How can mental models theory account for content effects in conditional reasoning? A developmental perspective

Pierre Barrouillet*, Jean-François Lecas

Laboratoire d’Étude des Apprentissages et du Développement, CNRS ESA 5022, Université de Bourgogne, Dijon, France

Received 13 November 1997; accepted 26 February 1998

Abstract

This article proposes a modification to Johnson-Laird’s mental models theory applied to the interpretation of conditional statements of the form ‘if...then’. The model suggests that this interpretation is based on the construction of mental models supplied by establishing a correspondence between the semantic spaces associated with the antecedent and consequent of the statements. The construction of the models and the interpretation of the statements would depend on the nature of the semantic spaces involved, the interpretative context and the subject’s knowledge and processing capacity. Three experiments show that the interpretation of conditional rules depends, for example, on whether or not the conditional rule possess binary terms (e.g. boy/girl). The developmental approach makes it possible to reveal phenomena which tend to remain hidden in studies of adult functioning. We show that the model accounts for a number of the reasoning biases described in the literature as well as for the interpretation of various conditional forms which do not have a truth-functional meaning. © 1998 Elsevier Science B.V. All rights reserved

Keywords: Cognitive development; Mental models; Reasoning; Conditional reasoning

1. Introduction

The mental models theory as applied to propositional reasoning (Johnson-Laird and Byrne, 1991; Johnson-Laird et al., 1992) has recently been the object of much criticism (O’Brien, 1993; Bonatti, 1994a; Bonatti, 1994b; O’Brien et al., 1994; Rips, 1994; Braine et al., 1995). Johnson-Laird et al. (1992) have suggested that propositional reasoning consists, as does the resolution of syllogisms (Johnson-Laird,
1983), in a manipulation of representations (mental models) of the statements to be processed, without the intervention of any logical rule. This conception is opposed to the ‘mental logic’ view which holds that human beings possess logical inference schemas in memory and use these to make deductions (Braine, 1990; Braine and O’Brien, 1991; Rips, 1983; Rips, 1994). The criticisms raised by Bonatti doubtless constitute the most complete argument against the mental models theory (Bonatti, 1994a; Bonatti, 1994b). One of its main thrusts is to argue that, contrary to the claims made by Johnson-Laird and Byrne (1991), this theory does not make it possible to account for content effects in any more satisfactory way than the theories based on mental logic (Bonatti, 1994b).

The aim of this article is to show that, even though Bonatti’s argument could be correct, the mental models theory can, if slightly modified, account for certain content effects which no other current theory appears able to explain. We shall start by analysing the reasons why the mental models theory proposed by Johnson-Laird (Johnson-Laird et al., 1992) is unable to account fully for content effects in conditional statements of the form ‘if p then q’. A modified model will then be proposed and tested, in three experiments relating to the interpretation of conditional sentences and production of inferences. Finally, the ability of this model to account for the major known effects (reasoning bias, suppression of valid and invalid inferences) will be evaluated. The model is limited to the interpretation of ‘if’. It is not therefore an alternative to Johnson-Laird’s theory, but instead, and rather more modestly, a suggested modification to this theory. We adopt a developmental approach. In effect, the study of the development of the processes which underpin adult reasoning helps reveal certain properties which remain inaccessible to an analysis of the performances achieved by a completed system.

1.1. The mental models theory and content effects

Although Bonatti has criticised many points relating to the mental models theory, we shall concentrate on the problem associated with content effects. This problem is central given that Johnson-Laird and Byrne (Johnson-Laird and Byrne, 1991; Johnson-Laird and Byrne, 1993) reject the formal, syntactical approach of the mental logic in favour of a semantic approach which could account for content effects: ‘reasoning is not a formal or syntactic process but a matter of understanding meanings’ (Johnson-Laird and Byrne (1993) p.181). However, Bonatti (1994b) argues that it is wrong to assert that ‘in contrast [to mental logic], the model theory has the machinery to deal with meaning’ (Byrne (1991) p. 77). Bonatti argues as follows: according to Johnson-Laird and Byrne (1991), mental models are constructed on the basis of propositional representations of statements by means of a set of semantic procedures. These procedures do not therefore apply to natural language but instead transform a propositional representation of statements into another representation (mental model). The propositional representations on the basis of which the models are constructed must make the propositional content and logical form of the messages accessible. The problem raised by Bonatti relates to the level at which the pragmatic and contextual aspects are processed during the course of comprehension.
If the mental models constitute the type of structure which is constructed during textual comprehension, they should represent the contextualised meaning of the message and not just its literal meaning. Since mental models are constructed on the basis of propositional representations, the latter should be clear, rich and devoid of structural ambiguity. They cannot therefore be the result of a semantic analysis which simply yields the literal meaning of the message to be processed, since they must also integrate pragmatic and contextual aspects which are liable to modify the literal meaning. These representations can therefore only result from an analysis of the message in context. Thus, the semantic, contextual and pragmatic aspects must be processed before the mental models are constructed in order to provide a correct propositional representation which is used as an input into the construction procedure. In short, the mental models should be constructed on the basis of a propositional representation which is identical to that acted on by the inference schemas postulated by mental logic. Thus the mental models theory would appear to be no more able than the mental logic theories to account for content and context effects.

What is more, Bonatti (1994b) points out that in the model theory, the meaning of the connectors (i.e. and, or and if) consists of a variation to the truth tables which are sensitive only to the truth value (i.e. ‘true’ and ‘false’) but not to the contents of propositions. Lowe (1993) has noted that ‘semantic’ logical methods (e.g. truth tables) are just as formal as so-called ‘syntactic’ methods and are no better at taking account of the contents of propositions (see also, O’Brien et al., 1994). This author concluded that ‘to this extent it is in fact misleading to describe model-theoretic methods as semantic’ (Lowe (1993) p. 221). Rips (1994) makes a similar comment.

In their response to the Bonatti (1994a) and O’Brien et al. (1994) articles, Johnson-Laird et al. (1994) do not discuss the specific point concerning the contents and associated effects. However, it is possible to make a response to the comments above which underestimate the importance of the fleshing-out process. In fact, one of the key components of the mental models theory is the idea that subjects represent only part of the information explicitly, thus avoiding a working-memory overload. The information which remains implicit can be made explicit (fleshed out) through the construction of mental models which are added to the initial model.

For example, Johnson-Laird and Byrne (1991) suggest that a statement of the form ‘if p then q’ is represented by an initial model

\[ p \rightarrow q \]

in which each line corresponds to a distinct model. The first model represents the co-occurrence of the antecedent (p) and the consequent (q). The second model (the three dots) remains implicit but indicates that other possibilities exist. The square brackets (marker of exhaustion) indicate that the value p cannot appear in any of these alternative models. This initial model can be fleshed out to produce either a biconditional interpretation (1),

\[ p \leftrightarrow q \]

1 In fact, Bonatti appears to neglect the structural differences between the two postulated propositional representations. In Johnson-Laird’s theory, the propositional representation is not as complete as Bonatti claims: ‘there is a propositional representation, which is close to the surface of the utterance’ (Johnson-Laird (1983) p. 407). However, arguing that the mental models for propositional reasoning, as presented by Johnson-Laird et al. (1992), are formal systems insensitive to contents seems to be correct.
where the symbol ‘\( \neg \)’ is a propositional tag which indicates the negation of the term to which it refers, or a conditional interpretation, since there was no exhaustion symbol around \( q \) which may appear in other models (2).

\[
\begin{array}{c|c}
[p] & [q] \\
[\neg p] & [\neg q] \\
\end{array}
\]  

(1)

Even though the psychological algorithm proposed by the mental models theory uses only initial models (i.e. without fleshing-out, Johnson-Laird et al., 1992), and even though the conditions which trigger fleshing-out remain unclear (Bonatti, 1994a; Bonatti, 1994b; Braine et al., 1995), the full meaning of a statement can only be represented by making the implicit models explicit. However, according to the theory, knowledge stored in long-term memory (LTM) can have an impact on these processes: ‘one factor that can assist the process of fleshing-out is existing knowledge about the contingencies to be added to the models’ (Johnson-Laird et al. (1992) p. 424). Since the fleshing-out process acts on models which have already been formed and is sensitive to knowledge in LTM, the theory allows for the possibility of content effects arising after the construction of the propositional representation.

However, this only applies to situations in which subjects possess factual knowledge in LTM concerning the contingencies associating antecedent and consequent. When these cases are excluded, the criticisms levelled by Bonatti (1994b), Lowe (1993) and Rips (1994) remain relevant.

1.2. The representation of negation in model construction

What are the aspects which make the mental models theory into a formal system (Holyoak and Spellman, 1993) which is unable to take account of content effects? One possible response would be: everything that relates this theory to formal logical systems and, in the case of mental models, formalization on the basis of truth tables. The theory supposes that subjects represent a proposition on the basis of all the cases which validate it in the truth table (cf. the ‘if’ example above). To do so, it introduces a representation of negation in the form of a propositional-like tag (\( \neg \)). The authors consider the use of such a symbol to be ‘innocuous’ (Johnson-Laird and Byrne (1991) p. 44). In contrast, O’Brien (1993) considers that the introduction of propositional-like tags implies that part of the mental models theory might consist of a propositional mental logic, in particular in so far as the processing of negation is concerned.

Here, we shall defend the idea that the introduction of these representations of negation is one of the elements which reinforce the formal aspect of the theory and which prevents it from accounting for content effects. In effect, the use of this symbol draws the psychological theory closer to logical systems. The existence of
such representations of negation (a) presupposes a logical operation on the part of the subject since the representation \( \neg p \) establishes an equivalence between all the cases for which the proposition \( p \) is not verified (O’Brien, 1993), and (b) makes fleshing-out a formal process which implements a pre-established representation structure and assigns a negation symbol to certain elements independently of their nature. Thus Sloman (1996) suggests that mental models exhibit two main characteristics which ally them with rule systems: they have a logical structure and contain variables.

It is debatable whether negation signs are compatible with the general theoretical framework of mental models (Levesque (1986) cited by Johnson-Laird and Byrne, 1991; Inder, 1987). According to Johnson-Laird, a mental model is ‘an internal model of the state of affairs that the premises describe’ (Johnson-Laird and Byrne (1991) p. 35). Mental models are therefore representations of situations, whereas the existence of the negation sign makes them representations of classes of situations, whence the necessity for a representation of exhaustion ([…]) which is itself problematic (see Braine et al., 1995). In accordance with the definitions set out by Johnson-Laird we shall propose that the initial model of the representation of if corresponds to the representation of a hypothetical state of affairs in which the propositions \( p \) and \( q \) are verified. The alternative models would then correspond to states of affairs in which \( p \) is not verified rather than to negated representations of a verified \( p \). To put it clearly, if \( p \) corresponds to ‘this table is made of wood’, the alternative cases would not necessarily consist of the representation of a wooden table indexed by a negation sign but might instead correspond to a representation of states of affairs in which the proposition \( p \) is not verified (e.g. a metal or plastic table, etc.).

The mental models theory would thus make it possible to imagine that the result of fleshing-out depends to a large extent on the semantic nature of the terms represented in the explicit models. Let us imagine that a given model contains the instance ‘boy’ (e.g. in the representation of ‘if it’s a boy then his classroom is on the first floor’) and that the subject constructs an alternative model. It is quite likely that the result will not be a model containing the value ‘boy’ marked by a negation sign but, more simply, the value ‘girl’. Indeed, Johnson-Laird and Byrne (1991) have proposed that subjects are biased to represent affirmative values in the models. In contrast, the alternative value to ‘first floor’ might remain indeterminate unless the subject possesses knowledge about the building in question (e.g. that it has only two levels) in which case this value might be made explicit as ‘ground floor’. Thus it is possible, in the mental models theory, to differentiate between models constructed as a function of the semantic nature of the terms involved and the subject’s knowledge.

The purpose of this article is to demonstrate that the extent to which the constructed representations can be made explicit has an effect on the way subjects interpret statements. In short, and contrary to Bonatti’s assertions, the mental models theory does indeed possess the ‘machinery’ necessary to account for a large number of content effects. These effects appear to act during the construction of the models and should therefore be inherent to the construction procedure as described by the
theory. For this, it is sufficient to suppose that the values which are marked with a
negation sign in the standard theory are either represented by affirmative values or
remain indeterminate. When an initial model is constructed to represent an atomic
proposition the negation of which can be expressed in an affirmative way (e.g. ‘the
pupil is a boy’), the alternative model or models contain a single value (i.e. girl). In
the opposite case (e.g. ‘the ball is red’), the alternative models consist of a set of
possible values and are thus largely indeterminate with regard to this proposition.

In this article, we designate terms whose negation in a given context takes a
determinate value (e.g. girl/boy for pupils) as binary (B) and terms whose negation
in a given context is indeterminate (e.g. red for a coloured ball), provided that the set
of possibilities is not constrained by any special knowledge, as non-binary (N). If
our hypothesis is correct, propositions containing binary terms will be processed
differently from those containing non-binary terms.

