Are conjunctive inferences easier than disjunctive inferences? A comparison of rules and models

J. A. García-Madruga, S. Moreno, N. Carriedo, and F. Gutiérrez
Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain

P.N. Johnson-Laird
Princeton University, Princeton, New Jersey, USA

We report four experiments investigating conjunctive inferences (from a conjunction and two conditional premises) and disjunctive inferences (from a disjunction and the same two conditionals). The mental model theory predicts that the conjunctive inferences, which require one model, should be easier than the disjunctive inferences, which require multiple models. Formal rule theories predict either the opposite result or no difference between the inferences. The experiments showed that the inferences were equally easy when the participants evaluated given conclusions, but that the conjunctive inferences were easier than the disjunctive inferences (1) when the participants drew their own conclusions, (2) when the conjunction and disjunction came last in the premises, (3) in the time the participants spent reading the premises and in responding to given conclusions, and (4) in their ratings of the difficulty of the inferences. The results support the model theory and demonstrate the importance of reasoners' inferential strategies.

Reasoning is an important human ability, which underlies science, mathematics, and thinking in everyday life. Naïve individuals—those untrained in logic—are able to make *valid* deductions, that is, those for which the conclusion must be true if the premises are true. They can also make such inferences about matters of which they have no general knowledge. Psychologists have proposed two main types of theory to explain such general deductive competence, and we will examine how these theories account for one of its core components: inferences based on sentential connectives, such as "if", "or", and "and".

The first type of reasoning theory depends on *formal rules* of inference akin to those of a logical calculus, couched in a form known as "natural deduction" (see, e.g., Rips, 1994; Braine &

Requests for reprints should be sent to Juan A. García Madruga, Dpto. de Psicología Evolutiva, Facultad de Psicología–UNED, 28040 Madrid, Spain. Email: jmadruga@psi.uned.es

We thank Dr. Lance J. Rips for providing us with the structure of some of the problems used in the present study. We also thank Ruth Byrne, Steve Newstead, and an anonymous reviewer for helpful comments on an earlier draft. The research was part of the "Comprehension and inference in conditional reasoning" project funded by the Spanish Ministry of Education and Science (PB94–0394).

http://www.tandf.co.uk/journals/pp/02724987.html DOI:10.1080/02724980042000309

O'Brien, 1998a, b). According to these theories, reasoning consists of a process of proof that proceeds in a series of steps, each permitted by a formal rule of inference. The main prediction of formal rule theories is that the length of a proof should correspond to the difficulty of the inference, though other factors, such as the complexity of rules, can also affect difficulty (see, e.g., Braine, 1990; Rips, 1994).

The second type of theory of reasoning is based on *mental models* (e.g., Johnson-Laird & Byrne, 1991; cf., Polk & Newell, 1995). According to this theory, naïve reasoners conceive the possibilities consistent with the premises: They construct mental models of them, where each model corresponds to a possibility, and they establish the validity of an inference by checking whether its conclusion holds in the models of the premises. A conclusion is necessary—it *must* be the case—if it holds in all the models of the premises; it is possible—it *may* be the case—if it holds in at least one of the models; and it is probable—it is *likely* to be the case—if it holds in most of the models. Assuming that each model is equiprobable, the probability of a conclusion depends on the proportion of models in which it holds (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). In order to minimize the load on working memory, reasoners represent as little information as possible.

The principal prediction of the model theory is that the greater the number of models that have to be constructed to make an inference, the harder the inference will be—it will take longer and lead to more errors. Table 1 shows the models predicted by the theory for assertions constructed from the main sentential connectives. Various phenomena corroborate the principal prediction; for example, inferences based on inclusive disjunctions (three models) are

TABLE 1
Models for the main sentential connectives

Connective	*Mental models*		*Fully explicit models*	
A and B	A	B	A	B
A or else B	A		A	¬B
		B	¬A	B
A or B, or both	A		A	¬B
		B	¬A	B
	A	B	A	B
If A then B	A	B	A	B
	. . .		¬A	B
			¬A	¬B
If and only if A then B	A	B	A	B
	. . .		¬A	¬B

Note: The central column shows the mental models postulated for human reasoners, and the right-hand column shows fully explicit models, which represent the false components of the true cases by using negations that are true: "¬" denotes negation and ". . ." denotes a wholly implicit model. Each line represents a model of a possibility. The mental models for "if" and for "if, and only if" differ in their mental footnotes; in either case, individuals tend to forget or to omit the implicit model when a conditional occurs in complex assertions.

harder than the corresponding inferences based on exclusive disjunctions (two models; see, e.g., Bauer & Johnson-Laird, 1993; Johnson-Laird, Byrne, & Schaeken, 1992).

