Memory & Cognition
2001, 29 (1), 165-175

Propositional reasoning: The differential contribution of "rules" to the difficulty of complex reasoning problems

FRANK RIJMEN and PAUL DE BOECK
University of Leuven, Leuven, Belgium

In Experiment 1, complex propositional reasoning problems were constructed as a combination of several types of logical inferences: modus ponens, modus tollens, disjunctive modus ponens, disjunctive syllogism, and conjunction. Rule theories of propositional reasoning can account for how one combines these inferences, but the difficulty of the problems can be accounted for only if a differential psychological cost is allowed for different basic rules. Experiment 2 ruled out some alternative explanations for these differences that did not refer to the intrinsic difficulty of the basic rules. It was also found that part of the results could be accounted for by the notion of representational cost, as it is used in the mental model theory of propositional reasoning. However, the number of models as a measure of representational cost seems to be too coarsely defined to capture all of the observed effects.

Deductive reasoning is the ability to draw logically valid conclusions from given information (the premises of a reasoning problem). One form of deductive reasoning is propositional reasoning: deductive reasoning with connectives, such as *if/then*, *and*, *or*, or *not*. For example, when the following propositions are given:

John is in Brussels.

If John is in Brussels, then Mike is in Dublin.

one can easily deduce

Mike is in Dublin.

According to *rule theories* of deductive reasoning (e.g., Braine, 1978; Braine & O' Brien, 1991; Braine, Reiser, & Rumain, 1984; Macnamara, 1986; Osherson, 1974, 1975; Rips, 1983, 1994; Rips & Conrad, 1983), human beings solve a logical problem in three stages. First, in the *encoding* stage, the logical structure of the problem is uncovered and represented in working memory. Second, in the *reasoning* stage, one or more basic reasoning rules are applied to this language-like structure. Finally, in the *decoding* stage, the result of the reasoning process is translated back into the form demanded by the task—for example, natural language. Rule theories concentrate on the second stage and on the basic rules used in that stage. The set of basic rules is assumed to be a repertory of syntactic, logically valid inferences. Throughout the text, the term *basic rule* refers to a psychologically elementary reasoning step, such as those proposed by rule theories. The term (*logical*) *inference*, on the other hand, is used as a more general term in two respects: There are logical inferences for which there is no basic rule (e.g., modus tollens), and the term *logical inference* is not confined to a particular psychological theory.

Each of the basic rules specifies how to reason with a particular sentence connective, such as *if*, *or*, *and*, or *not*. For example, the basic rule for modus ponens states that:

> Given the propositions of the form "IF p, THEN q" and "p"; conclude "q."

In a common notation:

```
if p, then q
p
----------------
q
```

According to rule theories, solving complex reasoning problems consists of solving consecutive simple reasoning problems for which basic rules are available. Rule theories assume that people will consider a conclusion to be true if they succeed in forming a chain of inferences linking the (true) premises with the conclusion to be evaluated.

In rule theories, problem difficulty is accounted for in terms of the reasoning operations: The length of the mental proof and the difficulty of the basic rules are assumed to determine the difficulty of a problem. The more basic rules there are that must be applied and the more difficult they are, the more difficult the problem. Hence, according to rule theories, problem difficulty is a function of the *number* and the *difficulty of the constituent basic rules*. Both Rips (1983) and Braine et al. (1984) tested this prediction, although a somewhat different method was used in each study. Rips (1983) tested a model

Frank Rijmen was supported by the Fund for Scientific Research, Flanders (FWO). We are also grateful to Walter Schaeken and Lance Rips for their helpful comments on an earlier draft of the manuscript. Correspondence concerning this article should be addressed to F. Rijmen, Department of Psychology, Tiensestraat 102, B-3000 Leuven, Belgium (e-mail: frank.rijmen@psy.kuleuven.ac.be).

 Copyright 2001 Psychonomic Society, Inc.

for complex reasoning problems that was based on availability parameters for the basic rules. Given that a basic rule is applied correctly if available, the difficulty of the basic rule lies in its availability. The parameters were estimated from the proportions of correct answers on the complex reasoning problems. The correlation between predicted and observed proportions was .93, corroborating the rule theory of Rips (1983).

Braine et al. (1984) have tested their theory by regressing the *perceived* difficulty of complex problems on the assumed constituent basic rules, coded as dummy variables. The resulting weights could be interpreted as the difficulty of the basic rules, since the perceived difficulty was modeled as a sum of basic rule weights. They obtained a multiple correlation of .92. When only the number of basic rules was used as a predictor, the correlation was only .79, which is significantly lower. When *actual* difficulty (an error index) was used as a dependent variable, the corresponding correlations were .73 and .57, respectively, the difference not reaching significance.

These high correlations do corroborate the rule theories of propositional reasoning, but the use of different parameters for each of the basic rules is a strong indication that the rules differ as to their difficulty. If there were differences, providing a basis for the differences would not only enhance the comprehensibility of rule theories, but would also render them more parsimonious by reducing the number of free parameters (from one for each basic rule to, ideally, one underlying parameter). In this respect, the early work of Osherson (1974), the first proponent of the rule theories, offers some interesting ideas. Osherson (1974) proposed different *performance models* (regression equations) to account for success rate in a broad range of reasoning problems. All that these models have in common is that the weight of the basic rules in the regression equation is conceptualized as the amount of *mental processing space* or *computing area* required by the application of the basic rule. On the one hand, because Osherson (1974) provides no account of how to determine the amount of computing space for the basic rules, we are left with no more than an indication of the problem. On the other hand, treatment of the differential contribution of basic rules to the difficulty of a complex reasoning problem as a differential amount of required computing area is related to the notion of representational cost. This is the way in which difficulty is conceived in the mental model theory, to which we now turn.

The *mental model theory* (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; Johnson-Laird, Byrne, & Schaeken, 1992) is another main approach to propositional reasoning. In mental model theory also, reasoning consists of three stages. In a first stage, called *comprehension,* a mental model is constructed for the problem, which is a mental representation of a situation consistent with the premises. This first stage can be considered as an encoding stage. Second, in the *description* stage, a putative conclusion is generated on the basis of this model. This putative conclusion contains information not yet explicitly asserted by the premises. If no putative conclusion can be generated, the answer is that nothing follows from the premises. The third stage, the only really deductive stage (Johnson-Laird & Byrne, 1991, p. 36), is one of *validation,* since participants are assumed to search for alternative models of the premises that falsify the putative conclusion. If a falsifying model is found, it is necessary to return to the second stage, to search for another putative conclusion that holds in the models constructed thus far. If no falsifying model is found, the conclusion is considered true.