1.3. Application to comprehension of the conditional

Legrenzi (1970) has shown that the introduction of a strictly binary situation (i.e.
in which the antecedent p and the consequent q can only assume two, mutually-
exclusive, values) impairs the interpretation of conditional statements. In this situa-
tion, statements are understood in such a way that the non-occurrence of the ante-
cedent implies the non-occurrence of the consequent. For example, this author
described to the subjects a mechanism in which the path taken by a ball was able
to switch on a light. When the mechanism consisted of only two paths (left and right)
and two lights (green and red) and the statement to be evaluated was ‘if the ball rolls
to the left, then the green lamp is lit’, the subjects considered the cases ‘left – green
lit’ (p, q) and ‘right – red lit’ (→p, ¬q) to be compatible with the rule and the cases
‘left – red lit’ (p, ¬q) and ‘right – green lit’ (→p, q) to be incompatible with it.

Marcus and Rips (1979) observed a similar phenomenon. Here we are again in the
presence of the frequently-observed biconditional interpretation of if (see Evans et
al. (1993) for a review).

The mental models theory allows us to explain the effect observed by Legrenzi
(1970) and Marcus and Rips (1979). Let us suppose (a) that p and q refer to strictly
binary situations as in the situation studied by Legrenzi, and (b) that the first alter-
native model constructed corresponds to the case where p and q are not satisfied (¬p,
¬q in the standard theory) as indicated by developmental research (Barrouillet,
1997; J.F. Lecas and P. Barrouillet, unpublished data). This leads to the following
mental models:

\[
\begin{array}{ll}
\text{left} & \text{green} \\
\text{right} & \text{red}
\end{array}
\]

(3)

Legrenzi’s results indicate that only these two models were constructed since the
subject’s interpretations were clearly biconditional (73% of responses). This sug-
gests that fleshing-out is stopped at this stage, probably because the subject obtains
two entirely explicit models which exhaust all the possibilities for both the ante-
cedent and the consequent (see below the ‘completeness principle’). The cases ‘left – red lit’ (p. ¬q) and ‘right – green lit’ (∼p. q) are then judged to be incompatible with the statement because they contradict the two models constructed.

Let us now consider a case in which there are numerous possibilities, both for the antecedent and the consequent. Imagine that, as in the last case, the subject constructs two models but without using the negation sign. The values in the second model are then indeterminate:

\[
\text{left green} \\
\text{Ind Ind}
\]

Let us imagine that the subject now has to judge the case ‘right – green lit’ (∼p. q). The ‘right’ can only correspond to the second model in which the value of the consequent is not specified. The incompatibility of this case with the statement should therefore be less obvious than in the last example and it should be rejected less frequently. The biconditional interpretation of the statement should therefore be less frequent in this case. Thus the only way to account for the predominance of biconditional responses in the situation designed by Legrenzi is to apply the principles underlying the construction of mental models.

However, why do subjects only construct two models in strictly binary situations? We suggest that this is due to an essentially relational conception of the conditional which results from two pragmatic principles: the principle of quantity proposed by Grice (1975) and another principle which we shall call ‘completeness’. Their application should enable subjects to maximise the relevance of the statement (Sperber and Wilson, 1986).

1.4. Relational nature of mental models for if

The hypothesis of a relational conception of if is compatible with the mental models theory (Johnson-Laird and Steedman, 1978; Johnson-Laird, 1983), and Evans (1993) has suggested introducing the idea of the directionality of the models from the antecedent to the consequent in the interpretation of if. Thus an if p then q statement would establish a relation between a variable P (the value of which is specified by the proposition p) and a variable Q (the value of which is specified by the proposition q). A statement of the type ‘if the ball rolls to the left, then the green lamp is lit’ informs the subject that the colour of the lamp which lights up (variable Q) depends on the path taken by the ball (variable P) and specifies the value assumed by Q (i.e. green) for one of the values of P (i.e. left).

In conformity with the mental models theory, the interpretation of the statement would not be limited to the values made explicit in the propositions p and q (i.e. the initial model). The processing of a simple conditional linking two or three atomic propositions would constitute one of the cases in which the initial model is fleshed out (Johnson-Laird et al. (1992) p. 424). Subjects would therefore attempt to flesh out this information by inferring from their existing knowledge the nature of the variables P and Q involved in the statement and the values which can be assumed by
for values of \(P\) not specified in this statement. These values might derive from the problem situation, the subject’s episodic memory or from the semantics associated with each of the terms involved (e.g. the variable ‘sex’ can only have one of two values, male and female). This would make fleshing-out an informal process since its result would depend on the semantic nature of the terms involved and on the subject’s knowledge. Calculation would lead subjects to represent directed relations which link determinate or indeterminate values from the antecedent variable (\(P\)) to the consequent variable (\(Q\)). Diagrams (3) and (4) can therefore be rewritten as (5) and (6)

\[
\begin{align*}
\text{left} & \rightarrow \text{green} \\
\text{right} & \rightarrow \text{red}
\end{align*}
\]

(5)

if there are only two possible values for \(P\) and \(Q\), or as

\[
\begin{align*}
\text{left} & \rightarrow \text{green} \\
\text{Ind Dir.} & \rightarrow \text{Ind Colour}
\end{align*}
\]

(6)

if multiple values are possible. In this latter case, the alternative model represents some indeterminate situations in which the possible directions (Dir.) not directly addressed in the conditional rule produce the lighting of unspecified coloured lamps (Colour). Note that, here, \(\text{Ind}\) is not a propositional tag acting on a representation (like the negation tag ‘\(\neg\)’ in the standard theory). The notation \(\text{Ind Dir.}\) refers to a possible set of directions retrieved from LTM or available in the current situation which probably result in lighting some lamp. Thus, the token \(\text{Ind Dir.}\) is not abstract but reflects specific semantic components, as suggested by Markovits et al. (1996).

In the same way, \(\text{Ind Colour}\) refers to values pertaining to the colours semantic space. As we will see, the directionality of the relation from Directions to Colours insures that \(\text{Ind Dir.}\) reflect values other than ‘left’, whereas \(\text{Ind Colour}\) does not exclude the value ‘green’.

In effect, this conception renders unnecessary the hypothesis of a representation of exhaustion in the representation of \(p\) in the initial model of \(\text{if } p \text{ then } q\). According to Johnson-Laird (1983), mental models for conditional are hypothetical models about possible states of affairs. Given that the conditional statement ‘if the ball rolls to the left, then the green lamp is lit’ suggests that the colour of the lit lamp depends on the trajectory of the ball, alternative models should represent additional hypothesis aimed to improve the information conveyed by the rule (see Sperber and Wilson, 1986). These alternative models should then represent the outcome of trajectories not referred to in the statement. However, there is no need to construct alternative models in which the ‘left’ direction reappears, because its consequence is already known (green lamp is lit). Therefore, other model with the value ‘left’ can not be an alternative hypothesis to ‘left – green’. The sign of directionality (\(\rightarrow\)) is devoted to capture the way models are constructed, i.e. from \(P\) to \(Q\), and not from \(Q\) to \(P\). Thus, exhaustion would be a by-effect of the entire process of constructing directional models from the sole available information: the left direction results on the green lamp lit. On the other hand, there is no reason that the same process implies

\[
\begin{align*}
\text{left} & \rightarrow \text{green} \\
\text{right} & \rightarrow \text{red}
\end{align*}
\]
systematic rejection of the value ‘green’ in constructing alternative models. Suppose that there are ten different possible directions for only two colours: it could be that another direction than ‘left’ results in the same effect (i.e. green lamp lit).

In summary, our suggestion is that mental models do not contain abstract truth-table-like components which require negation and exhaustion tags, but specific semantic components retrieved from LTM linked by directional relations. However, it is possible that such models could be reinterpreted and condensed in the more abstract format described by Johnson-Laird and Byrne (1991). Our intuition is that this last reinterpretive process would not be available until adulthood.

1.5. Pragmatic principles

In line with the mental models theory for syllogisms (Johnson-Laird, 1983), a mental models theory concerning the comprehension of the if connector should find a place for the effect of conversational principles. In particular, models should be structured in a way which reflects the principle of ‘quantity’ according to which speakers communicate all the useful information which they possess. In our example, the conditional statement is only meaningful if the result can be different from that described (i.e. green lit). In effect, if this is not the case, the result would no longer depend on the trajectory of the ball, and a statement such as ‘the green lamp always lights up’ would be more appropriate than the conditional statement.

It should be noted that formal logic does not require the existence of cases in which the consequent is not satisfied. If we suppose that the green lamp lights whatever happens, the statement ‘if the ball rolls to the left, then the green lamp is lit’ would remain true since there would be no other case in which the ball might have rolled to the left without the green lamp lighting. However, such a statement would not be informative since it would lead the listener to believe in the existence of a dependence between path and light which does not in fact exist. As a consequence, subjects should not represent a conditional by the alternatives \( p \rightarrow q \) and \( \neg p \rightarrow \neg q \). Thus, a conditional statement would inform the subject (a) of the existence of a link between two variables \( P \) and \( Q \) and (b) of the fact that the value of the variable \( Q \) indicated by the consequent \( q \) is not always verified. The interpretation of the statement would lead subjects to look for values of \( P \) which might result in alternative values of the consequent. As Sperber and Wilson (1986) have emphasised, subjects tend to choose a context which maximises relevance, i.e. which allows them to formulate new hypotheses. We suggest that the alternative new models represent these new hypotheses by assigning a value of \( Q \) to each value of \( P \). This explains the tendency to consider that the \( Ind \) value of the consequent in the second model of (6) must be a value other than ‘green’ (when the ball rolls to the left, the green lamp is lit, and other lamps will light for other directions). This would increase the informative nature of the statement but result in a biconditional interpretation of the connector.

Our second suggestion was that subjects attempt to increase the volume of information which they possess by constructing a ‘complete’ representation which links each of the \( P \) values to a specific value of \( Q \), thus establishing, wherever possible, a one-to-one correspondence. This suggestion is inspired by the results obtained by
Marcus and Rips (1979) and Markovits (1988). Using a similar experimental design to that used by Legrenzi, but in which the antecedent and consequent were able to assume more than two values, Marcus and Rips (1979) observed that subjects frequently tended to attribute a single value of $Q$ to each of the $P$ values. Markovits (1988) has also shown that two-thirds of subjects who exhibit a biconditional interpretation of an if $p$ then $q$ rule retained this interpretation even in cases where the observable data explicitly invalidated the possibility of a simple one-to-one correspondence (i.e. two different events result in the same consequence). This fact suggests the existence of a bias which leads subjects to adopt representations which simplify situations by establishing, wherever possible, a one-to-one correspondence between the possible values of the antecedent and the consequent.

This is compatible with the suggestions put forward by Sperber et al. (1995) who, basing their work on the relevance theory (Sperber and Wilson, 1986), suggested that subjects interpret statements in a way which maximises their relevance at the lowest possible cognitive cost. The establishment of a one-to-one correspondence maximises relevance since it permits the addition of new beliefs or hypotheses to those which are explicitly present in the statement: ‘if the ball rolls to the left, then the green lamp is lit’ and ‘if the ball rolls to the right then the red lamp is lit’. Moreover, this relevance is maximised at the least cognitive cost since a one-to-one correspondence implies that subjects construct only one model for each of the alternative possible values of the antecedent. This search for the least possible cognitive cost could result from the constraints inherent to constructing models in a capacity limited working memory.

When the statement involves two binary variables (e.g. left/right, green/red), subjects would establish links between alternative values not mentioned in the statement (i.e. right/red in the example above) in a second model. Since the obtained representation is complete (i.e. the values of $P$ and $Q$ are in bijection) the construction process would be halted, thus leading to the response pattern observed by Legrenzi (1970). In contrast, if this statement contains two non-binary terms, the number and type of the alternatives for $P$ and $Q$ are difficult to determine. If subjects have no prior knowledge of the fields of reference for the statement, the result would be an indeterminate alternative model which would be added to the initial model (see above).

To summarise, we believe that content effects are associated with the fleshing-out process of the initial model of if. In accordance with the mental models theory and relevance theory, this process would attempt to construct a representation of the situation described by the conditional statement at the least possible cognitive cost. This statement would be interpreted as a dependence relation between two variables $P$ and $Q$ of which one value $p$ and one value $q$ are explicitly linked by the conditional. The result would be an initial model representing the co-occurrence of $p$ and $q$. Subjects would attempt to complete this representation by constructing cases in which the alternative values of the variable $P$ are associated with the possible values of the variable $Q$. This process would not necessarily require the use of a propositional representation of negation as suggested by Johnson-Laird and Byrne (1991). The alternatives to the initial $p, q$ model would be constructed on the basis of LTM.
knowledge concerning the relation between $P$ and $Q$, the semantic nature of these variables, or the current situation to be processed. This construction would be guided by pragmatic principles (principles of ‘quantity’ and ‘completeness of representation’).

The use of the negation representation as well as the representation of the construction of the three models described by Johnson-Laird ($p.q$, $\neg p.q$, and $\neg p.q$) would therefore appear to be due to a level of understanding of the connector which is not achieved until adult age. Such a level is comparable to stage 4 of Moshman’s theory (1990) which corresponds to an explicit metalogic. In effect, the complete model of the conditional proposed by the standard mental model theory (i.e. a three-models representation with negation tag) presupposes an abstract formal representation of all the cases which are compatible with the conditional statement.