These results and others, however, do not count against formal rule theories. Such theories cannot in general make predictions about differences in difficulty between different rules of inference. Instead, they use experimental data to estimate the difficulty of each rule of inference. Hence, a problem in comparing the two theories is that neither is "fully and precisely defined" (Evans & Over, 1997, p. 27). Indeed, it is a rare psychological theory that is defined sufficiently explicitly to distinguish it from similar theories—their boundaries of application are notoriously underdefined. Fortunately, the two types of theory do make contrasting predictions about a central component of sentential reasoning. Psychology is difficult, but not impossible.

At the heart of sentential reasoning are inferences based on conjunctions and on disjunctions. Consider the following two examples (from Rips, 1990, 1994, pp. 365–9):

(1) p and q	(2) p or q
If p then r	If p then r
If q then r	If q then r
∴ r	∴ r

We will examine how the two main formal rule theories account for these inferences. Braine's theory of mental logic has three parts (see Braine, 1990; Braine & O'Brien, 1998b): A set of rules (inference schemas), a reasoning programme that applies the rules in a direct reasoning routine (DRR) and in some indirect reasoning strategies, and some pragmatic principles. Inferences (1) and (2) require only the automatic application of some basic rules by the direct reasoning routine. The conjunctive inference depends on using a rule to detach a conjunct from a conjunction (A and B, ∴ A) and then the rule for modus ponens (If A then B, A, ∴ B):

1) p and q
If p then r
If q then r
∴ p (detaching a conjunct)
∴ r (modus ponens)

In contrast, the disjunctive inference (2) depends only on a single application of the rule for *dilemmas*: A or B, If A then C, If B then C, ∴ C (see Braine & O'Brien, 1998b, p. 80).

Rips' (1994) PSYSCOP theory is also based on the notion of mental proof. The process of proving conclusions from premises depends on two types of rule. Forward rules work forward from the premises to derive a conclusion. Backward rules work backwards from a conclusion to yield subgoals—that is, intermediate conclusions or premises that suffice for the derivation of that conclusion (Rips, 1994, pp. 103–105). The theory's account of the conjunctive and disjunctive inferences is almost identical to Braine's account. The conjunctive inference (1) calls for the use of the forward AND rule, which detaches a conjunct, and then the forward IF-elimination rule, which carries out a modus ponens inference. Likewise the disjunctive inference (2) calls only the forward dilemma rule (Rips, 1994, p. 368). In short, according to formal rule theories, the conjunctive inference (1) calls for two steps using two rules, whereas the disjunctive inference (2) calls for just a single step using only the single rule for dilemmas. Table 2 summarizes the formal rules needed for these two inferences and for another two inferences that are similar in form, apart from the introduction of negations (3 and 4).

Granted the preceding analysis, which inference should be easier according to the formal rule theories? If the length of the derivation is the critical factor, then the disjunctive inference should be easier than the conjunctive inference. But, if the ease of application of the rules is the critical factor, then the disjunctive rule may be harder than the cumulative difficulty of the two rules needed for the conjunctive inference. There are, however, empirical results that provide estimates of the difficulty of using the relevant rules. Thus, according to Braine and O'Brien (1998b, p. 80), the difficulty weights for the two rules in the conjunctive inference are .41 for the rule for detaching a conjunct and .47 for the rule of modus ponens, whereas the difficulty weight for the dilemma rule needed for the disjunctive inference is .16. These weights were estimated from an empirical study asking people to rate the difficulty of 85 different problems (Braine, Reiser, & Rumain, 1984). According to these authors, the "difficulty of a problem should depend of the sum of the difficulty weights of the component reasoning steps" (Braine, Reiser, & Rumain, 1998, p. 104). Hence, Braine and his colleagues predict that the conjunctive inference should be harder than the disjunctive inference. In contrast, Rips (1994, pp. 368–9) argues on intuitive grounds that the rule for a dilemma is "somehow harder" to use given that it calls for coordinating three premises, whereas the rules for the conjunctive inference apply to less than three premises. He therefore concludes that there is no reason to suppose that one of the two inferences should be easier than the other.

