# When 'Or' Means 'And': A Study in Mental models

**Philip N. Johnson-Laird**
Department of Psychology
Princeton University
Green Hall
Princeton, NJ 08544
phil@clarity.princeton.edu

**Patricia E. Barres**
Department of Psychology
Princeton University
Green Hall
Princeton, NJ 08544

**Abstract**

We describe an algorithm that constructs mental models of assertions containing sentential connectives, such as <u>and</u>, <u>if</u>, and <u>or</u>. It performs at three levels of expertise depending on the completeness of the models it constructs. At a rudimentary level of performance, it constructs models that make explicit as little as possible. One unexpected consequence is that it produces the same explicit models for assertions of the form:

if p then q, <u>and</u> if r then s
if p then q, <u>or</u> if r then s
p and q, <u>or</u> r and s.

We initially suspected that there was a bug in the algorithm (or theory), but there was not. We therefore carried out two experiments with logically-untrained subjects. Their results confirmed the phenomena: for many individuals, a conjunction of conditionals is equivalent to their disjunction, which in turn is equivalent to a disjunction of conjunctions.

## Introduction

The theory of mental models postulates that reasoning -- deductive or inductive -- is a process in which individuals use their perceptual ability and their knowledge to construct mental models of the relevant situations (see Johnson-Laird, 1983; Johnson-Laird and Byrne, 1991; Legrenzi, Girotto, and Johnson-Laird, 1993; Newell, 1990). In reasoning from verbal premises, individuals first construct a mental representation of the truth conditions of assertions, and then use this representation together with general knowledge and a knowledge of context to construct models, which have a structure corresponding to situations. Thus, a conjunction of the form:

a and b

calls for a single model representing both a and b:

a b

An exclusive disjunction:

a or else b

calls for two models (one for each possibility):

a
   b

An inclusive disjunction calls for three models (one for a, one for b, and one for both a and b). In general, if a conclusion holds in all the models of the premises, reasoners judge it to be necessary; if it holds in most of the models, they judge it to be probable; if it holds in at least one model, they judge it to be possible; and if it holds in none of the models, they judge it to be impossible.

In contrast to the model theory, most cognitive scientists who have studied reasoning postulate that the mind has a set of formal rules of inference akin to those of a logical calculus, which are used to reason in a process akin to the derivation of a proof (see e.g. Osherson, 1974-6; Macnamara, 1986; Braine and O'Brien, 1991; Rips, 1994). In this case, the mental representation of an inclusive disjunction depends on an expression in a mental language that has the following form:

a v b

where 'v' denotes inclusive disjunction. This representation can then be used to reason by matching it to a premise in a formal rule of inference (such as P v Q, not P, therefore, Q).

Our aim in the present paper is, not to address the differences between the two sorts of theory head on, but rather to describe the latest computer implementation of the model theory (in Section 1). This implementation led us to an unexpected consequence: the theory predicts that certain different compound assertions will be interpreted in the same way by logically-untrained individuals (Section 2). We report two experiments that corroborate these predictions (Section 3).

## An algorithm for mental models

The model theory postulates several levels of expertise in sentential reasoning. In what follows, we describe three such levels implemented in a computer algorithm, concentrating on its interpretation of compound premises.

At a rudimentary level (comparable to a logically-untutored individual's performance), the algorithm interprets an inclusive disjunction, such as:

There is a 'P' on the board or there is a 'Q' on the board, or both.

by constructing the following three models (shown on separate lines):

p
   q
p q

The first model is a <u>partial</u> model because it does not make explicit that q does not occur, and the second model is also a partial model because it does not make explicit that p does not occur. Performance at this rudimentary level likewise represents a conditional, such as:

If there is a 'P' on the board then there is a 'Q' on the board

by the following two models:

p q
. . .

Individuals grasp that both letters may be on the board, but defer a detailed representation of the case where there is not a 'P' on the board. The ellipsis accordingly signifies an <u>implicit</u> model, i.e. one that has no explicit content. It here represents an alternative possibility to the first model. Reasoners need to note that this implicit model represents cases where the antecedent does not hold. They can do so by noting that the antecedent is exhaustively represented in the explicit model, i.e. it cannot occur in the implicit model

[p] q
. . .