1.6. The present experiments

The aim of the experiments presented here was to demonstrate the existence of effects linked to the nature (binary or non-binary) of the variables involved in conditional statements which present arbitrary relations between variables whose semantic structure is known to subjects. The adopted approach is essentially developmental. As pointed out by Karmiloff-Smith (1992), ‘a developmental perspective is essential to the analysis of human cognition […] and how knowledge changes over time can provide subtle cues to its final representation format in the adult mind’ (p. 26). The study of the way these representations develop with age should make it possible to reveal effects which are masked by fully-formed adult functioning.

There are four possible cases depending on the variables presented in the antecedent and consequent: BB (i.e. binary antecedent and consequent), BN (binary antecedent and non-binary consequent), NB (non-binary antecedent and binary consequent) and NN. When one variable, either the antecedent or the consequent, is binary, the value opposed to that specified in the statement represents the content of the alternative model (the notation for this value is $/p$, which means opposite to $p$). Note that in $/p$, the ‘$/$’ symbol does not refer to a propositional-like tag which affects a representation of $p$. Rather, $/p$ stands for the representation of the sole affirmative value opposite to $p$. When the variable is not binary, the alternative model contains an indeterminate value of this variable (notation $\text{Ind}$, see Table 1). A number of predictions concerning the structure and nature of the models can be derived from

<table>
<thead>
<tr>
<th>Types of conditional rule</th>
<th>BB</th>
<th>BN</th>
<th>NB</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>First model</td>
<td>$p \rightarrow q$</td>
<td>$p \rightarrow q$</td>
<td>$p \rightarrow q$</td>
<td>$p \rightarrow q$</td>
</tr>
<tr>
<td>Second model</td>
<td>$/p \rightarrow \text{Ind}$</td>
<td>$/p \rightarrow \text{Ind}$</td>
<td>$\text{Ind} \rightarrow q$</td>
<td>$\text{Ind} \rightarrow \text{Ind}$</td>
</tr>
</tbody>
</table>

B, binary; N, non-binary.
the manipulation of the type of variable involved in the statement.

We have emphasised the need to consider alternatives to the consequent in order to ensure that \textit{if} possesses a conditional meaning. An alternative model ensures a non-conjunctive interpretation of \textit{if} (Johnson-Laird and Byrne, 1991). According to our hypothesis, this model should be easy to produce when the consequent \( q \) refers to a binary variable which facilitates the evocation of \( /q \). Earlier work has shown that children often interpret \textit{if} in a conjunctive way (Paris, 1973; Taplin et al., 1974; Barrouillet, 1997; J.F. Lecas and P. Barrouillet, unpublished data). This is due to the fact that they construct only one model \((p - q)\). Our hypothesis holds that this interpretation should be less frequent when the consequent is binary (BB and NB vs. BN and NN).

Let us now imagine that subjects have gone beyond the conjunctive interpretation and that the mental models correspond to those presented in Table 1. The type of variable involved in the statement should also influence the conditional or biconditional nature of the interpretation. BB statements should result in more biconditional interpretations than NN statements. In effect, the binary character of the two variables facilitates the production of the alternative model \( /p - /q \). That is, a determinate value of the variable \( P \) (i.e. \( p \) and \( /p \)) is associated with each of the two possible values of the variable \( Q \) (i.e. \( q \) and \( /q \) respectively). This means that it is likely that subjects consider this representation to be complete. Only the cases \( p.q \) and \( /p-/-q \) are then compatible with the statement, thus resulting in the predicted biconditional interpretation. In contrast, when an NN statement is presented, the contents of the second model remain indeterminate. Since no specific value of the consequent is explicitly associated with the alternative values of \( p \), the acceptance of the case \( /p.q \) as being compatible with the rule should be facilitated, thus resulting in a higher level of conditional interpretation.

None of these predictions can be derived from the classical mental model theory since in this the alternative cases are constructed on the basis of the negation of the values present in the initial model and there is nothing to state that the application of this negation depends on the nature of the terms to which it applies. However, our model predicts content effects which only a mental models theory seems able to explain. Indeed, the theories belonging to the mental logic approach (Rips, 1983; Rips, 1994; Braine, 1990; Braine and O’Brien, 1991) do not predict this type of effect. These theories hold that the meaning of the connector is formed from a set of inference schemas which are syntactic in nature. In essence, these schemas are insensitive to the characteristics of the terms involved in the statement.

However, Braine and O’Brien (1991) have suggested that these inference schemas are added to pragmatic principles such as Grice’s conversational implicatures or the invited inferences proposed by Geis and Zwicky (1971). In particular, an invited inference of the type ‘\textit{if} \( p \) then \( q \) therefore \textit{if not} \( p \) then not \( q \)’ would result in a biconditional interpretation of \textit{if}, unless the subjects had a reason to think this inference inappropriate (Geis and Zwicky, 1971; Fillenbaum, 1977). However, there is no reason to predict that the binary nature of the variables involved in the statement should facilitate the production of this inference, or that the absence of any such binary nature should constitute a reason for thinking that it should be avoided.
On the contrary, the model which we are proposing might be able to explain the production of this inference. Thus, the mental models theory would be the only one able to account for certain content effects, contrary to the assertions made by Bonatti (1994b).

We present three experiments to validate the predictions made above. In the first, children and adolescents aged 9–15 years were asked to identify cases which violated \( \text{if } p \text{ then } q \) rules. In the second, subjects of the same age were asked to perform a production task involving cases compatible with a conditional rule. A developmental approach should make it easier to demonstrate the existence of content effects and reveal effects which cannot be observed in adult subjects. For example, the presence of binary terms in the BB rules (vs. NN) should facilitate the production of explicit alternatives to the initial \( p, q \) model. Thus the conjunctive interpretation of if, which is frequent in children, should be rarer in the case of BB than in NN. Such an effect, which is inherent in the way mental models are constructed, can only be demonstrated in children. In a third experiment, adolescents and adults were asked to perform a conditional syllogism task. The number of models which should be constructed has an influence on the type of inference produced. To ensure greater clarity, the predictions made possible by the model proposed above will be presented separately for each of the tasks used.

2. Experiment 1

The task consisted of identifying which of the four cases \( p, q, \neg p, q, p, \neg q, \neg p, \neg q \) violated a rule of the type \( \text{if } p \text{ then } q \). We used a fully within-subjects design, in which the rules varied depending on the type of variable involved in \( p \) and \( q \) (BB, BN, NB and NN). The negated values (e.g. not red) were represented by a different value (e.g. green) and not by a statement of the type ‘the object is not red’.

The conditionals with a binary consequent (BB and NB) should result in fewer conjunctive interpretations than the others (i.e. BN and NN) because the binary character of \( Q \) facilitates the production of the alternative case \( \neg q \) and the formation of a second model. Since the task consisted of identifying the cases which violated the rule, the conjunctive interpretation took the form of rejecting all the cases apart from \( p, q \). Earlier results in children have shown that the case \( p, \neg q \) is nearly always rejected, and that when the case \( \neg p, \neg q \) is rejected, case \( \neg p, q \) is also rejected (Barrouillet, 1997; J.F. Lecas and P. Barrouillet, unpublished data). Thus when subjects consider that case \( \neg p, \neg q \) violates the rule, their interpretation is almost always conjunctive. We therefore predicted that our subjects would reject the case \( \neg p, \neg q \) less frequently for the BB and NB conditionals than for the BN and NN conditionals, because the former conditionals facilitate the production of an alternative model which renders the case \( \neg p, \neg q \) compatible with the rule.

If subjects produce at least one alternative model, case \( \neg p, \neg q \) is always compatible with this model whatever the type of rule (Table 1). In effect, when the rule contains binary terms, the case \( \neg p, \neg q \) necessarily matches the content of the alternative model \( (\neg p, \neg q) \). When one or both of the values are indeterminate
they are not incompatible with the case to be judged. For example, concerning the Legrenzi’s mechanism and the rule ‘if the ball rolls to the left, then the green lamp is lit’, the case ‘right – blue’ is a kind of not p. not q case which is not incompatible with a model containing Ind Dir. – Ind Colour as in (6).

In contrast, the case not p. q is incompatible with the first model constructed in both possible situations (i.e. p – q) since this model associates q with p and not with not p. Thus the case not p. q should be rejected more often than the case not p. not q. This explains the frequency of the biconditional interpretation of the if connector (i.e. rejection of the cases p. not q and not p. q only).

However, the rejection level of the case not p. q should vary as a function of the type of conditional presented. This case is doubly incompatible with the mental models associated with BB (i.e. p – q and /p – /q) and, although to a lesser extent, with those associated with NB. The alternative model produced for NB conditionals (Ind P – /q) is indeterminate for P and might be instantiated by not p. However, it also associates the value /q with not p. Thus, although this model does not explicitly link /p to /q as in the case of BB conditionals and does not result in an explicit rejection of not p. q, neither of the models produced is compatible with such a case.

In contrast, the case not p. q is not incompatible with the mental models resulting from BN and NN. In the case of BN conditionals, the alternative value of P is determinate (i.e. /p) and is not linked to any determinate value of Q (the second model is /p – Ind). Thus even though neither of the two models which are constructed corresponds explicitly to the case not p. q, this case is not incompatible with the model /p – Ind. In the case of NN conditionals, the value not p can instantiate the indeterminate value of P which is not linked to any determinate value of Q (in this case, the second model is Ind – Ind). Thus, although not explicitly represented, the case not p. q is not incompatible with either of these two models.

In consequence, not p. q should be rejected less frequently for BN and NN conditionals than for BB and NB conditionals. Of the latter, the BB conditionals should result in more rejections of not p. q since this case is doubly and explicitly incompatible with the constructed models.

In consequence, the conditionals with a binary consequent (i.e. BB and NB) should result in fewer conjunctive interpretations than those which have a non-binary consequent (BN and NN). However, during development, interpretations of the BB rules should remain essentially biconditional, due to the rejection of the not p. q case. In contrast, NN conditionals should result in more conjunctive interpretations in very young subjects. However, once the subjects have gone beyond this interpretation, the majority of interpretations should be conditional in nature.

2.1. Method

2.1.1. Participants

Thirty-two pupils taken from each of grades 3 (mean (M) = 8.2 years, SD = 0.4 years), 6 (M = 11.3 years, SD = 0.4 years) and 9 (M = 15.0 years, SD = 0.3 years) took part in the experiment.
2.1.2. Materials and procedure

For each of the four types of if...then rule (i.e. BB, BN, NB, and NN), two rules were presented, preceded by an introductory scenario (see ). The rules were presented in writing and simultaneously read aloud by the experimenter, who then presented the subject with a board representing the four cases: \( p \land q \), \( \neg p \land q \), \( p \land \neg q \), and \( \neg p \land \neg q \) by means of drawings. For example, for the rule 'if the part (produced by the machine) is pierced, then it is a square', the presented cases took the form of a pierced square \( (p \land q) \), an unpierced square \( (\neg p \land q) \), a pierced circle \( (p \land \neg q) \), and an unpierced circle \( (\neg p \land \neg q) \). The subjects had to answer the question 'are there any of these cases that do not respect the rule?' and mark the illegal case or cases. Three distractor rules involving the connectors \( \text{and} \), \( \text{or} \), and \( \text{if and only if} \) were added to the eight experimental rules. The order of the presentation of the four rule-types and the order of appearance of the two rules within each type were counterbalanced. The order of presentation of the four cases to be judged on the boards was random. The experiment was administered individually with no time constraints.

2.2. Results

Table 2 presents the frequency of rejection for each case (i.e. identification of this

<table>
<thead>
<tr>
<th>Grade</th>
<th>BB</th>
<th>BN</th>
<th>NB</th>
<th>NN</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Third</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p \land q )</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \neg p \land q )</td>
<td>98</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>( \neg p \land \neg q )</td>
<td>63</td>
<td>69</td>
<td>58</td>
<td>80</td>
<td>67</td>
</tr>
<tr>
<td>( \neg p \land q )</td>
<td>92</td>
<td>88</td>
<td>88</td>
<td>97</td>
<td>91</td>
</tr>
<tr>
<td>Sixth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p \land q )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \neg p \land q )</td>
<td>100</td>
<td>97</td>
<td>100</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>( \neg p \land \neg q )</td>
<td>30</td>
<td>41</td>
<td>25</td>
<td>38</td>
<td>33</td>
</tr>
<tr>
<td>( \neg p \land q )</td>
<td>81</td>
<td>77</td>
<td>64</td>
<td>63</td>
<td>71</td>
</tr>
<tr>
<td>Ninth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p \land q )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \neg p \land q )</td>
<td>98</td>
<td>98</td>
<td>100</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>( \neg p \land \neg q )</td>
<td>6</td>
<td>14</td>
<td>11</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>( \neg p \land q )</td>
<td>61</td>
<td>33</td>
<td>44</td>
<td>44</td>
<td>45</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p \land q )</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>( \neg p \land q )</td>
<td>99</td>
<td>98</td>
<td>100</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>( \neg p \land \neg q )</td>
<td>33</td>
<td>41</td>
<td>31</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>( \neg p \land q )</td>
<td>78</td>
<td>66</td>
<td>65</td>
<td>68</td>
<td></td>
</tr>
</tbody>
</table>
As violating the conditional rule) as a function of the type of conditional and Table 3 indicates the type of interpretation resulting from these rejections. First of all, the results confirmed the developmental pattern of the interpretation of the connector which had already been observed for this type of task, developing from a conjunctive interpretation (rejection of all the cases except \( p \land q \), 64% of responses at grade 3) to a conditional interpretation (rejection of the case \( p \land \neg q \) only, 52% of responses at grade 9). Here, the biconditional interpretation represents an intermediate step (38% of responses at grade 6, see Table 3). This development conforms to the predictions of the mental models theory with increasing processing capacity making it possible to construct an ever greater number of models (one model for the conjunctive interpretation, two for the biconditional, and three for the conditional).