The mental model theory makes a simple direct prediction on the basis of the number of models that reasoners need to construct. The conjunctive inference should be easier than the

TABLE 2
The conjunctive and disjunctive inferences with their required formal rules of inference

	Inferences		*Rules*
One-model inferences (conjunctions)	(1) p and q if p then r if q then r ∴ r	(3) not-p and not-q if not-p then r if not-q then r ∴ r	Forward And elimination (detachment of conjunct) *P AND Q* P Forward If elimination (modus ponens) IF P THEN R *P* R
Multiple-model inferences (disjunctions)	(2) p or q if p then r if q then r ∴ r	(4) not-p or not-q if not-p then r if not-q then r ∴ r	Forward dilemma (core schema) P OR Q IF P THEN R *IF Q THEN R* R

disjunctive inference, because conjunctions call for only a single model whereas disjunctions call for at least two models. Indeed, Johnson-Laird et al. (1992) reported some corroboratory evidence in a re-analysis of results from Braine et al. (1984). The first premise, *p and q*, in the conjunctive inference (1) elicits the mental model:

```
P    Q
```

where P denotes a representation of the proposition expressed by p, and Q denotes a representation of the proposition expressed by q. According to the model theory, reasoners normally represent conditionals with two mental models; one is completely explicit and represents the possibility in which the antecedent is true, and the other is wholly implicit and represents the possibilities in which the antecedent is false. Reasoners who make a mental footnote that the implicit model represents the falsity of the antecedent can, in principle, flesh out the models to make them wholly explicit (see Table 1). Hence, the second premise, *If p then r*, elicits the mental models:

```
P    R
. . .
```

The result of combining the models for the two premises is:

```
P    Q    R
```

which supports the conclusion, r. (Reasoners who fail to notice this point may go on to combine this model with the models for the third premise, *if q then r*, then to yield the same overall model.) In contrast, the first premise, *p or q*, of the disjunctive inference (2) calls for the following mental models:

```
P
     Q
```

and reasoners may represent the possibility of the joint occurrence of p and q if they treat the disjunction as inclusive. If these models are combined with those for the second premise, *if p then r*, the result is the following set of models:

```
P         R
     Q
```

These models can, in turn, be combined with the models of the third premise, *if q then r*, to yield:

```
P         R
     Q    R
```

These models support the correct conclusion, r. The fully explicit (and therefore correct) models are as follows:

$$\begin{array}{ccc} P & \neg Q & R \\ \neg P & Q & R \end{array}$$

The conjunctive inference (1) calls for fewer models to be constructed than the disjunctive inference (2), and the same contrast occurs for inferences (3) and (4) containing negations (see Table 2). Hence, according to the model theory, the conjunctive inferences should be easier than the disjunctive inferences.

Rips (1994) remarks that the model theory's prediction is dubious, because of the results of an experiment that he carried out in which the participants evaluated the two sorts of inference (see also Rips, 1990). One group carried out the affirmative inferences (1) and (2), and another group carried out the negative inferences (3) and (4). Both groups also carried out two invalid filler inferences. The task was to decide whether each conclusion was "necessarily true" or "not necessarily true". The overall difference between one-model and multiple-model problems was not reliable. As Table 3 (on page 9) suggests, there was no reliable difference between the two affirmative inferences (in the first column of the table), whereas with the negatives (in the second column), the conjunctive inference was easier than the disjunctive inference, but the difference was only marginally significant. "These results", Rips wrote, "suggest that the need to keep track of multiple models was not the source of the subjects' difficulties" (Rips, 1994, p. 368). They also suggest that formal rule theories are right—granted the assumption that the dilemma rule is harder to use than the two rules needed for the conjunctive inferences.