In the mental model theory, more attention is paid than in rule theories to the first stage, the encoding or comprehension stage, in which the problem is represented. Given the conditional "if p, then q," one can construct four possible situations (corresponding to the rows of a truth table), of which only the situation "p not q" is inconsistent with the conditional. However, given that one wants to minimize the representational cost because of the limited capacity of working memory, human beings are assumed to represent initially as much information as possible in an implicit way, instead of representing all three possible models explicitly right from the beginning. The assumed initial representation consists of one explicit model, in which it is indicated that "p" is exhausted with respect to "q" ("p" within square brackets)—that is, that "p" cannot occur with "not q"—and of an implicit model indicating that there might be cases in which "p" is not present (the three dots):

```
[p] q
...
```

If "p" is given as a second premise, the first model is selected, and the implicit model eliminated, with the result of

```
[p]    q
```

This leads to the conclusion "q," which cannot be falsified. If "not q" is given as a second premise, the explicit model is eliminated, and all three models that are consistent with the conditional have to be made explicit (or "fleshed out"):

```
p      q
not p  q
not p  not q
```

Only the third model is consistent with the premises, resulting in the conclusion "not p."

Problems differ as to the number of explicit models needed in the comprehension (encoding) stage to reach a subsequent conclusion, and this is what determines the representational cost of a problem. For example, the empirical observation that modus ponens is easier than modus tollens can be accounted for by the mental model theory because the first requires only one explicit model (the initial one) and the latter requires three (all three that are consistent with the conditional). Johnson-Laird et al. (1992) report several experiments that confirm the cen-

tral prediction of the mental model theory that a problem is harder the greater the number of explicit models that have to be constructed and represented in working memory.

In contrast with rule theories, the mental model theory does predict differences in difficulty between logical inferences that correspond to single basic rules according to rule theories, since, depending on the basic rule, a different number of explicit models is involved. As has already been illustrated, modus ponens requires the construction of only one explicit model. A disjunctive syllogism, however, requires the construction of two explicit models. The disjunction "p OR q" has the following initial representation (following Johnson-Laird et al., 1992):

```
p
        q
```

The categorical premise "NOT p" eliminates the first and completes the second:

```
¬p      q
```

The conclusion "q" cannot be falsified. Also, a disjunctive modus ponens requires the construction of two explicit models. Following Johnson-Laird et al. (1992, p. 428), the premise "IF p OR q, THEN r" is initially represented as

```
[p]             r
        [q]     r
...
```

The categorical premise "p" eliminates the second model and the implicit model, which yields "r." Again, this conclusion cannot be falsified. Thus, the mental model theory predicts that modus ponens is easier than both disjunctive modus ponens and disjunctive syllogism.

In an experiment (Johnson-Laird et al., 1992) in which people had to generate conclusions (as opposed to the conclusion-evaluation paradigm of the rule theories), a modus ponens inference was found to be simpler (made more often) than a disjunctive syllogism inference with an exclusive disjunction. In conclusion, mental model theory does offer an explanation for differences in difficulty between "basic rules" in terms of working memory load.

It is our aim to test further whether it is necessary to assume that basic rules differ in their contribution to the difficulty of complex reasoning problems. If so, rule theories will have to be complemented with a principle that can account for these differences and, hence, for the difficulty of complex reasoning problems.

Insofar as differences between basic rules are found, we will compare these differences with the predictions of the mental model theory. Independent of the outcome, we will also test the mental model theory on the same data. If the differences are as predicted, this will certainly strengthen the claim of mental model theorists that the difficulty of reasoning problems is primarily affected by representational cost (the number of explicit models required to solve the problem).

EXPERIMENT 1

Analogous to Braine et al. (1984) and Rips (1983), we constructed complex reasoning problems by combining the basic rules, and we used a conclusion-evaluation paradigm (as opposed to the conclusion-generation paradigm used by Johnson-Laird et al., 1992). In contrast with the studies just mentioned, we used only a limited number of basic rules. This allowed us to combine the basic rules factorially, while keeping the total number of problems to be solved reasonable. By combining the basic rules factorially, correlations and tradeoffs between parameters were avoided as much as possible. The following basic rules were focused on in this study: *conjunction* (p, q ∴ p AND q), *modus ponens* (IF p, THEN q; p ∴ q), *disjunctive modus ponens* (IF p OR q, THEN r; p ∴ r), *disjunctive syllogism* (p OR q; NOT p ∴ q), and *solving a contradiction* (p, *not* p ∴ incompatible).

We used problems with a reasonable error rate to avoid a ceiling effect, and we worked with a larger number of participants than in the previous studies, in order to obtain reliable success proportions per problem.

Apart from the basic rules, four other factors were manipulated: *content*, *order of presentation of the premises*, *truth–value of the conclusion to be evaluated*, and *directness of the reasoning problems*. The last factor requires some explanation.

Braine et al. (1984) made a distinction between problems that can be solved by applying one or more basic rules in a quasi-automatic way (direct reasoning problems), which are supposed to be universally available, and problems that require the application of some strategy, which is considered not to be universally available and, hence, not a part of the repertory of basic rules (indirect reasoning problems). One such strategy is *reductio ad absurdum*, which, according to rule theories, is used to solve a modus tollens (a logically valid inference for which there is no basic rule assumed to account for its greater difficulty):

given:	(1) IF p, THEN q	
	(2) NOT q	
reasoning:	(3) suppose p	*supposition*
	(4) q	*modus ponens* 1+3
	(5) incompatible	*contradiction*: 2 and 4 cannot both be true
	(6) not p	*reductio ad absurdum*, a supposition leading to a contradiction must be false.

In half of the problems, the modus ponens, which is a direct reasoning problem (there is even a basic rule for it), will be replaced with a modus tollens, which requires an indirect way of reasoning.

Method

Participants. Two hundred fourteen high school students between 16 and 19 years of age participated in the experiment.