Our first hypothesis concerned the effect of rule type on the frequency of rejection of the not \( p \), not \( q \) case and the conjunctive interpretation which results from it. We performed a 3(grade: third, sixth, and ninth) \( \times \) 2(nature of the consequent: B or N) ANOVA with repeated measures on the last two factors for the rejection level for not \( p \), not \( q \). The only significant main effects were those associated with grade \( F(2,93) = 21.83, P < 0.001 \), Mean Square error (MSe) = 1.81, and the type of consequent \( F(1,93) = 25.77, P < 0.001 \).
MSe = 0.18, and there was no significant interaction. The number of rejections of the case $\neg p \land \neg q$ fell as academic level increased (67%, 33% and 12% for grades 3, 6 and 9, respectively). As predicted by the hypothesis, conditionals with a binary consequent (BB and NB) resulted in significantly fewer rejections of the case $\neg p \land \neg q$ (32.8% for BB, and 31.3% for NB) than conditionals with non-binary consequent (41.1% for BN, and 44.8% for NN). This effect was significant at all the academic levels investigated, grade 3: $F(1,31) = 9.85, P < 0.02, MSe = 0.26$, grade 6: $F(1,31) = 13.65, P < 0.001, MSe = 0.13$, grade 9: $F(1,31) = 4.25, P < 0.02, MSe = 0.15$. The variability of the rejection level of $\neg p \land \neg q$ as a function of the conditional type, resulted in more conjunctive responses (i.e. rejection of all cases except $p \land q$) for the BN and NN conditions (37 and 43.8%) than for the BB and NB conditions (30.7 and 29.7%) as the hypothesis predicted.

Our hypothesis also predicted a conditional type effect on the rejection level of the case $\neg p \lor q$ and, in consequence, on the number of biconditional response patterns. We expected biconditional patterns be more frequent in the BB and NB conditions than in the BN and NN conditions. Moreover, we expected more biconditional responses in the BB than in the NB condition because the case $\neg p \lor q$ is explicitly incompatible with the two mental models constructed in the BB condition (see Table 1). An analysis of variance using a similar design to that described above was conducted for the biconditional response pattern. The only significant main effects were those of the antecedent type $F(1,93) = 8.29, P < 0.01, MSe = 0.25$, and the consequent type $F(1,93) = 25.85, P < 0.001, MSe = 0.34$. As expected by the hypothesis, the biconditional interpretations were more frequent when the consequent was binary (46% and 35% for BB and NB respectively) than when it was non-binary (28% and 24% for BN and NN respectively). The conditionals with a binary antecedent also resulted in more biconditional interpretations than the others. This antecedent type effect, which was weaker than the consequent type effect, was primarily due to the fact that the BB rules resulted, as expected, in more biconditional interpretations (46%) than the NB rules (35%), $F(1,93) = 9.06, P < 0.01, MSe = 0.25$.

The results are confirmed if we consider only the non-conjunctive responses. In effect, the influence of the conditional type on the rejection of $\neg p \lor q$ cases cannot be observed unless the alternative models are constructed, i.e. if subjects go beyond the conjunctive interpretation. For the BB, NB, BN and NN conditions, the level of biconditional responses represented 67, 50, 44 and 43% of non-conjunctive responses respectively. This corresponded exactly to the predictions made possible by the model.

To summarise, the results fully confirmed the predictions derived from the model. The conditionals with a binary consequent (BB and NB) facilitated the production of an alternative model and resulted in fewer conjunctive interpretations than the others. Even in grade 9, where the conditional interpretation was the most frequent taken over all the
statements (52% of responses), biconditional interpretations remained the most frequent in the BB condition (55%). Thus the interpretations prompted by the various conditional rules varied depending on the nature of the antecedent and consequent.

3. Experiment 2

The purpose of this experiment was to examine the same hypotheses as were tested in Experiment 1. In this task, the subjects were asked to construct as many cases as possible that were compatible with a rule of the type \( \text{if} \ p \ \text{then} \ q \) using objects or drawings which concretised the instances \( p \) or \( \neg p \) on the one hand and \( q \) or \( \neg q \) on the other. As for Experiment 1, the type of rule (BB, NB, BN, and NN) was used as a within subjects factor. The hypotheses predicted that (a) production of the cases of the type \( p, q \) only (conjunctive interpretation) would be more frequent for the NN and BN rules than for the NB and BB rules, and (b) when the cases of the type \( \neg p, \neg q \) are produced, the production of cases of the type \( \neg p, q \) should be more frequent with NN and BN rules than with BB and NB rules. Among the latter, the production of \( \neg p, q \) cases should be rarer for BB than for NB. Thus, the NN condition should be the one to generate the most conjunctive interpretations. Biconditional interpretations should predominate in the BB condition whereas conditional interpretations should be the most frequent for the NN rule.

3.1. Method

3.1.1. Participants

Thirty-two subjects from each of the grades 3 (\( M = 9.0 \) years, SD = 0.3 years), 6 (\( M = 11.7 \) years, SD = 0.3 years) and 9 (\( M = 14.9 \) years, SD = 0.3 years) took part in the experiment. None of the participants of the former experiment took part in this experiment.

3.1.2. Materials and procedure

Eight rules of the type \( \text{if} \ p \ \text{then} \ q \) (two for each of the four types of rule BB, BN, NB, and NN) were presented, preceded by an introductory scenario (see ). Three distractor rules were added, as in Experiment 1. The material took the form of four sets of drawings or objects concretising the instance \( p \) or \( \neg p \) for the antecedent, and \( q \) or \( \neg q \) for the consequent. The subjects were told that they could construct cases by associating a possible value to the antecedent and a value to the consequent.

For example, for the rule ‘if you put on a white shirt then you will put on green trousers’, the material consisted of three white shirts (\( p \)), six yellow shirts (\( \neg p \)), six green trousers (\( q \)) and three blue trousers (\( \neg q \)) which the subjects had to link to one another. The number of objects of each type was calculated in such a way that the construction of the \( p, q \) cases followed by the \( \neg p, \neg q \) cases (order of construction followed by almost all subjects in preceding experiments) would leave an equal number of \( \neg p \) and \( q \) instances (three of each) over for the possible construction of
the not p, q cases. When subjects had completed the task and declared that there was no more possible case, the experimenter presented them with a number of instances of p and not q in order to check that the cases resulting from their conjunction (i.e. p, not q) were rejected, as the correct comprehension of the rule which they were manipulating implies. The instruction specified that subjects should construct as many cases as possible which were compatible with the presented rule. The order of presentation of the four types of rule (i.e. BB, BN, NB and NN) was counter-balanced and the number of objects proposed varied between rules. The experiment was administered individually with no time constraints.

3.2. Results

The responses were categorised on the basis of the interpretation which underpin them: conjunctive (production of p, q case only), biconditional (p, q and not p, not q cases), and conditional (p, q, not p, not q, and not p, q cases, see Table 4). This experiment revealed a developmental pattern identical to that observed in Experiment 1, with the number of conjunctive responses falling (47%, 8% and 6% for grades 3, 6, and 9 respectively) and the number of conditional responses increasing (5%, 21%, and 45%).

Firstly our hypothesis predicted that the conditionals with a non-binary consequent (BN and NN) would result in more conjunctive responses than conditionals with a binary consequent. We performed a 3(grade: third, sixth and ninth) × 2(nature
of consequent: B or N) \times 2 (nature of antecedent: B or N). ANOVA with repeated measures on the last two factors for the number of conjunctive responses. The grade effect was significant, \( F(1,93) = 29.90, P < 0.001, \) MSe = 0.93, and reflected the above-mentioned fall in the number of conjunctive responses with increasing age. In conformity with the hypothesis, the consequent type had a significant effect, \( F(1,93) = 21.52, P < 0.001, \) MSe = 0.15. The rules containing a non-binary consequent (i.e. BN and NN) provoked more conjunctive responses (25% of responses) than the rules with a binary consequent (16%). However, the antecedent type also had a significant effect, with rules containing a non-binary antecedent resulting in more conjunctive responses (25%) than rules with a binary antecedent (18%), \( F(1,93) = 19.95, P < 0.001, \) MSe = 0.13. The interaction was not significant, \( F(1,93) = 2.52, P = 0.12, \) MSe = 0.17, suggesting that the effect of these two factors is additive. Nevertheless, the effect of antecedent type was weaker than that of consequent type (41 vs. 52% of the effect due to crossing these two factors).

Thus, the presence of a binary antecedent or consequent made it easier for subjects to go beyond the conjunctive interpretation and produce \( \neg p, \neg q \) cases. This finding is compatible with our model in that binary terms facilitate the explicit representation of an alternative value to that supplied in the rule and induce subjects to construct an alternative model. When both terms were non-binary (NN rule), approximately one-third of interpretations (31%) were conjunctive.

Secondly our hypothesis predicted that when subjects go beyond the conjunctive interpretation, the rules with a binary consequent should primarily give rise to biconditional interpretations. Since the alternative to the consequent is unique and explicit (i.e. \( \neg q \)), this value should instantiate the consequent of the alternative model. Subjects would therefore construct two models. In contrast, the rules with a non-binary consequent should result in a conditional interpretation. We therefore conducted an ANOVA with the same design as that described above for the number of observed biconditional interpretations. The grade effect was significant, with the level of biconditional interpretations increasing from grade 3 (47%) to grade 6 (71%) and then falling again at grade 9 (49%), \( F(2,93) = 6.06, P < 0.01, \) MSe = 1.45. As expected by the hypothesis, rules with a binary consequent resulted in more biconditional interpretations (68%) than the rules with a non-binary consequent (43%), \( F(1,93) = 121.69, P < 0.001, \) MSe = 0.20. The antecedent type effect was also significant, \( F(1,93) = 100.84, P < 0.001, \) MSe = 0.19, with rules containing a binary antecedent resulting in more biconditional interpretations (67%) than the rules with a non-binary antecedent (43%). The interaction between the two factors was significant \( F(1,93) = 6.38, P = 0.01, \) MSe = 0.34, with the consequent type effect being greater for rules with a non-binary antecedent (NN: 28% vs. NB: 61%) than for rules with a binary antecedent (BN: 58% vs. BB: 76%). In consequence, the level of biconditional interpretations was all the greater the more binary terms there were present in the rule. Thus, 88% of the non-conjunctive interpretations were biconditional for the BB rules, whereas this level was only 41% for the NN rules. Thus, the nature of the terms involved in the statement considerably affected the type of interpretation evoked, with BB rules primarily inducing bicon-
ditional interpretations (76%) whereas the NN rules induced the greatest number of conjunctive interpretations (31%) and conditional interpretations (41%).

These results conformed to the model presented above, but also indicated that in the production task the nature of the antecedent was extremely important. This might be explained by the fact that the presence of a binary term, either in the antecedent or in the consequent, facilitates the production of an alternative term (/p or /q) and therefore the construction of an alternative model which allows subjects to go beyond the conjunctive interpretation which is based on the construction of a single model. In contrast, when an alternative model is constructed, the chance that fleshing-out will be blocked in the second model is all the greater the more binary terms the rule contains.

4. Experiment 3

The aim of this experiment was to test the prediction of the model concerning the production of inferences. In the conditional inference task, the subjects were given the opportunity to draw any of the following four inferences:

- modus ponens (MP) true antecedent implies true consequent
- Denial of antecedent (DA) false antecedent implies false consequent
- affirmation of consequent (AC) true consequent implies true antecedent
- modus tollens (MT) false consequent implies false antecedent.

In logic, only the forms MP and MT are considered to be valid whereas DA and AC are considered to be invalid. However, subjects endorse them and produce them frequently (Evans et al., 1993). In conformity with our model, the level of acceptance of the denial inferences (i.e. DA and MT), which require fleshing-out of the first model, should vary as a function of the nature (B vs. N) of the terms involved in the statement. This experiment presented only rules of the types BB and NN from which we expected to generate the following representations (7):

\[
\begin{array}{c|c|c}
\text{B/B} & \text{N/N} \\
\hline
p & q & p & q \\
\hline
/p & /q & \text{IndP} & \text{IndQ}
\end{array}
\]

The BB rules should facilitate the production of an explicit, alternative model (i.e. /p – /q) but should also block the fleshing-out process at this level, thus leading to the construction of two explicit models. These rules should therefore result in a high level of inference production based on the second model (i.e. MT and DA) and these levels should remain relatively stable irrespective of age. In contrast, the NN rules make the production of an alternative model difficult for young subjects who construct only the first model (see Experiments 1 and 2). Thus, among these subjects, MT and DA inferences based on the second model should be less frequent than the MP and AC inferences based on the first model. When the second model is constructed, the values remain indeterminate in the NN condition and calculations...
concerning these values (i.e. making the relation explicit) probably requires considerable levels of cognitive resources. Thus, MT and DA inferences should be produced less frequently in NN than in the BB condition.

