly, AC is produced more often than DA in children. We might suppose that an age-related increase in processing capacities should allow adults to set aside conversational interpretations (i.e. invited inferences underpinning both AC and DA), as suggested by Rumain et al. (1983) and Braine and O'Brien (1991). If AC is easier to produce than DA for children, it should also be more difficult to inhibit in adults. Accordingly, AC was produced more often than DA in adults.

However, this line of argument leads to a complex and rather odd conception of development. First, it might be supposed that the frequent production of DA in adolescence but not in childhood results from an increase in processing capacities from childhood to adolescence. Second, this same increase in processing capacities from adolescence to adulthood should account for invited inferences being inhibited in adults. In other words, the same factor (i.e. the increase in processing capacities) would lead to different results during developmental periods, sometimes accounting for the production of a given inference, sometimes accounting for its inhibition. Of course, this is not impossible. However, the mental models account is far more

simple. The response patterns observed both in childhood and adolescence were predicted by the hypothesis of an increase in the number of models the subjects can produce and coordinate. The response patterns observed in adults were either conditional or conjunctive, as predicted by our developmental model and the Johnson-Laird and Byrne (1991) theory of initial models, respectively.

In summary, none of the results presented in this article supported the proposals of the mental logic theory, whereas the functional and developmental predictions drawn from the mental models framework were confirmed. We do not claim that our results rule out the mental logic theory, but only that it needs some deep revision if it is to account for the main phenomena we observed.

5. General discussion

In discussing the mental logic–mental models controversy, Bonatti (1998, p. 435) stressed that “one way to make progress is to concentrate on a small domain that both theories address, and thoroughly examine which theory seems to be more likely to capture human reasoning”. Conditional reasoning is one of these small domains, and the results of the three experiments presented here were more compatible with the mental models theory than with its rival. Of course, it could be claimed that many of our predictions do not issue from the standard mental models theory, which does not include proposals about directionality of models nor explicit developmental assumptions. However, the Evans (1993) suggestions clearly constitute an attempt to improve the general framework provided by the standard theory rather than make a strong claim against mental models. In the same way, the Barrouillet and Lecas (1998) developmental model results from straightforward predictions permitted by the Johnson-Laird and Byrne (1991) conception of conditional reasoning. As a consequence, we assume that our results provide evidence for a mental models account of conditional reasoning in children and adults.

On the other hand, it could be claimed that these results do not rule out the mental logic approach. Indeed, with the exception of the Marcus and Rips (1979) consistency model, mental logic does not really predict anything about the time course of inference production. As pointed out by O’Brien et al. (1998), mental logic does not predict systematic differences among AC, DA, and MT. However, the mental models theory does, and these predictions are confirmed both by our chronometric and developmental studies. It seems clear that, although being outside the scope of a mental logic theory, the reasoning processes underpinned by invited inferences or pragmatic principles need a precise functionalist description. Indeed, we remain unsatisfied by the rather vague mental logic prediction that AC, DA and MT will lead to a mixture of *nothing follows* responses and expected determinate responses (O’Brien et al., 1998, p. 414). In this article we have made some suggestions about the way mental logic could account for the chronometric results, but the mental logic predictions concerning development seem more difficult to reconcile with the observed evolution of the inference production rates.

Finally, our results could be limited in scope, due to the nature of the conditional

relations we studied (i.e. artificial indicative conditionals). Obviously, further studies on deontic or familiar relations are needed. Furthermore, we cannot assume that the processes involved in tasks that require speedy decision making (Experiments 1 and 2) are the same as those involved in everyday reasoning. However, the fact that both the RT paradigm in Experiments 1 and 2 and the self-paced inference production task in Experiment 3 provided convergent evidence lends support to the mental models approach of human reasoning.

To conclude, both the mental models and the mental logic theories need further specifications to account for the results we observed. As far as the mental models theory is concerned, it seems that an assumption about the directionality of models (Barrouillet & Lecas, 1998; Evans, 1993) is required in order to account for the slower endorsement of AC and MT compared to MP and DA. Note that previous versions of the mental models theory included proposals about directionality of models in syllogistic reasoning (e.g. Johnson-Laird & Steedman, 1978). Furthermore, this theory needs a developmental facet which makes it necessary to modify some of the principles of the standard version such as the presence of exhaustion tags in initial models and the use of implicit models of the conditional (Barrouillet & Lecas, 1998). As we noted earlier, the mental logic theory should give a better account of AC and DA endorsement and production and of their links with the cognitive rule for MP. At the very least, the difference in nature between MP, supposedly underpinned by an inference schema, and the other inferences is not clearly evident from the results we gathered. Instead, the chronometric and developmental evidence corroborates the mental models theory.

Acknowledgements

The authors would like to thank Arnaud Desarménien for collecting the data of Experiment 1.

Appendix A

English translation of the conditional premises used in Experiment 1 with the antonyms used for denial inferences.

A.1. A conditionals

1. If the card is long (short), then the number is even (odd).
2. If a lemur is in captivity (free), then it is infertile (fertile).
3. If the door is open (closed), then he is in (out).
4. If it is a child's (adult's) bike, then he puts it outside (inside).
5. If it is a domed (ridged)* tent, then it is waterproof (it leaks).
6. If the man is married (bachelor), then he precedes (follows) the statue.
7. If the flowers are artificial (real), then he puts them on the right (left).
8. If it was a steam (sailing) ship, then it left in the morning (afternoon).

*The introductory text specified that there were only two kinds of tents.