To the best of our knowledge, Braine and his colleagues have not studied the conjunctive inference (1). However, Braine, Reiser, and Rumain (1998, p. 139) reported a study in which participants rated the difficulty of the disjunctive inference (2) in an evaluation task. The mean rating for the inference was 2.61 (where 1 signified "easiest" and 9 signified "hardest"). Likewise, Braine et al. (1998, p. 158) reported that almost all the participants in two experiments responded correctly to the disjunctive inference (Experiment 1: 100%, and Experiment 2: 97%). However, the order of the premises and the nature of the task in these experiments is unclear—two factors that, as we will see, turn out to be crucial. The aim of the present paper is accordingly to re-examine the relative difficulty of the conjunctive and disjunctive inferences and to report some new experiments comparing them.

EXPERIMENT 1

The inferences studied by formal-rule theorists tend to be relatively easy, and the participants usually have to evaluate given conclusions (Rips, 1994; Yang, Braine, & O'Brien, 1998). Evaluation makes inferences still easier because one is saved the labour of putting conclusions into words, and has a good chance of guessing the right answer. An important part of deduction, however, is the ability to frame one's own conclusions. The task can also be informative because it may reveal biases in reasoners' formulation of conclusions, whether correct or not, such as the figural bias in syllogistic reasoning (see, e.g., Johnson-Laird & Byrne, 1991).

Granted the relative ease of evaluating a given conclusion, Rips' experiment might not have been sensitive enough to detect a difference between the conjunctive and disjunctive inferences in Table 2. For these inferences, the formal rule theories do not predict any major difference between evaluating given conclusions and constructing one's own conclusions. In both cases, reasoners should apply the same formal rules. According to Braine and O'Brien's theory, the direct reasoning routine is applied automatically whenever the conditions of application are satisfied—that is, independently of whether the task is to evaluate a conclusion or to generate it. Rips' theory (1994) makes a similar claim. Forward rules can be directly applied to premises in order to derive conclusions. Moreover, forward rules are automatically used so often that it "seems reasonable to activate them as soon as possible whenever a triggering assertion appears in the database" (Rips, 1994, p.122).

The model theory, in contrast, places a much greater emphasis on the role of strategies in reasoning. It suggests that reasoners may adopt different strategies in the two kinds of task. When reasoners have to evaluate a given conclusion, they can work backwards from the conclusion—that is, they can consider its models in order to determine whether they hold within the models of the premises. Such a strategy is impossible to apply when reasoners have to construct their own conclusions unless they first make a guess at a likely conclusion. One consequence is that the evaluation task should be easier than the construction task. Hence, the difference between the conjunctive and disjunctive inferences is more likely to occur in the construction task. The purpose of Experiment 1 was to test whether there was any difference between the evaluation and the construction of conclusions for the conjunctive and disjunctive inferences.

Method

Participants

A total of 91 students participated voluntarily in the experiment. One group of 53 participants carried out the evaluation task, and another group of 39 participants carried out the construction task. The participants had received no training in logic and had not been previously tested in any experiment on reasoning.

Design

Two independent groups, to which the participants were assigned at random, carried out a set of inferences. One group evaluated given conclusions, and the other group constructed their own conclusions from the premises of the same set of problems. Each participant in both groups carried out one version of the four types of problem in Table 2 (i.e., two conjunctive and two disjunctive inferences) and a further four "filler" inferences that did not support valid conclusions.

Materials

The content of the problems, as in Rips' experiment, concerned people and cities. The experiment was carried out in Spanish. An example of one of the problems in the evaluation task (translated into English) is as follows:

Ana is in Granada and Pablo is in Barcelona.
If Ana is in Granada then Teresa is in Toledo.
If Pablo is in Barcelona then Teresa is in Toledo.
Is Teresa in Toledo?
Yes [ ]   No [ ]

The only difference in the construction task was that the final question was replaced by:

What can you conclude? ("¿qué se puede concluir?")
Conclusion: ....

In the negative inferences, each clause in the premises except for the conclusion was negative (e.g., Ana is not in Granada and Pablo is *not* in Barcelona). The invalid "filler" problems were similar to the valid ones except for the polarity of the antecedents of the two conditionals, for example:

Ana is in Granada and Pablo is in Barcelona.
If Ana is not in Granada then Teresa is in Toledo.
If Pablo is not in Barcelona then Teresa is in Toledo.