The frequency of MT production should increase with age since this inference is possible on the basis of a single alternative model. The case of the DA inference is more complex. The difficulty of constructing an alternative model in the NN condition should impair the production of DA in the youngest subjects. Next, its frequency of production should increase when only two models are constructed, and then decrease when the adult subjects are able to construct more than two models and imagine that an alternative value of the antecedent can be linked to the value of the consequent presented in the rule (i.e. case of not p, q), thus blocking the production of DA. Thus, the mental models theory permits a counterintuitive prediction. In adults the frequency of production of DA in the NN condition should be close to its level in the youngest subjects but for different reasons: in younger subjects, because the model capable of supporting DA is rarely constructed, and in adults because subjects are able to construct more than two models. To summarise: (a) the explicit nature of the alternative model induced by the BB rules should facilitate production of denial inferences DA and MT and (b) developmental effects should be weaker for the BB than for the NN rules.

Moreover, the type of negation used in the minor premise should result in specific effects. The negation of a term (e.g. the card is red) may be implicit (e.g. the card is green) or explicit (i.e. the card is not red) in nature. In formal approaches to reasoning, effects linked to the type of the negation are difficult to take into account. However, this is not the case for our model. In effect, in our model, the minor premise forms part of the context of interpretation (Sperber et al., 1995). The type of negation it contains modifies the interpretation of the statement if p then q by providing an indication of the nature of the relevant alternative values. Let us imagine that subjects have to reason about a statement such as ‘if it’s a square then it’s red’. The values which can be assumed by the Colour variable therefore depend on the value which can be assumed for the Shape variable. If the second premise is ‘it’s a circle’, this informs the subjects that many different possible values of the Shape variable have to be taken into consideration, probably with many different values for the colour variables. In contrast, if the second premise is ‘it’s not a square’, this indicates that only the opposition square/not square is relevant. Thus, the explicit negation results in a binary Shape variable since it only offers a single alternative for consideration. On the other hand, this is not the case for implicit negation which opens the way for multiple alternatives (i.e. circle, but also triangle, rectangle, lozenge, etc.).

The type of negation should not have any effect on the BB rules since here the implicit and explicit negations are equivalent (not married equals bachelor). In contrast, for the NN rules, the two types of negation are not equivalent (i.e. not red is not the same as green). The binary structure suggested by the explicit negation should induce a biconditional interpretation (see Experiment 2) which should be characterised by the more frequent production of DA and MT. In effect, not p (or not q) incorporated in the Ind value of the variable P (or Q) and the Ind value corre-
sponding to the variable \( Q \) (or \( P \)) should be interpreted as not \( q \) (or not \( p \)). In contrast, an implicit negation of \( p \) (or of \( q \)) makes one of the possible values of the alternative model of \( P \) (or \( Q \)) explicit. The indeterminate nature of the alternative model corresponding to \( Q \) (or \( P \)) should make it difficult to produce a precise value. In consequence, the conclusion ‘nothing follows’ should be more frequent. To summarise, the type of negation acting on the minor premise (implicit vs. explicit) should interact with the rule type (BB vs. NN).

Finally, the developmental approach should allow us to test one of the main points of the mental models theory. According to Johnson-Laird’s model, the production of DA necessitates the construction of two models and should therefore be less frequent than that of the AC which can be produced on the basis of a single model. In fact, however, the frequency of production or acceptance of AC does not seem to be greater than that of DA (see Evans et al. (1993) for a review). O’Brien et al. (1994) emphasise this point (see also Evans, 1993) and use it as an argument against the mental models theory.

The mental models theory predicts that, because of their limited processing capacity, young children tend to construct only one model \((p – q)\) and that development, by increasing this capacity, makes the construction of alternative models easier and more likely (Barrouillet, 1997). Moreover, the construction of the second model should be facilitated by the BB type rules. Thus, AC (which necessitates a single model) should be produced more frequently than DA (which necessitates two models) by young subjects and this difference should tend to disappear with development. In addition, this interaction should be more pronounced for NN rules (especially for implicit negation) than for BB rules.

4.1. Method

4.1.1. Participants

Twenty-four pupils from each of the grades 6 (\( M = 12.1 \) years, SD = 0.4 years), 9 (\( M = 15.0 \) years, SD = 0.4 years) and 24 adults in their first year at the Université d’ Bourgogne (\( M = 22.5 \) years, SD = 1.4 years) took part in the experiment.

4.1.2. Materials and procedure

We constructed 24 rules of the form if \( p \) then \( q \), presenting an artificial relation between the antecedent and the consequent: 12 rules for which the terms \( p \) and \( q \) were non-binary (NN rules) and 12 rules in which the terms were binary (BB rules, see ). Of the BB rules, eight contained a term for which there was a semantic opposite (e.g. boy/girl) for the antecedent and the consequent. The other four contained binary propositions the negation of which could not be expressed in terms of a semantic opposite. For example, the variable ‘presence of a lift’ is binary (i.e. there is or there is not a lift), the negation of ‘there is a lift’ being ‘there is no lift’. This latter type of rule was used in order to verify whether the effects associated with the binary variables were specifically due to the existence of a semantic opposite or to the fact that only one alternative is possible. We formed no hypothesis regarding this question.
These rules were presented accompanied by a minor premise making it possible to produce one of the four inferences MP, AC, DA, or MT. In the case of the DA and MT inferences, the minor premise (i.e. not \( p \) and not \( q \) respectively) contained either an explicit negation (‘the pupil is not a boy’), or an implicit negation (‘the pupil is a girl’) when this was possible.

In the NN situation each participant studied 12 pairs of premises (determined by the 12 NN rules) from which it was possible to produce two MP inferences, two AC, two implicit DA, two explicit DA, two implicit MT, and two explicit MT inferences. In the BB situation involving a semantic opposite (eight rules), the material allowed subjects to produce one MP, one AC, one explicit DA, two implicit DA, one explicit MT, and two implicit MT inferences. In the BB situation without a semantic opposite (four rules), the material made possible the production of one MP, one AC, one explicit DA, and one explicit MT inference. Thus, each subject was able to produce an inference for each of the 24 rules. The MP, AC, explicit DA, implicit DA, explicit MT, and implicit MT inferences could be produced four times each, twice in the NN and twice in the BB condition, and the number of situations supporting a valid (i.e. MP or MT) or invalid (i.e. AC or DA) inference was the same.

For each type of rule (NN and BB) and for each subject, the rules were associated with each inference type (MP, AC, DA or MT) by circular permutation in such a way that each rule was presented an equal number of times for each type of inference. The order of presentation of the rules was random. Eight situations which permitted the production of a disjunctive syllogism (\( a \) or \( b \) and not \( a \), what follows?) were used as distractors.

The task was administered by groups. Each subject received a 32-page notebook with one problem presented on each page. Each 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 syllogisms, that the conclusion was a new item of information which could be inferred with certainty from the information provided by the premises. Participants were not subject to any time constraints.

4.2. Results

An ANOVA was performed on the number of denial inferences (i.e. MT and DA) produced with 3(levels: sixth grade, ninth grade, and adults) \( \times \) 2(types of rule: BB and NN) \( \times \) 2(types of negation: implicit and explicit) \( \times \) 2(types of inference: DA and MT), with repeated measures on the last three factors. As predicted by our hypothesis, the main effect of rules was significant, DD rules eliciting more denial inferences (67%) than NN rules (50%), \( F(1,69) = 33.35, P < 0.001, MSe = 0.51 \) (Table 5). This effect was true for both DA (BB: 60%, and NN: 44%) and MT (BB: 75%, and NN: 57%). The effect of rule interacted with the level, \( F(2,69) = 10.69, \)
As predicted by the model, the level interacted significantly with the type of rule for MT inferences, $F(2,69) = 13.17, P < 0.001$, MSe = 0.44. There was a strong developmental effect for NN rules (29%, 67%, and 74% in sixth grade, ninth grade and adults respectively), whereas it disappeared on BB rules (75%, 77%, and 73% in sixth grade, ninth grade and adults respectively). Concerning DA inferences, the predicted quadratic trend appeared on NN rules (sixth grade: 32%, ninth grade: 64%, and adults: 35%), $F(1,69) = 11.38, P < 0.001$, MSe = 0.99, but also for BB rules (59, 73, and 47% respectively), $F(1,69) = 5.32, P < 0.001$, MSe = 0.94. As predicted by mental models theories, the rates of DA production for NN rules were close in sixth grade and adults (i.e. 32 and 35%), whereas the rate of MT was higher in adults (74%) than in sixth grade (29%).

In line with our hypothesis, explicit negation elicited more denial inferences (62%) than implicit negation (55%), $F(1,69) = 9.81, P < 0.01$, MSe = 0.27, and this effect interacted significantly with the type of rule, $F(1,69) = 17.77, P < 0.001$, MSe = 0.25. The effect of negation was virtually nil for BB rules (implicit negation, 68%, and explicit, 66%), whereas explicit negation elicited more inferences (58%) than implicit negation (42%) on NN rules. This interaction was observed both for MT (BB rules: 74% for explicit and 76% for implicit negation, NN rules: 64% and 49% respectively) and on DA inferences (BB rules: 59% for explicit and 60% for implicit negation, NN rules: 52% and 35% respectively). The Negation × Rule × Inference interaction was not significant, $F < 1$.

At grade 6, the BB rules resulted in more denial inferences than the NN rules, irrespective of whether the negation was explicit or implicit (in both cases, $P < 0.001$). This can be explained by the ease with which subjects were able to construct an alternative model in the BB condition. At grade 9 and in the adult participants, there was no difference between the two types of rules when they mobilised an explicit negation ($F < 1$). In the case of implicit negations, the BB rules continued to provoke more denial inferences than the NN rules, both in grade 9 ($P < 0.01$) and adult participants ($P < 0.05$).

As far as the relative frequency of production of AC and DA for the NN rules is concerned, AC was more frequent (75%) than DA (32%) in grade 6, with this difference diminishing at grade 9 (77 and 64% respectively) and in adult subjects (46 and 35%). As expected, this difference was accentuated by the use of implicit negations (75 and 25% in grade 6, 77 and 54% in grade 9, and 46 and 27% in adult subjects). As the model predicted, the difference between AC and DA was less for BB than for NN rules from grade 6 onwards (75%/32% for NN vs. 79%/59% for DD) and disappeared altogether at grade 9 (71 and 73%) and in adult subjects (50 and 47%). These results confirmed the predictions derived from the model theories: the inference DA is more difficult to produce than AC when the production of an explicit second model is impaired by the nature of the rules (i.e. NN) or by the subject’s limited capacity (e.g. in children).

Unfortunately, the contrast between BB rules with and without a semantic opposite was uninformative. In effect, the BB rules without a semantic opposite resulted in fewer inferences than the rules with a semantic opposite for DA (51 vs. 67%) and
MT (69 vs. 78%). These percentages were close to those obtained in the NN condition (52 and 64% for DA and MT, respectively). This might suggest that the absence of a semantic opposite hindered the construction of a second model and thus the production of the inferences which it permits in the same way as the NN condition. However, the BB rules without a semantic opposite induced fewer AC and MP inferences than the others. The MP inference was produced in 72% of cases, as against 97% in DD with a semantic opposite and 96% in the NN condition. The same phenomenon affected the production of AC (60, 74, and 66% respectively).

Thus, it was not so much the inferences supported by the second model (i.e. DA and MT) that were produced less frequently in the absence of a semantic opposite, but rather all the possible inferences including the MP and AC inferences. We possess no explanatory hypothesis to account for the lower level of inference production for this particular type of rule.

Globally, the results of this experiment therefore confirmed the predictions derived from our model. The MT and DA inferences were facilitated by the BB rule. The type of the negation (implicit vs. explicit) had an effect in conjunction with NN rules but not with BB rules, the two types of negations being equivalent for a binary term. The developmental effects were attenuated when the problems involved BB rules. However, the effect of rule-type (BB vs. NN) diminished with development and in adults was only evident for the DA condition. This suggests that adults process the two types of rule in the same way and possess the capacity necessary to construct more than two models, since DA and AC were produced less frequently than at grade 9. The results which we obtained therefore conform better to Johnson-Laird’s standard model than to our own.