Material. Each participant received a booklet, with 2 pages of instructions and 10 pages with problems. On each of the latter, 3 problems were presented, for a total of 30 problems. Under each problem, four response alternatives were presented:

—necessarily true

—necessarily not true

—undecidable

—I don't know

It was explained in the instructions that the participants should choose "necessarily true" if they believed that, given true premises, the conclusion must be true; "necessarily false" if they believed that the conclusion must be false; "undecidable" if they believed that, given the premises, the conclusion could be true, but also could be false (hence, that not enough information was given); and "I don't know" if they could not work it out. It was explained to the participants that "undecidable" is a genuine conclusion that could be correct for some of the problems, whereas "I don't know" means that one cannot choose among the other three response alternatives or has given up solving the problem. "I don't know" was offered as an alternative response in order to avoid guessing behavior. Four random sequences were generated for the problems, equally divided over the participants.

Design. Each participant had to evaluate the conclusion of 30 problems. For 12 problems, the correct conclusion was "necessarily true"; for 12 others, the correct conclusion was "necessarily not true"; for the other 6, the correct conclusion was "undecidable." The 24 problems with either a true or a false conclusion were constructed by combining premises as illustrated in Table 1, resulting in six problem types. A letter represents an elementary proposition, possibly containing a negation. It was never necessary to perform a double negation (e.g., for Problem Type 3 [see Table 1], "q" never represented a proposition with a negation).

Modus ponens versus modus tollens was the first factor in the design. For 12 reasoning problems, one inference was a modus ponens (left column; direct reasoning problems). For the other 12, this inference was a modus tollens (indirect reasoning problems).

The second factor of the design, which refers to the inferences that are combined with modus ponens or modus tollens, had three levels: *modus ponens + conjunction, disjunctive syllogism,* and *disjunctive modus ponens*. This factor is orthogonal with the first and makes a distinction between three sets of eight problems. For the first set of eight reasoning problems, problems of *level modus ponens + conjunction* (first row of Table 1), the second inference was a modus ponens (requiring the application of one basic rule). This modus ponens inference could be made independently from the modus ponens or modus tollens inference already discussed. The results of both conditional inferences subsequently had to be combined with a conjunction, which was then the third inference (again requiring one basic rule). For the second set of eight problems, problems of *level disjunctive syllogism* (second row of Table 1), the second (and last) inference to be made was a disjunctive syllogism (requiring one basic rule). The categorical proposition of the disjunctive syllogism was the result of the modus ponens or modus tollens inference from the first step. The third set of eight problems, problems of *level disjunctive modus ponens* (third row of Table 1), was similar to the second set, except that the second inference was now a disjunctive modus ponens (also requiring one basic rule).

A third factor, orthogonal to the first two, was the truth value of the conclusion to be evaluated: *true or false*. For 12 problems, the conclusion to be evaluated was true; for the other 12, the conclusion to be evaluated was false. For the latter, the conclusion to be evaluated contradicted the conclusion that followed from the premises. Solving this contradiction was an additional inference (requiring one basic rule) for these problems. To avoid the possibility that the presence of a negation would act as a cue for the correct answer, some propositions of the premises contained a negation, although no problem required an inference involving a double negation.

The fourth factor, manipulated orthogonally, was the *content of the problems*. Half of the problems were about people in cities or countries, for example "John is in Paris." The other half were about a functioning factory—for example "The green light flashes." The orthogonal manipulation of these factors resulted in $2 * 3 * 2 * 2 = 24$ cells in the design, each corresponding to exactly one problem.

A fifth factor, not orthogonally manipulated because this would render the experiment too long, was the *order of presentation of the premises*: grouped versus ungrouped. For each type of content, the premises of a given inference were presented in a group in half of the problems (as presented in Table 1) and separated by another premise in the remaining half of the problems (e.g., for Type 1: "p" became the first premise and "IF p, THEN q" the third, with a premise ["IF r, THEN s"] of the other modus ponens in between). For the content of a factory, these problems were of Types 2, 3, and 6, and for the content of people in cities or countries, these problems were of Types 1, 4, and 5.

Apart from the 24 problems just described, 6 reasoning problems with undecidable conclusions were added as filler items. These problems were constructed by replacing the modus ponens or modus tollens inference with an inference involving an affirmation of the con-

Table 1
The Six Problem Types of Experiment 1

	Modus Ponens	Modus Tollens
Modus ponens + conjunction	1. IF p, THEN q p IF r, THEN s r --- q AND s	2. IF p, THEN q p IF r, THEN s NOT s --- q AND NOT r
Disjunctive syllogism	3. p OR q IF r, THEN NOT q r --- p	4. p OR q IF q, THEN r NOT r --- p
Disjunctive modus ponens	5. IF p OR q, THEN r IF s, THEN q s --- r	6. IF p OR NOT q, THEN r IF q, THEN s NOT s --- r

sequent or a denial of the antecedent, respectively (neither is valid in classical logic). In this way, we ensured that the correct conclusion was sometimes something other than "necessarily true" or "necessarily not true."

Procedure. The experiment was conducted in groups of 20 participants, more or less. At the beginning of the experiment, the instructions were briefly explained. The participants were asked not to guess but to choose the response alternative "I don't know" in cases in which they had not made one of the three evaluations: "necessarily true," "necessarily not true," or "undecidable." It was stressed that they would not be allowed to write anything down while solving the problems. The participants had to complete the task within 50 min, and they all succeeded in doing so.

Results

We will concentrate now on the analysis of the difficulty of the 24 reasoning problems. The six filler problems were not included in the analysis. The range of the difficulty of the 24 problems, expressed in proportion of correct answers, goes from .44 to .98, with a mean of .72. The proportions of correct answers to the 24 problems are given in Table 2. These proportions were converted into their logit:

$$\text{logit}(p) = \ln [p/(1-p)],$$

where p = proportion of correct answers.

Working with the logit avoids the problem of a restricted range between 0 and 1 and compression in the neighborhood of these limits. The logit was the dependent variable in the subsequent regression analyses, following the regression equations of Braine et al. (1984). Since the length of the problem (expressed in number of words) did not correlate significantly with the objective difficulty of the problem, as expressed in the logit measure ($r = .30$, n.s.), it was not included in the predictor set. Given that the logit of the proportion correct answers is an inverse measure of difficulty, it is clearly not the case that problems with a longer wording are more difficult.

In a first regression equation, the predictor was the number of basic rules necessary to solve the problem. Hence, each basic rule was considered to be of equal difficulty. Appendix A describes how the value on this predictor was computed for the different problem types. The number of basic rules explained only 16% of the variance, which is not significant [$F(1,22) = 4.14$, n.s.].