A.2. *N conditionals*

9. If its waterpolo, then its Monday morning.
10. If they're bears, then she gives carrots.
11. If he goes to Marseilles, then he's an engineer.
12. If its a wardrobe, then its made of cherry.
13. If she has a statistics exam, then she wears shorts.
14. If the musician is Polish, then he plays the bassoon.
15. If its the month of March, then its skirts.
16. If the car has 5 horsepower, then its a Peugeot.

Appendix B

English translation of the conditional premises used in Experiment 2 with the synonyms and the antonyms used. Example: for the Conditional 1, the major premise was 'If the tree is tall, then the mushrooms are poisonous', and the four minor premises were 'the tree is high', 'the tree is low', 'the mushrooms are bad', and 'the mushrooms are good' for MP, DA, AC, and MT, respectively.

1. If the tree is tall (high/low), then the mushrooms are poisonous (bad/good).
2. If the piece is modern (contemporary/classical), then the musicians are happy (pleased/angry).
3. If the wire is thin (narrow/wide), then it is boiling (very hot/cold).
4. If the rod is twisted (bent/straight), then it is hollow (empty/solid).
5. If the soil is damp (wet/dry), then the hole is occupied (inhabited/abandoned).
6. If the patient is nervous (agitated/calm), then the dose is increased (larger/normal).
7. If the supervisor is salaried (paid/a volunteer), then the courses are for a number of participants (for groups/for individuals).
8. If the animal has soft (silky/rough) hair, then it is well (healthy/sick).
9. If the weather is bad (wet/sunny), then the drinks are gratis (free/paid for).
10. If a horse was clever (intelligent/stupid), then it lost (was beaten/it won) the race.
11. If a worker is drunk (intoxicated/sober), then his pay is cut (reduced/unchanged).
12. If the text is comical (funny/serious), then the questions are difficult (complex/easy).

References

- Barrouillet, P. (1997). Modifying the interpretation of if...then sentences in adolescents by inducing a structure mapping strategy. *Current Psychology of Cognition*, 16, 609–637.
- 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*, 4 5(4), 289–302.
- Bonatti, L. (1998). What the mental logic–mental models controversy is not about. In M. D. S. Braine, & D. P. O'Brien, *Mental logic* (pp. 435–446). Mahwah, NJ: Erlbaum.
- Braine, M. D. S. (1990). The natural approach to reasoning. In W. F. Overton, *Reasoning, necessity, and logic: developmental perspectives* (pp. 135–158). Hillsdale, NJ: Lawrence Erlbaum.
- 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.
- Braine, M. D. S., & Rumain, B. (1983). Logical reasoning. In J. H. Flavell, & E. Markman, *Handbook of child psychology, vol. 3. Cognitive development* (pp. 263–339). New York: Wiley.
- Case, R. (1985). *Intellectual development: birth to adulthood*. New York: Academic Press.
- Case, R. (1992). *The mind's staircase: exploring the conceptual underpinnings of children's thought and knowledge*. Hillsdale, NJ: Lawrence Erlbaum.
- Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, 17, 391–416.
- Evans, J. St. B. T. (1993). The mental model theory of conditional reasoning: critical appraisal and revision. *Cognition*, 48, 1–20.
- Evans, J. St. B. T., & Over, D. E. (1997). Rationality in reasoning: the problem of deductive competence. *Current Psychology of Cognition*, 16, 3–38.
- Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: the psychology of deduction*. Hillsdale, NJ: Lawrence Erlbaum.
- Halford, G. S. (1993). *Children's understanding: the development of mental models*. Hillsdale, NJ: Lawrence Erlbaum.
- Johnson-Laird, P. N. (1993). *Human and machine thinking*. Hillsdale, NJ: Lawrence Erlbaum.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Lawrence Erlbaum.
- Johnson-Laird, P. N., & Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology*, 10, 64–99.
- Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review*, 3, 418–439.
- Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1994). Why models rather than rules give a better account of propositional reasoning: a reply to Bonatti and to O'Brien, Braine, and Yang. *Psychological Review*, 101, 734–739.
- Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J. P. (1999). Naive probability: a mental model theory of extensional reasoning. *Psychological Review*, 106, 62–88.
- 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.
- Legrenzi, P. (1970). Relations between language and reasoning about deductive rules. In G. B. Flores D'Arcais, & W. J. M. Levelt, *Advances in psycholinguistics*. Amsterdam: North Holland.
- Marcus, S. L., & Rips, L. J. (1979). Conditional reasoning. *Journal of Verbal Learning and Verbal Behavior*, 18, 199–223.
- Markovits, H. (1993). The development of conditional reasoning: a reformulation of mental models theory. *Merrill-Palmer Quarterly*, 1, 131–158.
- 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.
- O'Brien, D. P., Braine, M. D. S., & Yang, Y. (1994). Propositional reasoning by mental models? Simple to refute in principle and in practice. *Psychological Review*, 4, 711–724.
- O'Brien, D. P., Dias, M. G., & Roazzi, A. (1998). A case study in mental models and mental logic debate: conditional syllogisms. In M. D. S. Braine, & D. P. O'Brien, *Mental logic* (pp. 385–420). Mahwah, NJ: Erlbaum.
- Ormerod, T. C., Manktelow, K. I., & Jones, G. V. (1993). Reasoning with three types of conditionals: biases and mental models. *Quarterly Journal of Experimental Psychology*, 46A (4), 653–677.
- Paris, S. G. (1973). Comprehension of language connectives and propositional logical relationships. *Journal of Experimental Child Psychology*, 16, 278–291.
- Pascual Leone, J. (1988). Organismic processes for neo-piagetian theories: a dialectical causal account of

- cognitive development. In A. Demetriou, *The neo-piagetian theories of cognitive development: toward an integration* (pp. 25–64). Amsterdam: North Holland.
- Rips, L. J. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.
- 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*, 4, 471–481.
- 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.