There are two possible explanations for this phenomenon. Firstly, as we mentioned in our introduction, with development subjects might be able to construct

<table>
<thead>
<tr>
<th>Sixth grade</th>
<th>Ninth grade</th>
<th>Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB</td>
<td>NN</td>
<td>BB</td>
</tr>
<tr>
<td>MT explicit</td>
<td>79</td>
<td>40</td>
</tr>
<tr>
<td>MT implicit</td>
<td>71</td>
<td>19</td>
</tr>
<tr>
<td>MT mean</td>
<td>73</td>
<td>29</td>
</tr>
<tr>
<td>DA explicit</td>
<td>58</td>
<td>40</td>
</tr>
<tr>
<td>DA implicit</td>
<td>60</td>
<td>25</td>
</tr>
<tr>
<td>DA mean</td>
<td>59</td>
<td>32</td>
</tr>
<tr>
<td>AC</td>
<td>79</td>
<td>75</td>
</tr>
<tr>
<td>MP</td>
<td>79</td>
<td>90</td>
</tr>
</tbody>
</table>

*BB refers to rules with binary antecedent and consequent, NN to rules with both non-binary antecedent and consequent.*

MT (69 vs. 78%). These percentages were close to those obtained in the NN condition (52 and 64% for DA and MT, respectively). This might suggest that the absence of a semantic opposite hindered the construction of a second model and thus the production of the inferences which it permits in the same way as the NN condition. However, the BB rules without a semantic opposite induced fewer AC and MP inferences than the others. The MP inference was produced in 72% of cases, as against 97% in DD with a semantic opposite and 96% in the NN condition. The same phenomenon affected the production of AC (60, 74, and 66% respectively). Thus, it was not so much the inferences supported by the second model (i.e. DA and MT) that were produced less frequently in the absence of a semantic opposite, but rather all the possible inferences including the MP and AC inferences. We possess no explanatory hypothesis to account for the lower level of inference production for this particular type of rule.

Globally, the results of this experiment therefore confirmed the predictions derived from our model. The MT and DA inferences were facilitated by the BB rule. The type of the negation (implicit vs. explicit) had an effect in conjunction with NN rules but not with BB rules, the two types of negations being equivalent for a binary term. The developmental effects were attenuated when the problems involved BB rules. However, the effect of rule-type (BB vs. NN) diminished with development and in adults was only evident for the DA condition. This suggests that adults process the two types of rule in the same way and possess the capacity necessary to construct more than two models, since DA and AC were produced less frequently than at grade 9. The results which we obtained therefore conform better to Johnson-Laird’s standard model than to our own.

There are two possible explanations for this phenomenon. Firstly, as we mentioned in our introduction, with development subjects might be able to construct
representations which are more abstract than those postulated by our model (Mar-
kovits, 1993) or which mobilise negation signs, as proposed by Johnson-Laird and
Byrne (1991). These abstract representations might be the result of a process of
reflective abstraction operating on earlier forms of reasoning, as Moshman (1990)
suggested. However, the same phenomena might be explained by an increase in
processing capacity which allows adults to go beyond the construction of the two
models induced by the BB rules without modifying the nature of the representations
involved. Neither of these interpretations conflicts with the general framework of the
mental models theory.

5. General discussion

The results of the three experiments conformed to the predictions of the model.
The construction of mental models depended on the type of term involved in the
statement: the binary terms facilitated the production of an alternative second model
but hindered the latter fleshing-out process. This frequently resulted in a bicondi-
tional interpretation of the BB rules at all the ages observed (Experiments 1 and 2).
The ease of production of this second model also resulted in the more frequent
production of the DA and MT inferences which, recording to the mental models
theory, are underpinned by a second model (Experiment 3).

The developmental approach allowed us, on the one hand, to demonstrate clearly
the content effect predicted by the model and, on the other, to reveal a development
in the interpretation of conditional statement with age. During the first stage, these
statements generated an interpretation close to the conjunctive one: rejection of
all cases apart from \(p \land q\) as being incompatible with the statement, construction of
only this case in the production task and more frequent production of MP and
AC inferences than DA and MT inferences. During the second stage, the interpre-
tation was biconditional (rejection of the case \(\neg p \land q\) and \(\neg p \land \neg q\), construction of
cases \(p \land q\) and \(\neg p \land \neg q\)), and all the inferences (MP, AC, DA, and MT)
were produced frequently. During the third stage, the conditional interpretation
prevailed. This developmental pattern was observed irrespective of the task used.
It can be simply explained within the framework of the mental models theory:
The increase of processing capacity with age (Case, 1985; Pascual Leone, 1988;
Halford, 1993; Swanson, 1996) allows subjects to construct an increasingly large
number of models. The conjunctive stage, which appears at about 9 years, is
explained by the formation of a single model. Similarly, Johnson-Laird et al.
(1986) have shown that 9-year-old children were only able to resolve the syllogisms
which required the construction of a single model. The biconditional stage is
explained by the construction of a second model which contains values opposed
to those present in the initial model while the final stage is made possible by
the construction of additional models which allow subjects to conceive the
possibility of the \(\neg p \land q\) cases unless the contents of the rule to be processed hinder
this construction (e.g., BB rules). We believe that only Johnson-Laird’s mental
models theory is able to explain the entirety of the observed results in such an
In the remainder of this article, we shall demonstrate that the model presented in the introduction also makes it possible to account for numerous effects described in the literature. First of all, let us return to the problem of content effect.

5.1. Theories of reasoning and content effects

Bonatti claimed that the standard mental models theory was unable to account for the way in which pragmatics and general knowledge affect the interpretation of statements since this theory takes the form of a variation of the truth tables principle, ‘and truth tables are only sensible to truth values, not to content connections relevance’ (Bonatti, 1994a; Bonatti, 1994b). Although this argument is correct (see Section 1), the model that we have put forward makes it possible to account for such effects. Now, the clarity of Bonatti’s argumentation reduces the heuristic character of the mental logic hypothesis.

Based on the studies conducted by Braine (Braine et al., 1984; Braine and O’Brien, 1991), Bonatti (1994b) holds that two distinct processes are involved in problem solving, the first being a comprehension and the second a reasoning process. After having acknowledged that mental logic is unable to explain the role of contents in reasoning (Bonatti (1994b) p. 20), he attempts to show that the same is true of the mental models theory. Within the framework of mental logic, an initial process is thought to provide a syntactic analysis of the linguistic signal and the identification of its logical form together with an initial semantic analysis which covers its literal meaning. Next, pragmatic principles and general knowledge are thought to help subjects select a particular logical form as the input signal to the rule system. At this point ‘representations possibly sharply different from the first semantic analysis are passed onto a processor blind to content and pragmatics’ (Bonatti (1994b) p. 20). Thus, according to Bonatti, the input to the inference rules is provided by a comprehension process every bit as complete as that which he claims to be necessary for the construction of mental models. Moreover, Bonatti emphasises that in the mental models theory the analysis of input ‘has no difference from the mental logic’ (Bonatti (1994b) p. 24) insofar as the role of pragmatics and meaning is concerned.

However, this raises one question. For Braine and O’Brien (1991) the MP rule and the schema for conditional proof constitute the lexical entry for if, i.e. the meaning of the if…then connector, since they believe the lexical entry to be the ‘information about the meaning of the words carried in semantic memory’ (Braine and O’Brien (1991) p. 183). However, according to Bonatti, these rules are thought to be applied to an input provided by a complete process of statement comprehension. This comprehension process is doubtless based on the pragmatic comprehension processes and the knowledge possessed by the subject which constitutes what Braine and O’Brien refer to as the construal in context of the word ‘if’. The question is thus: since the rules apply only to the result of a complete comprehension process, how can the meaning of a conditional statement be calculated before the lexical entry for the connector has been accessed, since it is the inference rules which constitute the
meaning of if? If Braine and O’Brien’s model is to account for context and content effects in terms of statement comprehension processes which precede the application of inference rules, then these rules are superfluous.

If we accept that the only role of this rule system is to produce inferences (i.e. the actual reasoning process, according to Bonatti), then we must necessarily consider that the problem of statement comprehension (rather than simply the pragmatic and contextual effects bearing on the meaning of the message) is independent of the problems raised by the psychology of reasoning. But how can we imagine a psychology of reasoning which has nothing to say about the way in which human beings attribute meaning to a statement of the form if p then q? In demonstrating that the standard mental models theory is no better placed than mental logic to account for content effects, Bonatti seems also to have shown that neither of the two theories is able to account for the interpretation of conditional statements. Both are based on formal systems which are insensitive to contents which make it necessary to calculate the meaning of the statement before these systems can be applied. In contrast, our model holds that the construction of mental models depends on the subject’s knowledge, the semantic nature of the variables involved and the context in which the statement is uttered.

It might be possible to make the standard mental models theory compatible with our results by arguing that each term to which a negation is applied is replaced by an affirmative value wherever possible. However, such an adaptation is not sufficient to account for the fact that the interpretation of the BB rules is more often biconditional than the NN rules in children and adolescents. In the standard theory the conditional interpretation derives from the construction of three models, two of which possess an antecedent of the same value (i.e. \(\neg p \land \neg q\) and \(\neg p \land q\)). Why should such a construction become more difficult when the value \(\neg p\) can be made affirmative (i.e. \(p\)), thus requiring the subjects to construct two models with the same antecedent (i.e. \(p \land \neg q\) and \(p \land q\))? The standard mental models theory does not seem to be able to account for the major phenomenon observed in our three experiments, at least in children’s reasoning. Nevertheless, the modifications that we have made to Johnson-Laird’s theory are minimal. They make it possible (a) to answer certain criticisms which have been levelled at this theory, (b) to explain the existence of reasoning biases, (c) to account for inference suppression phenomena and (d) to understand why the interpretation of if is so frequently biconditional.

5.2. Variability of the models for negated conditionals

One of the first criticisms addressed by Bonatti (1994a) to Johnson-Laird’s theory relates to the variable nature of the representation of conditional statements: ‘The representation of the conditional is extremely flexible’ (p. 726). In effect, according to the theory, the representation is different for ‘if A then B’, ‘if A then not B’ or ‘if not A then B’. According to Bonatti, the mental models would be so flexible that they could be made consistent with practically anything at all. However, the fact that the constructed mental models vary when negations are present in the statement is problematic only if we hold on to the idea of invariable lexical entry which defines
the meaning of if. Our model is totally consistent with, and reinforces, the explanation proposed by Johnson-Laird to justify the effect of negations: ‘a negation is likely to call to mind the affirmative alternative’ (Johnson-Laird and Byrne (1991) p. 67). In fact, a negation always provides an explicit and easily-representable alternative. It follows that the models which are constructed on the basis of statements containing negations might well take the following form when \( p \) and \( q \) are not binary:

\[
\begin{array}{cccc}
AA & NA^* & AN & NN \\
if \ p \ then \ q & if \ not \ p \ then \ q & if \ p \ then \ not \ q & if \ not \ p \ then \ not \ q \\
p & q & Ind & q & p & Ind & Ind & Ind \\
Ind & Ind & p & Ind & Ind & q & p & q \\
\end{array}
\]

* \( A \) for affirmative and \( N \) for negative: for example, \( NA \) refers to a rule with a negative antecedent (not \( p \)) and an affirmative consequent (\( q \)).

The hypothesis of the existence of these various representations is in no way ad hoc, but derives from the application of the proposed model. It makes it possible to account for two reasoning biases which are frequently described in the literature: the matching bias and the negative conclusion bias.

5.3. The matching bias

Our model explains the matching bias effect pointed out by Evans (1989) in much the same way as Johnson-Laird’s theory. In truth-value tasks, subjects have to judge whether a given case confirms, invalidates or says nothing about a conditional rule of the form \( if \ p \ then \ q \). Normally the four cases which can be derived from the affirmation (T for true) or the negation (F for false) of the antecedent or consequent (TT, FT, TF, and FF) are presented to the subject. In logic, only the TF case (i.e. true antecedent and false consequent) violates the rule, whereas the others (i.e. TT, FF and FT) confirm it. In fact, the FT case is frequently considered to be irrelevant, or to violate the rule (see Experiments 1 and 2). Introducing negations of the antecedent or consequent of the rule to be evaluated means that a given case (e.g. TT) can consist of two affirmations (e.g. for an AA rule such as if the letter is R the number is 4, the TT case is R4), two negations (for an NN rule such as if the letter is not R, the number is not 4, the TT case might be F6) or an affirmation and negation (for the rules AN and NA). The matching bias consists of a tendency to consider those cases whose content matches that of the conditional rule for evaluation as relevant. For example, for a rule of the form AA (e.g. if the letter is R, the number is 4), the TT case (i.e. R4) will be judged to be more relevant than the FF case (e.g. G6), none of whose terms corresponds to the content of the rule for evaluation. Similarly, for a rule of the type NA (e.g. if the letter is not R, the number is 4), the R4 case (which then corresponds to FT) will be judged to be more relevant than the others when the rule is evaluated.