We devised eight sets of different contents, each containing the names of people and places, and rotated them over the eight inferences in order to construct eight versions of the evaluative and the constructive tasks. Each participant was assigned at random to one of the eight versions of a task.

Procedure

The participants were tested together in a single group. The instructions and the problems were assembled into booklets, which were assigned to them at random. The instructions on the first page of the booklet explained the nature of the task. For the evaluation task, the participants were told that they had to decide whether or not the conclusions of the inferences were "necessarily true"—that is, whether they *must* be true given that the premises were true. Likewise, for the construction task, the participants were told that they had to state what, if anything, followed necessarily from each set of premises—that is, what *must* be true given that the premises were true. If the participants considered that nothing followed from the premises, then they had to write that nothing follows. There were two example problems, one based on modus ponens and one based on affirmation of the consequent, which were used to illustrate the meaning of "necessarily true". The participants were told that they could take as much time as they needed on a problem, but that they should not return to a problem once they had answered it.

Results and discussion

Table 3 presents the percentages of correct conclusions drawn in the experiment, and, for purposes of comparison, the results from Rips' experiment. Overall, our results showed that the participants were reliably more accurate for conjunctive problems (93% correct) than for disjunctive problems (79% correct; Wilcoxon test, $n = 91$; $z = 2.71$; $p < .003$, one-tailed). But, as predicted, there was a reliable interaction: the effect of number of models was significantly greater in the construction task than in the evaluation task (Mann–Whitney U test, $n_1 = 53$, $n_2 = 39$, $z = 1.86$, $p < .04$, one-tailed). Indeed, there was no reliable difference between the conjunctive inferences (91% correct) and the disjunctive inferences (85% correct) in the evaluation task (Wilcoxon test, $n = 52$, $z = 1.19$, $p > .2$, one-tailed), either for the affirmative or for the negative inferences. In contrast, as predicted, the conjunctive inferences (95% correct) were reliably easier than the disjunctive inferences (72% correct) in the construction task (Wilcoxon test, $n = 39$, $z = 2.58$, $p < .01$, one-tailed), and the difference was reliable for both the affirmative and the negative inferences (Wilcoxon tests, $n = 39$, $z = 2.24$, $p < .02$, one-tailed, and $z = 2.35$, $p < .01$, one-tailed, respectively).

We carried out an error analysis of the responses in the construction task. There was a total of 26 errors—that is, an overall error rate of 16%. Most of the errors (81%) were "no valid

TABLE 3
The percentages of correct responses in Experiment 1

	Inferences	% Correct Evaluate	% Correct Construct	Inferences	% Correct Evaluate	% Correct Construct
One-model inferences (conjunctions)	p and q if p then r if q then r ∴ r	94 (89)	97	not-p and not-q if not-p then r if not-q then r ∴ r	89 (95)	92
Multiple-model inferences (disjunctions)	p or q if p then r if q then r ∴ r	87 (89)	77	not-p or not-q if not-p then r if not-q then r ∴ r	83 (81)	67

Note: The column labelled "Evaluate" gives the results for the task in which the participants evaluated given conclusions, and it shows the results from Rips' (1994) experiment in parentheses. The column labelled "Construct" gives the results for the task in which the participants constructed their own conclusions.

conclusion" responses. Three errors were made in problems with negations by changing the polarity of the correct answer (saying *not-r* instead of *r*). Only two errors occurred outside these categories.

These results replicated Rips' study of the evaluation task, but they suggest that its failure to corroborate the mental model theory may depend on the nature of the task. In the case of the construction task, our results confirmed the mental model theory: The conjunctive inferences were easier than the disjunctive inferences. One possibility is that the evaluation task is not sensitive enough to detect a difference, because it is easy to guess the right answer. Another possibility is that reasoners adopt different strategies in the evaluation task—they can, for example, work backwards from the given conclusion, whereas this strategy is not possible in the construction task unless reasoners somehow guess the correct conclusion. Our second experiment accordingly examined a condition that may affect reasoners' strategies.