This notation is equivalent to storing a mental 'footnote' on the implicit model to indicate that the antecedent does not hold in it:

p q
. . . {¬p}

where '¬' signifies negation. The explicit model captures what happens when the antecedent occurs, and the footnote (in parentheses) represents that the antecedent, p, does not occur in the implicit model The advantage of such footnotes is that they are easy to model computationally, and so we have adopted them in our algorithm.

The computations required to interpret compound premises are simple. Thus, given the conjunction:

if p then q, <u>and</u> p

the algorithm uses a procedure for 'and' to combine each of the models for the conditional, if p then q, with the model for p. According to this procedure, two explicit models combine to yield an explicit model that avoids unnecessary duplications:

p q     <u>and</u> p   yield     p q

If one explicit model is inconsistent with another, or with the content of a footnote on an implicit model, then no new model is formed from them, i.e. the output is the null model:

. . . {¬p} <u>and</u> p   yield     nil

where 'nil' represents the null model, which is akin to the empty set. The procedure for conjoining two sets of models thus forms their Cartesian product, i.e. it multiplies each model in one set by each model in the other set according to the principles summarized in Table 1.

As an illustration, consider the conjunction of two conditionals: if a then 2 <u>and</u> if b then 3. The two conditionals have the following models:

a 2
. . . {¬a}

and:

b 3
. . . {¬b}

The conjunction of the two sets of models yields:

| | | |
|---|---|---|
| a 2 | | (by rule 2) |
| | b 3 | (by rule 2) |
| a 2 | b 3 | (by rule 1) |
| . . . {¬a ¬b} | | (by rule 3) |

Although the combination of models is guided by the footnotes, their content does not surface in explicit models at this rudimentary level of performance.

In contrast, at the next level up in the algorithm's performance, which corresponds to the ability of highly competent individuals, the content of footnotes <u>is</u> made explicit in any models resulting from the conjunction of implicit and explicit models. One consequence is that at this level the algorithm is able to make a modus tollens deduction. That is, given the premises:

if p then q
not q

The conjunction of the two sets of models proceeds as follows:

p q     <u>and</u> ¬q yields   nil
. . . {¬p} <u>and</u> ¬q yields   ¬p

Table 1: The principles for combining models

1. If the two models are explicit, then they are joined together eliminating any redundancies, unless one model contradicts the other, i.e. one represents a proposition and the other represents its negation, in which case the output is the null model. (When people combine two separate premises, they tend to drop propositions that they know categorically, e.g. if they know that p is the case, then the model, p q, combines with p to yield q alone.)
2. If one model is explicit and the other model is implicit, then the result is the explicit model unless its content contradicts the model in the footnote, in which case the output is the null model.
3. If both models are implicit, then the result is an implicit model, which conjoins the footnotes on the two implicit models unless the two footnotes contradict one another, in which case the output is the null model.

where there is no need to build in the content of the categorical premise, because it can be taken for granted. The resulting model yields the conclusion:

not p

Performance at this level also makes explicit the negated elements in disjunctive models, and indeed it is probably as accurate as possible given the use of implicit models.

Finally, at its most powerful level of performance, which exceeds the ability of untrained individuals, the algorithm fleshes out the contents of implicit models wholly explicitly. For example, fleshing out the implicit model of the conditional above:

p q
. . . {¬p}

calls for ¬p to occur in every new model, whereas separate models need to be made for q and ¬q because the footnote does not constrain them. The result is accordingly:

p q
¬p q
¬p ¬q

When the models of a set of premises are wholly explicit, they correspond to the rows that are <u>true</u> in a truth table of all the premises. Implicit models for conditionals and partial models for disjunctions are therefore devices that allow the inferential system to represent explicitly as little as possible, i.e. just that information which is of the first priority. Models can be made explicit but at the cost of fleshing them out. The notation can be used recursively -- as it is in the computer program implementing the algorithm -- to accommodate propositions of any degree of complexity.

## When 'or' means 'and'

In testing the performance of the computer program described in the previous section, we examined its interpretation of the inclusive disjunction of two conditionals:

if a then 2 <u>or</u> if b then 3.

At the rudimentary level of performance, it produced the following explicit models:

a 2<br>
    b 3<br>
a 2 b 3<br>
. . .