Our model holds that if the statement contains a negation in the antecedent or consequent, the alternative model contains an explicit representation to which the negation applies (see above). Subjects tend to consider cases to be all the more
relevant, the more terms they contain which are explicitly represented in the constructed models. In contrast, the cases involving values which remain indeterminate are judged to be irrelevant by subjects since they are not explicitly represented. Thus, the model predicts that a case will more frequently be considered to be consistent with the rule when the two terms which form it are the object of an explicit representation in the same model. Explicitly-represented terms always correspond to the lexical contents of the rule, thus resulting in the matching bias. Thus, psychologically, the case TT confirms the AA rule more strongly than it confirms the NN rule because it is explicitly represented in the first model (first model: $p - q$) but remains implicit in the second (first model: $\text{Ind} - \text{Ind}$), whereas the opposite is true for the FF case.

Similarly a case will more often be considered to violate a rule when it consists of two terms which are explicitly represented, but in two different models, since it is extremely unlikely that they will co-occur. The cases TF and FT will more frequently be judged to be in violation of the rule in the AN and NA conditions respectively. Cases which contain values which refer to representations which remain indeterminate in the constructed models will always be judged to be irrelevant (e.g. FF in AA condition, FT in AN, TF in NA or TT in NN). All these predictions concur with the data reported by Evans et al. (1996).

The model also makes it possible to explain why the matching bias tends to disappear when explicit negations are used to construct cases for evaluation (e.g. ‘the letter is F’ is an implicit negation of ‘the letter is R’, whereas ‘the letter is not R’ is an explicit negation). The use of explicit negations induces an artificial dichotomisation of the variables and immediately causes the subjects to instantiate the indeterminate models (see Experiment 3). In this case, statements are less frequently judged to be irrelevant to the rule and are more often thought to confirm it when the values match one of the two constructed models and are more often thought to violate it when they match two different models. Once again, this is what was observed by Evans et al. (1996). Finally, as predicted by our hypothesis, the matching bias effect should disappear when the two variables are binary. This is what was indeed observed by Oaksford and Stenning (1992).

The matching bias observed in Wason’s (Wason, 1968) selection task can be explained using the same argumentation. Subjects tend to select instances which are explicitly represented in the representation which they construct for the rule for verification (however, see also the remarkable article by Sperber et al. (1995) for a detailed analysis of relevance effects in the selection task).

5.4. The negative conclusion bias

The negative conclusion bias is observed when subjects have to draw inferences from negated conditionals. When the four inference forms MP, MT, DA, and AC (see Experiment 3) are applied to the four types of if...then rule, obtained by the negation of the antecedent or the consequent, the conclusions obtained from these inferences are either affirmative or negative (see Table 6).

Numerous experiments (Evans, 1977; Wildman and Fletcher, 1977; Pollard and Evans, 1980) have shown that subjects more readily accept and produce negative
conclusions (i.e. not p or not q) than positive conclusions (p or q). More recently, Evans et al. (1995) have shown that this bias is limited to the denial inferences, DA (i.e. if p then q and not p, conclusion not q) and MT (i.e. if p then q and not q, conclusion not p). The DA inference is drawn less frequently from rules with a negative consequent (i.e. if p then not q and if not p then not q) since the conclusion is then q, and MT is drawn less frequently from rules with a negative antecedent (i.e. if not p then q and if not p then not q) since the conclusion is then p. As suggested by Evans et al. (1995), the negative conclusion bias might therefore be described as a double negation effect (not not p equals p, and not not q equals q).

Evans has discussed the possibility of explaining this effect in terms of the mental models theory or the mental logic theory (Evans, 1993; Evans et al., 1995). While the latter tells us little about reasoning biases in general, Johnson-Laird’s theory might be able to explain the bias, but, in order to do so it would have to be modified in such a way that it would no longer be able to account for the matching bias (Evans et al. (1995) p. 668). We believe that our model is able to explain this bias.

Evans et al. (1995), Experiment 2, used an inference-production task involving the contents of a card on which a letter was written on the left-hand side and a number on the right-hand side. The problems presented to the subjects took the following form: if the letter is R, the number is 4, the letter is not R, conclusion? Four types of conditional rule were constructed through the negation of the antecedent or the consequent, with the minor premise being either affirmative (i.e. the letter is R, the number is 4) or negative (the letter is not R, the number is not 4).

In conformity with our model, the statement informs the subjects that there is a relation between the letters and the numbers. The comprehension of the major premise should therefore consist of the establishment of a correspondence between two non-binary semantic spaces (letters and numbers) which are retrieved from LTM. The values presented in the antecedent and consequent should be linked in a directional way with the representation of this relation varying as a function of the determinate (i.e. R, 4) or indeterminate (not R, not 4) nature of the values. The indeterminate values should be represented by a set of letters or numbers, as indicated in Fig. 1. It is clear at first glance that there are major differences between the four situations: for example, the major premise may make a large number of rela-

<table>
<thead>
<tr>
<th>Type of rule</th>
<th>Given</th>
<th>Concl</th>
<th>Given</th>
<th>Concl</th>
<th>Given</th>
<th>Concl</th>
<th>Given</th>
<th>Concl</th>
</tr>
</thead>
<tbody>
<tr>
<td>If p then q</td>
<td>p</td>
<td>q</td>
<td>not p</td>
<td>q</td>
<td>p</td>
<td>not q</td>
<td>not p</td>
<td>not q</td>
</tr>
<tr>
<td>If not p then q</td>
<td>not q</td>
<td>not p</td>
<td>not q</td>
<td>not q</td>
<td>p</td>
<td>q</td>
<td>not q</td>
<td>not q</td>
</tr>
<tr>
<td>If p then not q</td>
<td>q</td>
<td>p</td>
<td>not p</td>
<td>not q</td>
<td>p</td>
<td>q</td>
<td>not p</td>
<td>not q</td>
</tr>
<tr>
<td>If not p then not q</td>
<td>not q</td>
<td>p</td>
<td>not q</td>
<td>not q</td>
<td>p</td>
<td>q</td>
<td>not p</td>
<td>not q</td>
</tr>
</tbody>
</table>

Predicted negative conclusion biases are underlined.
tions (e.g. if not p then q, Fig. 1c) explicit or none at all (e.g. if not p then not q, Fig. 1d).

This approach is of interest because it demonstrates that the space occupied by
values which remain available for non-explicit matchings varies between statements which are similar in appearance. Thus, the rule if the letter is R, then the number is not 4 (Fig. 1b) links the letter R to one of the numbers in the set ‘not 4’, whereas the rule if the letter is not R, then the number is 4 (Fig. 1c) indicates that all the letters different to R are linked to 4. In this second case (Fig. 1c), DA consists of drawing the conclusion not 4 from R. If the letter is R, all the numbers apart from 4 are available, and some type of matching is easy to conceive. In the first case (Fig. 1b), DA consists of drawing the conclusion 4 from not R. Now, if the letter is not R, the value 4 is clearly available. However, all the numbers other than 4 are also available apart from the indeterminate number which is associated with R. In consequence, the conclusion the number is not 4 in c is more likely than the conclusion the number is 4 in b, thus resulting in the negative conclusion bias.

The opposite phenomenon applies to MT. In b, the minor premise is the number is 4 and in c the number is not 4. In the former case, all letters except R are available to be matched with 4. The conclusion the letter is not R (MT) is highly probable. It is enough for subjects to imagine that the link to 4 originates from one of the terms in the set not R. In the latter case, in order to produce MT the subject should imagine that all the available numbers have to be linked to a single letter (i.e. R). Now, the problem situation and the principle of completeness (one-to-one correspondence) suggests that a given letter is linked to one (and only one) number. Thus, it would appear far from certain that the letter has to be R for any number different from 4, even if this reasoning is false. Thus, MT is more probable on the basis of if the letter is R, then the number is not 4 (negative conclusion not p) than on the basis of if the letter is not R, then the number is 4 (affirmative conclusion p).

The rule if the letter is R, then the number is 4 (Fig. 1a) specifies the relation R4. The association of letters different from R with any value other than 4 should be simple, as should be the inverse operation, thus resulting in production levels of DA and MT which are close to those observed in situations c and b respectively. It should be noted that in both cases the conclusion is negative (not 4 and not R).

The results obtained by Evans et al. (1995) as well as those documented by Evans et al. (1993) are perfectly consistent with our model. However, this model is less relevant in the latter situation, if the letter is not R, then the number is not 4, since it predicts that the subjects will represent the alternative case R4 in an explicit way. A totally explicit alternative model should result in the frequent production of the denial inferences DA and MT. However, the results show that these inferences are rarely produced (Evans et al., 1995; see Evans et al. (1993) for a review). It is possible that in this situation the difficulty subjects experience in constructing the first model (which links two indeterminate values) and the double negation effect proposed by Evans explain why it is difficult to produce inferences.

Thus, even though our model is not fully satisfactory, it accounts for the majority of the facts relating to the negative conclusion bias without losing its ability to explain the matching bias. According to our model the negative response bias would be due to the fact that it is easier to imagine that an isolated element is matched with any of the elements in a set (which necessarily leads to a negative
conclusion) than to imagine that each element in a set is matched with a single isolated element (which results in an affirmative conclusion). This phenomenon can be explained by what we have termed the principle of completeness which also explains the biconditional interpretations of BB situations.

This hypothesis also makes it possible to explain why the bias does not exist (or only exists in a weaker form) for AC (Evans et al., 1995). We believe that it is more difficult to imagine that each element of a set is associated with a single element (e.g. Fig. 1b for AC, affirmative conclusion R) than that an isolated element is matched with any element of a set (e.g. Fig. 1c for AC, conclusion not R). So far this effect is compatible with the negative conclusion bias. However, for the same reasons it should be easy to match two isolated elements as in the case if \( p \) then \( q \) (Fig. 1a, affirmative conclusion R for AC!). Thus, AC should be accepted and produced more frequently for the rules if \( p \) then \( q \) and if not \( p \) then \( q \) than for the rule if \( p \) then not \( q \). This can be observed, in varying degrees, in all the experiments conducted by Evans et al. (1995), i.e., respectively, 88 and 95 versus 81% in Experiment 1, 88 and 88 versus 84% in Experiment 2, 70 and 75 versus 55% in Experiment 3, as well as in Evans (1977), 75 and 81 versus 31%, and in Pollard and Evans (1980), Experiment 1, 66 and 72 versus 37%. This might explain why the negative conclusion bias is not observed in the AC condition. Once again the predictions concerning if not \( p \) then not \( q \) which can be derived from the model are unclear. It is conceivable that the impossibility of representing a precise relation (Fig. 1d) would result in a low level of AC production. However, this effect seems to be unsystematic.

5.5. The mental models theory should predict invalid inferences

Bonatti (1994a) claims that the psychological algorithm in Johnson-Laird’s theory permits invalid inferences. The premises:

If there is an A, then there is a B
There is a C

let subjects conclude that there is an A, a B and a C since, because they are reasoning on the basis of initial models (not fleshed out), the coordination of the premise C with the first model of the conditional (i.e. A – B) leads to the model A – B – C, whereas its coordination with the second model of the conditional, which has remained implicit, has not produced any model (Johnson-Laird and Byrne, 1991). Reading the only remaining model (i.e. A – B – C) results in the erroneous conclusion that there are simultaneously an A, a B, and a C. At this point we can add a comment to the response made by Johnson-Laird et al. (1994).

There are problems involved in considering, as Johnson-Laird does, that the coordination of the premise C with the first model of the conditional (i.e. A – B) leads to the representation of the conjunction of the three terms. In effect, the two items of information do not have the same status. In themselves, statements of the form if \( p \) then \( q \) do not supply precise information about the current state of affairs (see below for a number of exceptions to this rule). Models drawn from the if do not
therefore have the same status as those made possible, for example, by and or a
simple assertion of the type there is a C. This distinction should characterise the
models themselves and a mental models theory must make it possible to integrate
the status of particular models. Perner (1993) has suggested that representations
might index a specific model and indicate its nature (e.g. given vs. hypothetical).
The same suggestion was made by Johnson-Laird: ‘mental models [...] may be
intended to represent a true situation, a possible situation, or an imaginary situation.
Individuals must keep in mind the status of a model, but they can do so easily...’

It might be possible to imagine that, among a set of hypothetical models (such as
those generated by an if) constructed on the basis of the alternative values provided
by the semantic spaces of the antecedent and consequent, only models which are
matched with a premise would be retained and the others would be eliminated. This
goes back to Wittgenstein’s idea that elementary propositions having the same
coordinates within a semantic field would be mutually exclusive. In effect, the
models contain a variety of values in what we have designated as variables, in
both the antecedent and the consequent. If a premise informs subjects of the value
assumed by one of these variables (e.g. this is blue), the other values (e.g. this is
red) are ipso facto excluded. In Bonatti’s example, since the premise C does not
correspond to any term in the hypothetical A – B model, nothing can be derived
from their coordination. This would make it possible to arm the theory against the
possibility of coordinating models drawn from premises which are in no way
linked.

5.6. Inference suppression and the biconditional interpretation of if

Following Evans (1993), we have suggested that the meaning of if is essen-
tially relational. The statement if p then q would inform the subjects about a
relation between two variables P and Q by associating one of the possible values
of the first with one of the possible values of the second. We also suggest that, at
least in the experimental situations used in psychology, the statement fixes the
discourse universe within which the specific relation p – q has to be understood
and determines the variables to be considered. This would make it possible to
understand the frequency of biconditional interpretations of if in statements such
as ‘if it rains then the grass is wet’ and the increase in conditional interpretations
due to subjects remembering other possible causes, such as watering (Byrnes and
Overton, 1988).