EXPERIMENT 2

A robust result in reasoning is that modus ponens inferences are easier than modus tollens inferences (see, e.g., Evans, Newstead, & Byrne, 1993): Modus ponens can be drawn from the mental models of the conditional premise, whereas modus tollens calls for these models to be fleshed out to make them fully explicit. However, as Girotto, Mazzoco, and Tasso (1997) have demonstrated, the difference is susceptible to a strategic manipulation in a way that was predicted by the model theory. As these authors argued, if the categorical premise is presented prior to the conditional premise, modus ponens should be unaffected, whereas modus tollens should be easier. When the categorical premise is presented second (in the usual order), reasoners have already interpreted the conditional premise, and the categorical information merely eliminates the explicit model of the conditional. It then seems that nothing follows from the premises, which is a common erroneous response. In contrast, when the categorical premise is presented first, reasoners represent it, and so they do not have to represent the model of the conditional in which the antecedent and consequent are true. They are more likely to flesh out the other models of the conditional, and hence to reach the valid conclusion.

Granted that the order of premises can influence deductive performance, we presented the premises in Experiment 2 in a different order. The conditional premises occurred prior to the conjunctive or disjunctive premise. We manipulated the order of the premises in the next experiment, but we held it constant in the present experiment because our main aim was to replicate the effect of constructing conclusions as opposed to evaluating them. When the conditional premises occur first, they are likely to preoccupy working memory while reasoners are interpreting the third premise. The additional load should have little or no effect on the inferences with conjunctions as third premises, because conjunctions call only for a single model—as was the case with modus ponens in the Girotto et al. (1997) study. But, the additional load is likely to affect inferences with disjunctions as third premises. The effect should be reliable, but it is unlikely to be massive because the disjunctive inferences—unlike modus tollens—can be made without having to flesh out mental models into fully explicit models. In sum, the conjunctive inferences should be easier than the disjunctive inferences for both the evaluation and the construction tasks.

Formal rule theories make no such predictions. On the contrary, granted the direct and automatic application of the relevant rules, a change in the order of the premises should not have a major effect on performance. In the case of the disjunctive inference, there should be no effect, because "the order of propositions in numerators of schemas is immaterial" (Braine & O'Brien, 1998b, p. 81). In the case of the conjunctive inferences, however, the schema of modus ponens needs to be "fed" with the output from the schema for detaching a conjunct. Hence, there might be a small deleterious effect of presenting the conjunction after the conditionals. Rips' theory makes comparable claims. Forward rules are automatically triggered by the occurrence of appropriate premises.

Method

Participants

A total of 92 students voluntarily participated in the experiment. One group of 44 participants carried out the evaluation task, and another group of 48 participants carried out the construction task. The participants had received no training in logic and had not been previously tested in any experiment on reasoning.

Design, procedure, and materials

These were identical to those of Experiment 1 except that the premises were re-ordered so that the conjunctive and disjunctive premises occurred last, for example:

If Ana is in Granada then Teresa is in Toledo.
If Pablo is in Barcelona then Teresa is in Toledo.
Ana is in Granada and Pablo is in Barcelona.

As in Experiment 1, there were two groups of participants: One group evaluated given conclusions for conjunctive and disjunctive inferences, and the other group constructed their own conclusions from the premises of the same set of problems. Each participant carried out one version of the four types of problem summarized in Table 2 (i.e., two conjunctive and two disjunctive inferences) and a further four problems that were invalid "filler" inferences.

Results

Table 4 presents the percentages of correct conclusions drawn in the experiment. The results showed that the conjunctive problems (97% correct) were reliably easier than the disjunctive problems (78% correct; Wilcoxon test, $n = 92$; $z = 3.29$; $p < .0005$, one-tailed). Moreover, as Table 4 reveals, the results were similar for both groups. Thus, as predicted, the conjunctive inferences (100% correct) were reliably easier than the disjunctive inferences (78% correct) in the evaluation task (Wilcoxon test, $n = 44$, $z = 2.8$, $p < .005$, one-tailed), and, as in Experiment 1, the conjunctive inferences (95% correct) were reliably easier than the disjunctive inferences (79% correct) in the construction task (Wilcoxon test, $n = 48$, $z = 1.96$, $p < .03$, one-tailed). In each group, the differences were reliable for both affirmative and negative problems (Wilcoxon tests with p values ranging from $p < .05$ to $p < .005$). The overall error rate in the construction condition was 13%, and all the errors were "no valid conclusion" responses.