According to Braine et al. (1984), a modus tollens inference requires four reasoning steps (see the introduction). It is unclear, however, whether it is also appropriate to count four basic rules, as we did. First, a problem involving a modus tollens inference is an indirect reasoning problem (Braine et al., 1984). It cannot be solved using the repertory of universally available basic rules (see the introduction). Second, Rips (1994, p. 116) conceptualizes three of the reasoning steps involved as one, although cognitively complex, rule. Third, when the conditional is interpreted as a biconditional, a simple solution exists for a modus tollens inference (Braine et al., 1984, p. 343). For all these reasons, it seems worth trying out an alternative that leaves the number of rules for a modus tollens inference unspecified.

In order to assess the predictive value of the number of rules independently from the number of basic rules assumed for a modus tollens inference, a second regression analysis was conducted with two predictors: a binary variable, which coded for modus ponens (value of one) versus modus tollens (value of zero; see below for the way the factors of the design were coded into binary predictors), and the number of basic rules minus the number of basic rules for modus ponens/modus tollens. The two predictors explained 28% of the variance [$F(2,21) = 4.01$, $p < .05$]. The reduced number of basic rules had no significant effect in the regression analysis ($t = 0.722$, n.s.). Hence, the number of basic rules does not seem to have much predictive value beyond the difference between modus ponens and modus tollens, at least for the set of problems under consideration here.

To investigate whether the relation between the number of basic rules and the logit of the proportion of correct answers (an inverse measure of difficulty) is monotonic (without having to make assumptions about the particular form, as in the linear regression analysis), we computed the Spearman rank correlation. The correlation was −.42, which although not very high, is significantly smaller than zero ($p < .05$). Nevertheless, this is not much better than the Pearson correlation between number of basic rules and difficulty (in logit; $r = -.40$), indicating that the monotonic relation between the number of basic rules and the logit of the proportion of correct answers does not depart strongly from linearity.

The next step in the analysis was to allow for a differential weighting of the basic rules, as in Braine et al. (1984). Therefore, the three factors of the design referring to basic rules were coded into four binary variables as follows:

P1: 1/0 if first inference modus ponens/modus tollens.
P2: 1/0 if second and third inference modus ponens + conjunction/other.
P3: 1/0 if second inference disjunctive syllogism/other.

Table 2
Proportion of Correct Answers for the 24 Problems of Experiment 1

	People in Cities or Countries		Functioning Factory	
Problem Type	True Conclusion	False Conclusion	True Conclusion	False Conclusion
1	.95	.87	.98	.92
2	.77	.82	.86	.68
3	.59	.53	.79	.51
4	.50	.49	.78	.47
5	.88	.83	.95	.84
6	.73	.70	.44	.51

P4: 1/0 if conclusion false/true.

Problems with a value of zero on both P2 and P3 were problems with a disjunctive modus ponens as the second inference.

By performing a multiple regression analysis with P1 to P4 as predictors, different weights are allowed for the basic rules.

The four predictors, P1–P4, explained 74% of the variance [$F(4,19) = 13.82, p < .001$]. It was possible to perform a direct comparison between the model incorporating P1–P4 as predictors and the one with only the number of basic rules as a predictor, since the latter is nested within the former. The improvement of the prediction was significant [$F(3,19) = 14.50, p < .001$]. The standardized regression weights for P1–P4, in descending order of absolute value, were .51 (P1), −.38 (P3), .35 (P2), and −.30 (P4; all $p < .05$). Positive coefficients indicate that problems with a coding of 1 are easier.

These results imply that the contributions of the basic rules to the difficulty of the total problem are not equal but depend on the kind of basic rule. Furthermore, a reasoning problem with a modus tollens is more difficult than a problem with a modus ponens, and a reasoning problem with a disjunctive syllogism or a disjunctive modus ponens is more difficult than a problem with a modus ponens + conjunction; a reasoning problem with a disjunctive syllogism is more difficult than a reasoning problem with a modus ponens + conjunction or a disjunctive modus ponens, and finally, a problem with a false conclusion is more difficult than a problem with a true conclusion.

In addition, to assess directly the difference between a problem with a disjunctive syllogism and a problem with a disjunctive modus ponens, P3 (disjunctive syllogism vs. not) was replaced with P3′, which was coded as follows:

P3′ = 0 if a modus ponens plus conjunction were the second and third inferences.

= 1 if the second inference was a disjunctive syllogism.

= −1 if the second inference was a disjunctive modus ponens.

The standardized regression weight of P3′ in the regression analysis with P1, P2, P3′, and P4 as predictors was −.33 ($p < .05$). Hence, a reasoning problem with a disjunctive syllogism is more difficult than a reasoning problem with a disjunctive modus ponens.

In order to test the effects of the other factors of the design (content and presentation order of the premises), these factors were also coded into binary variables, as follows:

P5: 1/0 if content of functioning factory/content of people in cities or countries.

P6: 1/0 if premises of the same inference not presented together/presented together.

When either P6 or P5 was added to the predictor set, these factors were not significant. Note that P6 is not orthogonal to the other Ps. We also tested each of the pairwise interactions for its additional contribution. Among these interactions, only one was significant ($p < .01$), P1 * P3, with a standardized regression weight of −.46. This indicates that the difference between modus ponens and modus tollens becomes smaller when a disjunctive syllogism is also part of the problem.

Discussion

The objective of Experiment 1 was to investigate whether different basic rules have different effects on the difficulty of a complex reasoning problem. The answer is yes. The number of basic rules as such explains only 16% of the variance. One can explain 74% of the variance in the difficulty of the reasoning problems if one allows a differential weighting of the basic rules. Hence, it appears that basic rules differ as to their difficulty. However, rule theories offer no explanation as to why there are differences between basic rules. Only two of the main effects, found in the analysis with P1 to P4 as predictors, can be explained by rule theories without allowing for a differential difficulty in the basic rules: A modus tollens is harder than a modus ponens (four basic rules[1] vs. one), and a problem with a false conclusion is harder than a problem with a true conclusion (detecting a contradiction requires the application of an additional basic rule). The other main effects cannot be accounted for by rule theories if no differential difficulty is allowed for the basic rules: A modus ponens plus conjunction is easier than a problem with a disjunctive modus ponens or a disjunctive syllogism (despite the first problems' requiring two basic rules, vs. one for the latter problems), and a disjunctive syllogism is more difficult than a disjunctive modus ponens (despite the fact that both require the application of one basic rule).