This method, which is known as ‘contramanding’ has been used to demonstrate
that adolescents and some adults do not understand if in a biconditional way but that
their responses are due a conversational interpretation (Rumain et al., 1983). The
presence of contramanding would have the effect of blocking this interpretation.
Even though this is plausible, two questions have to be asked: (a) why do subjects
who know very well that there are more things than rain that make grass wet
interpret the statement as biconditional, and (b) does the contramand condition
really result in a conditional interpretation?
As far as the first question is concerned, we suggest that, apart from the relation between rain and the wetness of the grass, the statement determines the nature of the variables to be taken into consideration. Here, only the presence of rain would be taken into account since the statement only refers to rain. In effect, when all other factors are discounted, the grass is dry because it has not rained. Why do subjects not integrate information which they possess (e.g. watering wets grass) into their mental models? Because the value ‘sprinkler operating’ is not a modality of the variable ‘presence of rain’ which is the only one presented in the statement (i.e. it has rained or it has not rained). The fact, that watering can wet grass is therefore largely irrelevant to subjects within the framework of the stated problem, since the alternative mental models can only be instantiated by modalities of the variables involved. Since, as far as the subject is concerned, the experimenter is respecting the principle of quantity and has not referred to a cause such as ‘watering’ or ‘flooding’ or ‘morning dew’ there is no need to take account of any of these other values. The constructed mental model would therefore be:

<table>
<thead>
<tr>
<th>Rain</th>
<th>Wet grass</th>
</tr>
</thead>
<tbody>
<tr>
<td>No rain</td>
<td>Dry grass</td>
</tr>
</tbody>
</table>

The effect of the contramand condition would be to indicate that multiple causes have to be taken into account (e.g. watering or flooding) and not simply the variable mentioned in the statement (presence of rain). In this case, the value rain is not understood as a modality of the variable ‘presence of rain’, which maximised relevance, but of the variable ‘possible causes of wet grass’. Can we then consider that this condition results in a conditional interpretation of the statement ‘if it rains then the grass is wet’? This is not certain, because, in this case, subjects are no longer processing the relation between rain and the state of the grass, but the relation which obtains between, on the one hand, the presence of rain or watering and, on the other, the state of the grass. We can see that this new relation can be processed by means of a simple one-to-one correspondence:

<table>
<thead>
<tr>
<th>Rain</th>
<th>Wet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watering</td>
<td>Wet</td>
</tr>
<tr>
<td>Flood</td>
<td>Wet</td>
</tr>
<tr>
<td>Nothing</td>
<td>Dry</td>
</tr>
</tbody>
</table>

Behaviour would change not because the interpretation changes but because the situation to be processed and the conditions of verification of the statement ‘if it rains the grass is wet’ are modified. In other words the conditional statement as it is interpreted in psychological experiments would define its own context and specify both the nature of the relations and the variables which would have to be taken into account when verifying it.
This has been clearly demonstrated by Byrne (1989) who has shown that an inference of type MP can be suppressed by the presentation of a supplementary premise. On the basis of the conditional:

\[
\text{If she meets her friend then she will go to a play}
\]

and the premise:

\[
\text{She meets her friend}
\]

the subjects accepted the MP conclusion in 96% of cases:

\[
\text{Therefore, she will go to a play.}
\]

This percentage fell to 38% when the major premise introduced a supplementary variable:

\[
\begin{align*}
\text{If she meets her friend} & \text{ then she will go to a play} \\
\text{If she has enough money} & \text{ then she will go to a play}
\end{align*}
\]

The subjects accept MP in the first case for the same reason that they interpret the rule concerning the rain as a biconditional. Since the initial statement refers to only a single condition (meeting a friend), the fulfilment of this condition implies the consequent. Other possible conditions (e.g. having money, not being ill, that the actors are not on strike etc.) are not taken into account since the statement fixes the nature of the variables to be taken into consideration at the same time as their relation. The introduction of a supplementary variable leads subjects to construct a mental model which integrates the two variables on the basis of the knowledge which they possess. This knowledge permits them to distinguish between additional antecedents (which all have to be satisfied) and alternative antecedents (only one of which has to be satisfied in order to produce the consequent, Byrne and Johnson-Laird (1992)). In the case of going to the play, the antecedents are additional and the satisfaction of only one of them does not necessarily entail the consequent.

5.7. Atypical conditionals

Braine and O’Brien (1991) describe special conditionals in which the consequent informs subjects of a state of affairs and is therefore always satisfied, as for example in:

\[
\text{If you want some water, there is a fountain in the hall}
\]

Braine and O’Brien explain that this type of conditional cannot give rise to the counterposition (if there is no fountain in the hall then you do not want any water). To achieve this they describe a complex system of principles which are considered to inhibit reasoning (principle 3b: a supposition can be the antecedent of a conditional conclusion only if it is consistent with prior assumptions, and principle 3c: an assumption reiterated into a conditional argument cannot contradict the supposition that is to be the antecedent of the conditional).

Our model accounts for the processing of this type of statement in a more eco-
nomical way. Listeners’ knowledge informs them that the presence of a fountain cannot depend on their thirst. Since the dependency relation is excluded, the statement cannot be a true conditional and there is no need to construct an alternative model. The proposition ‘there is a fountain...’ should therefore be processed as a certain rather than merely a possible item of information. It results in the construction of a single model which is not hypothetical. It can be compared with ‘if you want water I can get you some’ which, for obvious reasons, mobilises a different type of processing.

This model applies equally well to the special form of conditionals represented by the negations of the type ‘if this singer is a good tenor then my name is Pavarotti’, or ‘if England beat the All-Blacks, I’ll eat my hat’, generally emitted in response to some affirmation or insinuation (’what a good tenor!’). The implausibility of the consequent informs the listener that the model induced by the statement does not correspond to reality for the speaker and that this reality should be looked for in the alternative model which does not verify the antecedent (this singer is a bad tenor, England will not win). This type of conditional therefore produces an implicit negation of the antecedent. It should be noted that this type of statement mobilises an implication paradox: a conditional with a false antecedent is always true. What the speaker is communicating in this case is that anything can be derived from a false affirmation.

6. Conclusion

The developmental approach sheds light on the development of the representational systems which underpin the understanding of statements of the form \( \text{if } p \text{ then } q \) and the production of the inferences derived from them. The facts assembled from our three experiments are entirely compatible with a mental models theory which supposes that the comprehension of \( \text{if } \) is underpinned by the construction of representations which establish a relation between the semantic structures specific to the antecedent and the consequent. This construction would depend on cognitive capacity, which increases with age, thus permitting the explicit representation of an ever-growing number of relations, the nature of the semantic structures between which relations are established, and the context. Binary structure would facilitate the production of an alternative model in the youngest subjects but would tend to limit the fleshing-out process to the construction of two models only (i.e. \( p \rightarrow q \) and \( /p \rightarrow /q \)), thus inducing a biconditional interpretation of the statements.

Mental logic explains differences in the interpretation of \( \text{if } \) in terms of pragmatic principles and developmental effects in terms of the increasing age-related ability to suppress the logic of conversation when doing laboratory reasoning problems (Rumain et al., 1983; Braine and O’Brien, 1991). However, these pragmatic principles are mobilised independently of the existence of inference rules (Braine and O’Brien, 1991).

In contrast, our model holds pragmatic effects and content to be a consequence of the processes involved in the construction of mental models on which reasoning is
based. It therefore represents a possible way of integrating all the different phenomena which affect the interpretation of *if* and the reasoning which is derived from it. As stressed by Sperber et al., 'the study of 'content effects' is the study of sound cognitive processes that are by no means out of place in subjects' performance' (Sperber et al. (1995) p. 44). There are three main differences between the model proposed here and that advanced by Johnson-Laird. Firstly, the abandoning of the hypothesis of a negation sign which acts as an index for the representations in the alternative models to \( p \rightarrow q \) (at least in children) makes it possible to understand the model construction process as an informal process which is sensitive to the nature of contents for processing and the context of the statement. The low level of sensitivity of adults to the variation of content introduced in Experiment 3 might be due to an ability to reformulate, by means of negation, the indeterminate models when conclusions have to be produced. The use of explicit negations in the premises appears to be one of the factors which is able to trigger this reformulation (cf. for example the fact that in the NN condition, adults produce more DA and MT conclusions in the presence of explicit negations, 44% and 79%, than implicit negations, 27% and 69% respectively).

Secondly, the mental models would appear to result from the establishment of a matching between two variables, one of the values of which is constituted by the antecedent and the consequent. The number and nature of the values between which relations can be established would depend primarily on the semantic structure in long term memory of each of the two concepts, as well as on the statement context or any factual knowledge which might constrain the nature of the values to be considered. Finally, the models constructed in response to *if* would be processed as hypothetical models by subjects. The production of an inference on the basis of *if* would consist of selecting, from the set of constructed models, that (or those) which is (are) compatible with subjects’ information about reality.

A consequence of this model is that the lexical entry for *if* would not be neither a set of rules nor a pre-established set of representational forms capable of being instantiated by the available values. We suggest that the ‘meaning’ of *if* lies solely in the described procedure for establishing a correspondence between semantic structures. *If* tells the listener nothing apart from the way in which the semantic spaces occupied by the values \( p \) and \( q \) present in the statement \( \text{if } p \text{ then } q \) have to be coordinated. The result of establishing this correspondence would depend on the nature of the linked semantic structures. Of the distinction made by Braine and O’Brien (1991) between the lexical entry for a word (i.e. the meaning stored in semantic memory) and the construal-in-context of the same word (i.e. the way the word is understood on some occasion of usage), we retain only the second term and argue that the role of a theory of reasoning is to account for this construal-in-context. The modus ponens inference which appears to define the meaning of *if* would simply be the only inference which is always available when the minimum meaning of the *if* \( p \) \text{ then } \( q \) statement is constructed (construction of the first model \( p \rightarrow q \)). There would therefore be no necessity for the difference in nature between the various inferences (e.g. MP involved as a lexical entry rule whereas DA and AC result from pragmatic principles) postulated by mental logic.
Appendix A. Rules used in Experiment 1

Appendix A.1. BB rules

If the pupil is a boy then he wears glasses.
If the car has an aerial then it pulls a trailer.

Appendix A.2. BN rules

If the part is pierced then it is a square.
If the house has a chimney then it has two windows.

Appendix A.3. NB rules

If it’s a rose then the paper is transparent.
If the shirt is red then it has short sleeves.

Appendix A.4. NN rules

If there is a rabbit in the cage then there is a chicken.
If there is a B on the card then there is a 3.

Appendix A.5. Example scenario (rule 1)

In an imaginary country, Alan, a school pupil, watches his friends in the playground and states that he has noticed the following rule concerning pupils and glasses.

If the pupil is a boy then he wears glasses.

Are there any cases which don’t correspond to the rule proposed by Alan?

Appendix B. Rules used in Experiment 2

Appendix B.1. BB rules

1 If he has a little nose then he has big ears.
2 If the plumage is multicoloured then it’s a male.

Appendix B.2. BN rules

3 If the sweater has short sleeves then it is yellow.
4 If the player is an amateur then he trains on Thursdays.
Appendix B.3. NB rules

5 If it is a rose then the paper is transparent.
6 If the manufactured part is a square then it is pierced.

Appendix B.4. NN rules

7 If you put on a white shirt then you put on green trousers.
8 If you plant daffodils then you plant lilies.

Appendix B.5. Example scenario (rule 8)

A boy is working with his father in the garden and asks him what flowers he can plant in the flower beds. His father says to him, ‘you are going to plant two types of flower in the flower beds, one coloured and one white-coloured, while obeying the following instructions: in each flower bed,

if you plant daffodils then you plant lilies.

Indicate all the flower beds that the child can plant while following the rule given by his father.

Appendix C. Rules used in Experiment 3 (inference production)

Appendix C.1. NN rules

If they are roses then we use white paper.
If the film is a western then it is shown on Thursday.
If there is a guinea pig then there is a canary.
If you buy a strawberry tart then you buy some Orangina.
If the pot is red then we plant geraniums.
If he is a violinist then he puts his things in wardrobe H.
If the customer buys a pullover then the present is a tie.
If you put on a white shirt then you put on green trousers.
If he is a builder then he has a Ford truck.
If the customer buys caramels then there is a 10% discount.
If it’s a 3rd year class then she goes to the swimming pool on Thursdays.

Appendix C.2. BB rules with semantic opposites in p and q

If the bird is female then it has light plumage.
If the farmer owns his farm then he has more than 12 cows.
If the house is new then it has a flat roof.
If the letter is a vowel then the number is even.
If the player is a professional then he is rich.
If the red light is lit then the door is open.
If the tree is big then the leaves are slender.
If the lecturer is a bachelor then he pays for his meal.

Appendix C.3. BB rules without semantic opposites
If the machine has aerials then it has a trailer.
If the pupil has a grant then he wears glasses.
If the chair has a back then it has arm rests.
If the building has a lift then it has a roof terrace.

References


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.