They are identical to the models of a <u>conjunction</u> of the two conditionals (illustrated above). Our first thought was that the program contained a bug, but we had instead stumbled upon an unforeseen consequence of the theory. The inclusive disjunction is of the form, X <u>or</u> Y, which calls for the following models:

X<br>
  Y<br>
X Y

where X models the conditional: if a then 2, and Y models the conditional: if b then 3. The models for X are:

a 2

and an implicit model. The models for Y are:

b 3

and an implicit model. The models for X <u>and</u> Y, as we saw in the previous section, are:

a 2<br>
    b 3<br>
a 2 b 3<br>
. . .

But, the first two of these models and the implicit model have already been constructed to represent X and Y respectively, and so it is necessary to add only the model:

a 2 b 3

The interpretation of a disjunction of conditionals:

a and 2 <u>or</u> b and 3

also calls for the same explicit models, but no implicit model.

In short, at its rudimentary level, the program constructs the same explicit models for a disjunction of conditionals, a conjunction of conditionals, and a disjunction of conjunctions. These interpretations arise from the interactions between two components of the theory: the use of implicit models in the representation of conditionals, and the use of partial models in the representation of disjunctions. Which assertions are erroneously interpreted according to these predictions? The answer is shown by the full sets of explicit models for the three sorts of assertion, which are shown in Table 2 (as constructed by the program at its highest level of performance). The disjunction of the conditionals is true in many more cases than its explicit models (shown in bold print in Table 2), the conjunction of the conditionals is true in some more cases than its explicit models, but the disjunction of conjunctions <u>is</u> accurately represented by its explicit models, though they do not enumerate the different ways in which each conjunction can be false.

## Empirical tests of the predictions

We have carried out two tests of the model theory's predictions about the interpretation of compound assertions. The first study was carried out as a class exercise in an undergraduate course in which each student was an experimenter and tested a separate logically-untrained individual. (None of the acting experimenters knew the predictions of the model theory.) The materials were the following four assertions:

Table 2: The complete set of models for three assertions. The items in bold are those that the theory predicts will be generated as true instances by logically-untrained individuals.

| If a then 2 or if c then 3 | If a then 2 and if c then 3 | a and 2 or c and 3 |
|---|---|---|
| **a 2** c ¬3 | | **a 2** c ¬3 |
| **a 2** ¬c 3 | **a 2** ¬c 3 | **a 2** ¬c 3 |
| **a 2** ¬c ¬3 | **a 2** ¬c ¬3 | **a 2** ¬c ¬3 |
| **a 2 c 3** | **a 2 c 3** | **a 2 c 3** |
| a ¬2 **c 3** | | a ¬2 **c 3** |
| ¬a 2 **c 3** | ¬a 2 **c 3** | ¬a 2 **c 3** |
| ¬a ¬2 **c 3** | ¬a ¬2 **c 3** | ¬a ¬2 **c 3** |
| ¬a 2 c ¬3 | | |
| ¬a 2 ¬c 3 | ¬a 2 ¬c 3 | |
| ¬a 2 ¬c ¬3 | ¬a 2 ¬c ¬3 | |
| ¬a ¬2 c ¬3 | | |
| ¬a ¬2 ¬c 3 | ¬a ¬2 ¬c 3 | |
| ¬a ¬2 ¬c ¬3 | ¬a ¬2 ¬c ¬3 | |
| a¬ 2 ¬c 3 | | |
| a ¬2 ¬c ¬3 | | |

1. If there is a 'A' on the board then there is a '2', and if there is a 'C' on the board then there is a '3'.
2. If there is a 'D' on the board then there is a '5', or if there is a 'E' then there is a '6', and both conditionals may be true.
3. There is a 'J' on the board and there is a '9', or there is a 'L' on the board and there is a '7', and both conjunctions may be true.
4. There is a 'G' on the board and there is a '8', and there is an 'H' on the board and there is a '4'.

For each assertion, the subjects wrote down a list of the possible circumstances in which the assertion would be true, i.e. pairings of letters and numbers, such as 'A 2'. When they had completed this task, they went through each assertion again, and wrote down a list of the possible circumstances in which the assertion would be false. The model theory predicts that the first three assertions will elicit the same list of true circumstances (corresponding to the explicit models presented in the previous section). The fourth assertion is a control that should be treated as true only in one case, namely, when all four items co-occur.