Before we interpret these results as *intrinsic* to the basic rules themselves (such as the representational cost of a basic rule), we consider the possibility that they are merely a reflection of the particular way the different inferences are combined in the problems of Experiment 1 (i.e., that they are *extrinsic* to the basic rules). If the latter were the case, the rule theories would still have to be supplemented to account for our results, but an explanation would then more likely be found in a specification of principles concerning the way people combine basic rules than in a principle concerning the intrinsic difficulty of basic rules. For each of (1) the difference between a disjunctive modus ponens and a disjunctive syllogism and (2) the difference between these two and a modus ponens plus conjunction, an alternative explanation can be provided in terms of the particular way the inferences of Experiment 1 were combined. We shall call these possible effects *combination effects*.

First, the difference between disjunctive modus ponens and disjunctive syllogism could be a consequence of a spontaneous preference for starting with a particular inference. For the reasoning problems with dependent inferences, a modus ponens or a modus tollens inference had to be made first, before a disjunctive syllogism or a disjunctive modus ponens could be applied. Dekeyser, Schroyens, Schaeken, Spitaels, and d'Ydewalle (2000)

observed that some people have a *preference for starting with the disjunction* when a problem contains a disjunction. If a majority of the participants in Experiment 1 had such a preference, this would have misled them in Problem Types 3 and 4 (see Table 1), because the disjunctive syllogism can be applied only after drawing the conditional inference (modus ponens or modus tollens). Hence, a strategy of starting with the disjunction would render these problems more difficult.

Second, the difference in difficulty between problems of Types 1 and 2 (see Table 1: modus ponens plus conjunction) and the four other problem types (with a disjunctive syllogism or a disjunctive modus ponens) could be caused by the fact that, for the latter, the second inference requires the result of the first, whereas for the former the inferences are independent (see Table 1). Hence, an alternative explanation for the observation that the problems of Types 1 and 2 are the easiest problems is that their inferences are independent. Note, however, that in order to rule out the existence of intrinsic differences between basic rules, the combination effect *inferences dependent versus independent* would have to be substantial: If there are no intrinsic differences between basic rules, the finding in Experiment 1 that problems with a disjunctive syllogism or a disjunctive modus ponens are more difficult than problems with a modus ponens plus conjunction (although the latter requires an additional basic rule) can only be accounted for if the combination effect on the difficulty is larger than the effect of a basic rule.

A second experiment was devised to test for these alternative explanations and to replicate the findings of Experiment 1 with a slightly different set of problems. Comparison of the results of Experiment 1 against the predictions of the mental model theory will be deferred until the alternative explanations have been tested.

EXPERIMENT 2

To test the possibility that the obtained difficulty effects are combination effects rather than the reflection of intrinsic differences between basic rules, for each of the eight problems of Experiment 1 with a disjunctive syllogism or a disjunctive modus ponens (Problem Types 3–6) and with a factory-related content, two variants were constructed. In one, the inferences remained dependent, but their order was reversed; in the other, the inferences were independent (hence, their order did not matter). This resulted in 3 * 8 problems: the eight problems of Experiment 1 plus two variants of each. We retained only eight problems from Experiment 1 because we wanted to keep the number of problems manageable for the participants and equal across experiments. Since the content type had no effect in Experiment 1, we retained only the problems with a factory-related content. In addition, since the effect of a preference for starting with a disjunction can be tested only when a disjunction is part of the problem and the effect of dependence between inferences can also be tested with such problems, we excluded the problems involving modus ponens plus conjunction.

For the eight problems identical to those in Experiment 1 with dependent inferences, modus ponens or modus tollens was the first inference. In Variant 1, we reversed the order of inferences; more specifically, disjunctive modus ponens or disjunctive syllogism was the first inference to be performed (independent of the presentation order of the premises). In Variant 2, the inferences were independent (in which case, the order of inferences did not matter). A problem from Experiment 1, with its two variants, is presented in Table 3. Other factors, such as presentation order of the premises, negations, and content, were, as far as possible, held constant within each triplet of problems.

Method

Participants. Two hundred and twenty-nine high school students between 16 and 19 years of age participated in the experiment.

Design and Procedure. Each participant had to solve 30 reasoning problems. Twenty-four problems were constructed using the following design.

The first factor was *modus ponens versus modus tollens* (12 problems of each type). The second factor was *disjunctive modus ponens versus disjunctive syllogism* (12 problems of each type). The third factor was the *truth-value of the conclusion* to be evaluated: true versus false (12 problems of each type).

The fourth factor was the *dependence and order of inferences*: Eight problems were dependent, as in Experiment 1, eight problems were dependent but with the reverse inference order, and eight problems were independent. The four factors were manipulated orthogonally.

The orthogonal manipulation results in 2 * 2 * 2 * 3 repeated measures.

As in Experiment 1, a fifth factor was the presentation order of the premises related to one inference. Premises of the same inference were presented either together or separately (12 problems of each type). Grouped premise presentation was used for problems with disjunctive syllogism and modus tollens and for problems with disjunctive modus ponens and modus ponens. Separate premise presentation was used for the other problems. This means that the

Table 3
Problem From Experiment 1 and Two Variants

Experiment 1	*Variant 1* (dependent, reversed order)	*Variant 2* (independent)
p	NOT p	p
q OR r	IF q, THEN r	r OR s
IF p, THEN NOT q	p OR q	IF p, THEN NOT q
----------	----------	NOT r
r	r	----------
		NOT q AND s

fifth factor is the interaction between *modus ponens versus modus tollens* and *disjunctive syllogism versus disjunctive modus ponens*.

The position and number of negations were held constant across variants as far as possible, with the restriction that the problems were always such as to avoid the necessity of performing double negations (according to rule theories).

As in Experiment 1, there were also six problems with an undecidable conclusion.

The procedure of this second experiment was the same as that of Experiment 1.

Results and Discussion

The alternative explanations for the two effects observed in Experiment 1 were tested by comparing the problems of Experiment 1 with the corresponding Variant 1 and Variant 2 problems. We used Wilcoxon tests for large samples. The alternative account for the greater difficulty of disjunctive syllogism, in terms of the order of the dependent inferences, was tested by comparing the number of correct answers on the four problems that had a disjunctive syllogism retained from Experiment 1 (disjunctive syllogism as second inference) with the number of correct answers on the four corresponding Variant 1 (disjunctive syllogism as first inference) problems. The mean number of correct answers was 2.7 versus 2.2. This difference is in the direction opposite to that predicted by the alternative account and was significant ($z = -5.23$, $p < .001$, two-tailed).