Because both experiments tested participants from the same population, it is instructive to compare their results, particularly in the evaluation task. The overall pattern is very similar, except that the difference between evaluating the conjunctive and disjunctive problems was greater in Experiment 2 than in Experiment 1 (Mann–Whitney U test, $n_1 = 53$, $n_2 = 44$, $z = 1.67$, $p < .05$, one-tailed). Thus, the change in the order of the premises had the predicted effect of yielding a difference between the two types of problem even when the participants had to evaluate given conclusions. The participants may have adopted a different strategy or their working memory may have been preoccupied by the conditionals when they interpreted the disjunctive premise.

TABLE 4
The percentages of correct responses in Experiment 2

		% Correct			% Correct	
	Inferences	*Evaluate*	*Construct*	*Inferences*	*Evaluate*	*Construct*
One-model inferences (conjunctions)	if p then r if q then r p and q ∴ r	100	96	if not-p then r if not-q then r not-p and not-q ∴ r	100	94
Multiple-model inferences (disjunctions)	if p then r if q then r p or q ∴ r	77	81	if not-p then r if not-q then r not-p or not-q ∴ r	80	77

Note: The column labelled "Evaluate" gives the results for the task in which the participants evaluated given conclusions, and the column labelled "Construct" gives the results for the task in which the participants constructed their own conclusions.

EXPERIMENT 3

The aim of this experiment was twofold. First, we introduced a new computer-controlled procedure that enabled us to record the participants' times to read each premise and to evaluate the conclusions. Second, given the results of Experiment 2, we manipulated the order of the premises. One order was the same as that in Rips' original experiment, with the conjunction or disjunction preceding the two conditional premises, and the other was the "inverse" order in which the two conditional premises preceded the conjunction or disjunction. The task in both cases was to evaluate given conclusions. Given our earlier argument, we can predict an interaction between the two variables: The difference between the conjunctive and disjunctive problems should be greater for problems in the inverse order than for problems in Rips' original order.

Method

Participants

A total of 42 students participated in the experiment for course credits. They were assigned at random to one of the two groups (with 21 participants in each group). The participants had received no training in logic and had not been previously tested in any experiment on reasoning.

Design and materials

There were two independent groups to which the participants were assigned at random. Both groups carried out evaluation tasks of the conjunctive and disjunctive inferences (shown in Table 2), but one group received the premises in the same order as that in Rips' experiment, and the other group received the premises in the "inverse" order with the two conditionals preceding the conjunction or disjunction. In addition to the two conjunctive and disjunctive inferences, there were four invalid filler inferences. The problems were presented in a different random order to each participant. The content of the problems was the same as that used in Experiments 1 and 2, and each participant was assigned at random to one of the eight versions of the task.

Procedure

The participants were tested individually in a quiet room using the MacLaboratory program running on a Macintosh computer. The instructions and practice problems were the same as those in the previous experiments. The participants were told that they had to decide whether or not the conclusions of the inferences were "necessarily true"—that is, that they *must* be true given that the premises were true. They were also told that they could take as much time as they needed on a problem, but that they would not be able to backtrack through the display. Each trial began with a screen on which a message asked the participant to press the "space" key. As soon as the key was pressed, the first premise appeared in the middle of the screen. It remained until the participant pressed the "space" key again, which led to the second premise, and so on, until the last screen of the trial presented the putative conclusion. The participant responded to the conclusion by pressing either a key labelled "yes" (the conclusion was valid) or a key labelled "no" (the conclusion was not valid). The program recorded the reading times for the three premises and the latency of the response.

Results and discussion

The experiment used three main dependent measures: accuracy, reading times, and latencies. Accuracy turned out to be relatively insensitive, but reading times and latencies corroborated the model theory's key predictions.

Table 5 presents the percentages of correct conclusions drawn in the experiment. The difference in accuracy between the conjunctive and disjunctive problems was in the direction predicted by the model theory for both orders of the premises, but none of the differences was statistically significant (Wilcoxon tests, z ranged from .92 to 1.08). Likewise, the apparent increase in the difference with the inverse order of premises was not significant (Wilcoxon test, $n = 21$, $z = .92$). We surmise that the presentation of the premises one by one, the use of an evaluation task, and a necessarily smaller number of participants, yielded a task that was not sensitive enough to yield reliable differences in accuracy.