The results corroborated the predictions. Table 3 summarizes the most frequent selections for the four assertions. The numbers of subjects who listed the responses in the order predicted by the algorithm were as follows: 9 for the first assertion, 17 for the second assertion, and 16 for the third assertion. On the assumption that there are 16 possible selections, the likelihood of making by chance these three selections in their given order is at the micro-probability level, i.e. the chance probability for the selection is $1/16 \times 1/15 \times 1/14 < 0.0003$, and so the binomial probability for, say, 9 out of 25 subjects making this selection is miniscule.

Our second experiment was carried out in a conventional way by a single experimenter. The materials again compared a conjunction of conditionals with a disjunction

Table 3: The most frequent responses in Experiments 1 and 2: the number of subjects who listed the circumstances shown as cases in which the assertions would be true. We have stated the assertions as though they had the same lexical materials.

Experiment 1 (n = 25)

| | If A then 2 and if B then 3 | If A then 2 or if B then 3 | A and 2 or B and 3 | A and 2 and B and 3 |
|---|---|---|---|---|
| A 2 | 21 | 23 | 23 | 0 |
| B 3 | 20 | 24 | 22 | 0 |
| A 2 B 3 | 20 | 24 | 19 | 24 |

Experiment 2 (n =21)

| | If A then 2 and if B then 3 | If A then 2 or if B then 3 | A or 2 and B or 3 | A or 2 or B or 3 |
|---|---|---|---|---|
| A 2 | 13 | 19 | | |
| B 3 | 11 | 20 | | |
| A 2 B 3 | 16 | 11 | | |
| A B | | | 18 | |
| A 3 | | | 17 | |
| 2 B | | | 17 | |
| 2 3 | | | 18 | |
| A | | | | 16 |
| 2 | | | | 16 |
| B | | | | 16 |
| 3 | | | | 16 |

of conditionals, but used a conjunction of disjunctions and a disjunction of disjunctions as control assertions. The task was the same as before except that we surreptitiously timed the interval from the presentation of an assertion until the subjects began to list the circumstances. There were two trials in each condition (with a different lexical content). Table 3 also presents the data from this experiment. In both experiments, the control sentences were accurately interpreted, and the task of generating false instances was much harder -- it took subjects about 2 seconds longer to make their initial responses, and they varied considerably in their responses. Few subjects accounted for all 16 contingencies in their combined lists of true and false instances.

## Conclusions

The model theory assumes that the bottle-neck in inferential processing is the capacity of working memory (Simon, 1982). Reasoners accordingly try to work with as few explicit models as possible: they do not represent the negative cases in their models of an inclusive disjunction of the form, p or q, and they do not represent explicit models of the case where the antecedent of a conditional, such as if p then q, is false. These two assumptions have an unexpected consequence, which we discovered from the implementation of the algorithm: assertions that seem superficially very different are likely to be interpreted in similar ways. Our results corroborated this prediction: a conjunction of conditionals is taken to be true in the same circumstances as their disjunction, which is taken to be true in the same circumstances as a disjunction of conjunctions. These results bear out the heuristic value of computer implementations of theories, and suggest that the theory of mental models gives a good account of the interpretation of sentential connectives. Alternative theories of reasoning, including those based on formal rules of inference, either do not deal with the meaning of connectives or else treat their meanings as equivalent to the rules of inference that govern them. Because the rules for disjunction differ from those for conjunction, which in turn differ from those for conditionals, such theories do not explain the cases in which 'or' means 'and', and 'if' means 'and'.

## Acknowledgements

This research was supported in part by the James S. McDonnell Foundation. We thank the students in Cog 200, 1993-4, for carrying out Experiment 1.

## References

Braine, M.D.S., & O'Brien, D.P. (1991). A theory of If: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, 98, 182-203.

Johnson-Laird, P.N. (1983). *Mental Models*. Cambridge: Cambridge University Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Legrenzi, P., Girotto, V., & Johnson-Laird, P.N. (1993). Focussing in reasoning and decision making. *Cognition*, 49, 37-66.

Macnamara, J. (1986). *A Border Dispute: The Place of Logic in Psychology*. Cambridge, MA: MIT Press.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Osherson, D.N. (1974-6). *Logical Abilities in Children, Vols. 1-4*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rips, L.J. (1994). *The Psychology of Proof*. Cambridge, MA: MIT Press.

Simon, H.A. (1982). *Models of Bounded Rationality, Vols. 1 and 2*. Cambridge, MA: MIT Press.