The alternative account for difficulty differences, in terms of whether or not the inferences are dependent, was tested by comparing the number of correct answers on the eight dependent problems retained from Experiment 1 with the number of correct answers on the eight problems of Variant 2 (independent inferences). The means were 5.5 and 5.7, respectively ($z = 1.82$, $p < .05$, one-tailed). Although the effect was significant and in the direction predicted by the alternative account, it is too small to explain the substantial difference found in Experiment 1 between modus ponens plus conjunction (when the mean for the eight problems was 6.84), on the one hand, and disjunctive syllogism and disjunctive modus ponens, on the other hand (means were, respectively, 4.66 and 5.87), as a combination effect.

Hence, the relevant results of Experiment 1 cannot be explained by the effects found in Experiment 2, which makes an interpretation in terms of intrinsic differences in the difficulty of basic rules more plausible. Nevertheless, our results indicate that combination factors do play a role in the difficulty of more complex problems: There is a rather substantial effect of the order of inferences, although this is in the direction opposite to our prediction, and a smaller effect of the inferences being dependent or not.

As in Experiment 1, we also tested the effect of the number of basic rules. The range in proportions of correct answers for the 24 problems of Experiment 2 was large: from .38 to .97, with a mean of .68 (see Table 4). Again, the logit of the proportion of correct answers was used in a further analysis. The logit did not correlate with the number of words in the problem ($r = -.027$, n.s.).

The number of basic rules explained 36% of the variance in the logit [$F(1,22) = 12.18$, $p < .01$; see Appendix A for the construction of the predictor]. The Spearman rank correlation between number of basic rules and the logit of the proportion of correct answers was computed to test for a monotonic (not necessarily linear) relation. The Spearman correlation was $-.50$ ($p < .01$, one-tailed). The 36% explained variance in this experiment is substantially higher than the 16% of Experiment 1, which is probably due to the particular basic rules involved. In Experiment 1, modus ponens + conjunction (two basic rules) was easier than both disjunctive syllogism and disjunctive modus ponens (one basic rule). This contradicts the prediction that the number of basic rules is positively related to difficulty. In Experiment 2, no modus ponens + conjunction problems were included, so that the number of basic rules was a better predictor. Hence, the better fit follows from the particular choice of basic rules.

The crucial difference between basic rules that both is unpredicted by the number of basic rules and can be tested in Experiment 2 is the difference between disjunctive syllogism and disjunctive modus ponens. This was tested in two ways. First, when a binary variable (1 for a disjunctive syllogism, -1 for a disjunctive modus ponens) was added as a predictor to the number of basic rules, the two predictors explained 44% of the variance. The standardized regression weight of the additional variable was $-.30$, which is significantly negative [$t(23) = -1.81$, $p < .05$, one-tailed], confirming the results of Ex-

Table 4
Proportion of Correct Answers for 24 Problems of Experiment 2

Problem Type		Problems of Experiment 1	Variant 1	Variant 2
	True Conclusion			
Disjunctive syllogism	modus ponens	.92	.55	.71
	modus tollens	.77	.66	.72
Disjunctive modus ponens	modus ponens	.97	.97	.89
	modus tollens	.53	.74	.67
	False Conclusion			
Disjunctive syllogism	modus ponens	.48	.62	.60
	modus tollens	.50	.38	.78
Disjunctive modus ponens	modus ponens	.90	.91	.71
	modus tollens	.40	.41	.58

periment 1. A second way to compare disjunctive syllogism with disjunctive modus ponens is to use Wilcoxon tests, as in the assessment of combination effects. The mean score for the 12 problems with a disjunctive syllogism was 7.69; for the 12 problems with a disjunctive modus ponens, the mean score was 8.68. The latter is significantly larger ($z = 6.22$, $p < .001$, one-tailed), again confirming our hypothesis.

The results of Experiment 2 corroborate the findings and conclusions of Experiment 1 concerning the existence of intrinsic differences in difficulty between basic rules. We have seen that rule theories offer no account for such differences. However, one may wonder how much of the variance in the difficulties of the problems can be explained on the basis of the mental model theory. This theory may be able to explain the intrinsic differences in difficulty between basic rules.

THE CONTRIBUTION OF THE MENTAL MODEL THEORY

In mental model theory, each explicit model *represents* a possible situation. Hence, the number of explicit models needed (to be "fleshed out") to solve an inference can be considered to be the representational cost of the inference. This notion of representational cost can be used to account for several results, including some of the intrinsic differences between basic rules inferred from our experiments. The mental model theory can account for the fact that modus ponens (one model) is easier than modus tollens (three models), a difference that is not due to an intrinsic difference between basic rules. Furthermore, it predicts that modus ponens + conjunction (one model; the conjunction to be made in this experiment does not require the construction of a model, but is merely the "reading" of a model containing two elementary propositions) is easier than disjunctive syllogism and disjunctive modus ponens (two models). The results of Experiment 2 suggest that the latter is an intrinsic difference between basic rules.

However, the mental model theory offers no clear prediction concerning the relative difficulty of disjunctive syllogism versus disjunctive modus ponens. Both require the construction of two models (according to Johnson-Laird et al., 1992). One might argue that the mental model theory predicts the difference in difficulty that we found between disjunctive syllogism and disjunctive modus ponens, because in a disjunctive syllogism, an inconsistency has to be detected (between one of the disjuncts and the categorical premise), whereas this is not needed in a disjunctive modus ponens. On the other hand, the two explicit models of the initial representation of the disjunction "p OR q" are smaller than those of the conditional "if p OR q, THEN r" (see the introduction for the initial representation). Hence, one could argue equally well that the mental model theory predicts the reverse of the obtained effect.

We will now try to complement the rule theories with the notion of representational cost, as conceived in the mental model theory. Of course, from the previous evaluation, it is clear that this will not result in a perfect account of the data. Nevertheless, it is useful to see how much we can improve upon our prediction of the difficulty and, also, to see where we fail when we take the notion of representational cost into account. A possible way of complementing the rule theories with the notion of representational cost is to weight the different inferences by their representational cost, as predicted by the mental model theory (the number of explicit models). The value of a problem on this new predictor of problem difficulty is then the sum of the representational costs, over the inferences to be made. Appendix B describes in detail how the value on this predictor for the different problems of Experiments 1 and 2 was counted.

The predictor explains 56% of the variance in Experiment 1 [$F(1,22) = 28.5$, $p < .001$] and 41% of the variance in Experiment 2 [$F(1,22) = 15.08$, $p < .001$]. The new predictor can be compared with the predictor *number of basic rules* by testing whether the correlation of the former with difficulty differs from the correlation of the latter with difficulty. For Experiment 1, the correlation with the logit of the proportion of correct answers is significantly more negative for the predictor *inferences weighted by the number of explicit models* than for the predictor *number of basic rules* [the correlations are $-.75$ and $-.40$, respectively; $t(21) = 3.908$, $p < .001$]. For Experiment 2, no significant difference was obtained [the correlations are $-.64$ and $-.60$, respectively; $t(21) = 0.76$].

The Spearman rank correlations of the predictor *inferences weighted by the number of explicit models* with the logit of the proportion of correct answers are $-.76$ ($p < .001$) and $-.58$ ($p < .01$) for Experiments 1 and 2.

In Experiment 1, weighting the inferences with the number of explicit models needed to perform them results in a considerable improvement over the *number of basic rules* predictor. Nevertheless, in Experiment 1, a large part of the variance remains unexplained (56% vs. 74% when a differential weight is allowed for different basic rules). This is not very surprising, because, as we have already pointed out, the mental model theory does not predict the difference between disjunctive modus ponens and disjunctive syllogism. If P3 (second inference is disjunctive syllogism or not) is added as a predictor to the mental model predictor, 74% of the variance is explained [$F(1,21) = 13.58$, $p < .01$], reaching the same level of prediction as the four factors of the design. When in addition, we add the only interaction effect that was found, P1 * P3 (the interaction between *modus ponens versus modus tollens* and *disjunctive syllogism or not*), the explained variance amounts to 83% [$F(1,20) = 10.68$, $p < .01$]. With respect to Experiment 2, when the factor coding for *disjunctive syllogism versus disjunctive modus ponens* is added, 49% of the variance is explained, which is slightly more than the percentage explained by

the mental model theory predictor [$F(1,21) = 3.61$, $p < .10$]. A considerable improvement of the latter prediction is obtained when the same interaction as that found in Experiment 1 (*modus ponens versus modus tollens* and *disjunctive syllogism versus disjunctive modus ponens*) is added as a predictor. The three predictors explain 71% of the variance, which is more than the percentage explained by the previous two predictors [$F(1,20) = 14.63$, $p < .01$]. The effect of the interaction is in the same direction as that in Experiment 1 (beta = −.46). Although we have no explanation for this interaction effect, it is clearly a combination effect: It results from the particular way the inferences are combined in the design of our experiments.

In conclusion, although the incorporation of the number of models in the regression equations results in a better prediction, a substantial amount of the variance remains unexplained. The number of models seems to be too coarse a notion of representational cost. The notion probably needs refining in order to account for the unexplained intrinsic difference between *disjunctive syllogism* and *disjunctive modus ponens*. Furthermore, an account of the combination effect that could be derived from the interaction between *disjunctive syllogism versus disjunctive modus ponens* and *modus ponens versus modus tollens*, would also add to the explanatory power. Finally, the other combination effects we have found in Experiment 2 also necessitate a further elaboration of the theory. A more detailed examination of the different processes entering into a reasoning task would probably add to the understanding of these effects. It might be that the differences (intrinsic as well as combination effects) can be captured in an elegant way by a more fine-grained notion of representational cost that takes into account what must be represented at a particular moment in time and how long it has to remain represented. More particularly, a protocol analysis of complex reasoning problems, such as those of Experiment 2, may result in a better understanding of the different inferential steps people take.

CONCLUDING REMARKS

Rule theories cannot account for the difficulty of the complex reasoning problems used in our two studies, unless they allow a differential weight for different basic rules. The differences in the weights themselves remain unaccounted for. On the basis of the principles of the mental model theory, we hypothesized that the notion of representational cost offers an explanation for at least part of these differences. Weighting the inferences to be made in a rather complex reasoning problem by the number of models required to represent them results in a better fit, but this measure seems too coarsely defined to capture all of the variance. Nevertheless, elaborating the notion of representational cost seems to be a promising approach. By examining the processes underlying reasoning in more detail, it should be possible to quantify the representational cost of a problem more precisely and to arrive at a better account of the data.

REFERENCES

Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, **85**, 1-21.

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

Braine, M. D. S., Reiser, B. J., & Rumain, B. (1984). Some empirical justification for a theory of natural propositional logic. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 18, pp. 313-371). New York: Academic Press.

Dekeyser, M., Schroyens, W., Schaeken, W., Spitaels, O., & d'Ydewalle, G. (2000). Preferred premise order in propositional reasoning: Semantic informativeness and co-reference. In A. Vandierendonck, G. d'Ydewalle, G. De Vooght, & W. Schaeken (Eds.), *Deductive reasoning and strategies* (pp. 73-95). Mahwah, NJ: Erlbaum.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.

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

Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review*, **99**, 418-439.

Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1994). Why models rather than rules give a better account of prepositional reasoning: A reply to Bonatti and to O'Brien, Braine, and Yang. *Psychological Review*, **101**, 734-739.

Macnamara, J. (1986). *A border dispute: The place of logic in psychology*. Cambridge, MA: MIT Press, Bradford Books.

Osherson, D. N. (1974). *Logical abilities in children: Vol. 2. Logical inference: Underlying operations*. Hillsdale, NJ: Erlbaum.

Osherson, D. N. (1975). *Logical abilities in children: Vol. 3. Reasoning in adolescence: Deductive inference*. Hillsdale, NJ: Erlbaum.

Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review*, **90**, 38-71.

Rips, L. J. (1994). *The psychology of proof*. Cambridge, MA: Routledge.

Rips, L. J., & Conrad, F. G. (1983). Individual differences in deduction. *Cognition & Brain Theory*, **6**, 259-289.

NOTE

1. As was explained earlier, the four reasoning steps to solve a modus tollens are, strictly speaking, not "basic rules," because they do not belong to the direct reasoning program. Equating the reasoning steps with basic rules, as we do for pragmatic reasons, does not change the prediction that modus tollens is harder than modus ponens, however. This prediction does not change if a biconditional interpretation is assumed, either.

APPENDIX A
The Value of the Problems on the Predictor Number of Basic Rules

For the problems of Experiment 1 with a true conclusion, the predictor was defined as follows for the six problem types (see Table 1):

Problem Type 1: 1 (modus ponens) + 1 (modus ponens) + 1 (conjunction to combine the two intermediate conclusions of the conditional inferences) = 3.

Problem Type 2: 4 (modus tollens) + 1 (modus ponens) + 1 (conjunction to combine the two intermediate conclusions of the conditional inferences) = 6.

Problem Type 3: 1 (modus ponens) + 1 (disjunctive syllogism) = 2.

APPENDIX A (Continued)

Problem Type 4: 4 (modus tollens) + 1 (disjunctive syllogism) = 5.

Problem Type 5: 1 (modus ponens) + 1 (disjunctive modus ponens) = 2.

Problem Type 6: 4 (modus tollens) + 1 (disjunctive modus ponens) = 5.

For the two sets of six problems of Experiment 1 with a false conclusion, solving the contradiction between the derived conclusion and the conclusion to be evaluated required the application of one more basic rule: contradiction. Hence, the values of the predictor for these problems were respectively 4, 7, 3, 6, 3, and 6 for the six problem types.

In Experiment 2, Problem Types 3–6 were used. The number of basic rules for the Variant 1 problems (dependent inferences, reversed order) is the same as for the original problems. For the Variant 2 problems (independent inferences), one additional basic rule has to be applied in order to combine the intermediate conclusions of the two other inferences (similar to the combination of the two independent inferences of the problems of Types 1 and 2 in Experiment 1).

APPENDIX B
The Predictor Constructed by Weighting Each Inference by the Number of Explicit Models Needed to Make the Inference

The mental model predictor for the complex reasoning problems was based on the account of Johnson-Laird et al. (1992) for *elementary* inferences, the crucial determinant being the number of explicit models to be constructed to solve these inferences (see the Introduction for the number of explicit models needed for the different inferences). For the problems of Experiment 1 with a true conclusion, the models were counted as follows:

Problem Type 1: 1 (modus ponens) + 1 (modus ponens) = 2.

Problem Type 2: 3 (modus tollens) + 1 (modus ponens) = 4.

Problem Type 3: 1 (modus ponens) + 2 (disjunctive syllogism) = 3.

Problem Type 4: 3 (modus tollens) + 2 (disjunctive syllogism) = 5.

Problem Type 5: 1 (modus ponens) + 2 (disjunctive modus ponens) = 3.

Problem Type 6: 3 (modus tollens) + 2 (disjunctive modus ponens) = 5.

The figures mentioned refer to the number of explicit models. In Experiment 2, Problem Types 3–6 were used. For the two variants of the problems also used in Experiment 1, the number of models is the same as that for the original problems, because the same inferences have to be made (except for a conjunction, but we assume no additional model for it, as will be explained next).

For problems of Types 1 and 2 of Experiment 1 and for the problems with independent inferences of Experiment 2, there is no additional explicit model assumed for the combination of the intermediate conclusions of the two first inferences, because the conjunction can be considered as a simple "reading" of the resulting model containing the two intermediate conclusions. Indeed, each model is isomorphic with a conjunction (Johnson-Laird et al., 1992, p. 436). Including an additional model for it would worsen the prediction: The percentages of explained variance drop to 35% and 36%, for Experiments 1 and 2, respectively.

If the conclusion to be evaluated was false, the value of the predictor was augmented by one, since the detection of the inconsistency between the conclusion read out from the resultant mental model and the given conclusion requires an additional reasoning operation (again, not including this additional model would worsen the prediction; the percentages of explained variance drop to 47% and 26%, for Experiments 1 and 2, respectively). Hence, the value of the predictor for the false problems were respectively 3, 5, 4, 6, 4, and 6 in Experiment 1 and 4, 6, 4, and 6 in Experiment 2 for the different problem types.

APPENDIX B (Continued)

When "if/then" is interpreted as a biconditional, only two explicit models are needed to perform a modus tollens (Johnson-Laird et al., 1992, p. 431). If we counted two (instead of three) models for a modus tollens, the mental model predictor explains 60% and 42% of the variance in Experiments 1 and 2, respectively, which is marginally better.

Johnson-Laird and Byrne (1991, p. 52) also assume that the process of fleshing out per se adds to the difficulty of a deduction, in addition to the representational cost of the fleshed-out models. For our problems, this is the case for the problems with a modus tollens inference. This can be incorporated into our predictor by augmenting the predictor by one for these problems. The percentages of explained variance are then 50% and 38%, for Experiments 1 and 2, respectively. The prediction is worse because the higher difficulty of problems with a modus tollens inference is already explained by a higher number of explicit models, so that further augmentation is not needed. In fact, it turns out to slightly diminish the quality of the prediction.

Note that the interpretation of the conditional (conditional vs. biconditional) and the issue of whether the need for fleshing out is an additional source of difficulty do not change the qualitative predictions for the elementary inferences, as was discussed above. In all cases, a modus tollens is predicted to be more difficult than a modus ponens.

While constructing the mental model predictor, we had to make some choices that are related to our problems being complex reasoning problems.

1. In Johnson-Laird et al. (1992), the order in which premises are represented is determined by the order of presentation (first premise is represented first, etc.). In a later publication, Johnson-Laird, Byrne, and Schaeken (1994) state that "reasoners will tend to start with a premise that is maximally informative (e.g., a categorical assertion), and they will tend to work in an order that maintains co-reference, drawing, when possible, informative intermediate conclusions" (p. 736). According to the latter, the mental model theory makes no differential predictions for the two presentation orders of the premises. This is corroborated by the results of Experiment 1, since the factor coding for the presentation order did not improve the prediction (see the Results section for Experiment 1). Therefore, the mental model predictor we constructed did not take into account the presentation order of the premises.

2. Johnson-Laird et al. (1992) used their "psychological algorithm" to determine the number of explicit models needed to solve complex reasoning problems. The psychological algorithm does not incorporate a fleshing-out procedure, however. In consequence, we could not follow the algorithm for problems involving a modus tollens inference, since these problems require that the initial model be fleshed out. Instead, we counted three models for a modus tollens inference, since this is the number of explicit models the mental model theory assumes for an elementary modus tollens inference.

(Manuscript received September 2, 1999;
revision accepted for publication June 2, 2000.)