Table 6 presents the reading times for the premises and the latencies of the responses. The latencies of the correct responses corroborated the model theory's predictions. The participants were significantly faster to respond to the conjunctive inferences (mean of 2.96 s) than to the disjunctive inferences (mean of 4.87 s; Wilcoxon test, $n = 42$; $z = 2.12$; $p < .02$, one-tailed). The interaction between type of problem and the order of premises for response times was not significant (Mann–Whitney U test, $n_1 = 18$, $n_2 = 20$, $z = -1.26$, $p = .10$, one-tailed). The two main comparisons, however, were as predicted: With the inverse order of premises, responses to the conjunctive inferences (3.04 s) were reliably faster than those to disjunctive ones (6.02 s; Wilcoxon test, $n = 21$; $z = 2.17$; $p < .02$, one-tailed), whereas with Rips' order of premises, there was no reliable difference between conjunctive inferences (2.87 s) and disjunctive inferences (3.78 s).

The reading times of the premises also revealed some significant differences. The participants read the conjunctive premises faster than disjunctions when these premises occurred first (Wilcoxon test, $n = 21$; $z = 2.97$; $p < .002$, one-tailed). Likewise, it took less time to read the first conditional premise following the conjunction than that following the disjunction, but the difference misses significance by a small amount (Wilcoxon test, $n = 21$; $z = 1.47$; $p = .07$, one-tailed). These results are compatible with the need to construct one model for

TABLE 5
The percentages of correct responses in Experiment 3 in which the participants evaluated given conclusions

	% Correct					
	Rips' order of premises			*Inverse order of premises*		
	Affirmative	*Negative*	*Overall*	*Affirmative*	*Negative*	*Overall*
One-model inferences (conjunctions)	76	86	81	90	71	81
Multiple-model inferences (disjunctions)	76	71	74	86	57	71

Note: In Rips' order of premises, conditionals were presented at the end, whereas in the inverse order they were presented at the beginning.

TABLE 6
The mean reading times of the premises and response latencies[a] in Experiment 3

		Rips' order of premises (conjunction or disjunction first)	*Inverse order of premises (conjunction or disjunction last)*
One-model inferences (conjunctions)	Conjunction	9.52	8.64
	First conditional	6.73	6.39
	Second conditional	6.27	10.80
	Response time	2.87	3.04
Overall		22.51	25.72
Multiple-model inferences (disjunctions)	Disjunction	11.62	8.38
	First conditional	7.76	8.80
	Second conditional	7.23	11.78
	Response time	3.78	6.02
Overall		26.61	28.97

[a]In s.

conjunctions but multiple models for disjunctions and the respective difficulty of incorporating new information into one model or multiple models. In the inverse order when the two conditionals came first, the second conditional took reliably longer to read than the first conditional before both a conjunction and a disjunction (Wilcoxon test, $n = 21$, $z = 3.15$, p .001; and $n = 21$, $z = 2.31$, $p < .01$, one-tailed, respectively). This result is compatible with the need to try to integrate two conditionals when they occur first. Finally, in Rips' order, the conjunctive inferences had a faster overall processing time than the disjunctive inferences (Wilcoxon test, $n = 21$; $z = 2.07$; $p < .02$, one-tailed). Thus, reasoners take more time to read the disjunction and to connect it with the first conditional than to read the conjunction and to connect it with the first conditional. The model theory, as we show in the General Discussion, makes good sense of these results.

EXPERIMENT 4

A common measure in studies of reasoning inspired by formal rule theories is the participants' ratings of the difficulty of inferences (see, e.g., Braine et al., 1984; Yang et al., 1998). If multiple-model inferences are more difficult than one-model inferences, then naïve individuals should be aware of the difference—that is, they should rate the disjunctive inferences as harder than the conjunctive inferences. Likewise, they should have some insight into the cause of the difficulty. The aim of Experiment 4 was to test both these predictions.

Method

Participants

A total of 40 students participated voluntarily in the experiment for course credit. They had no training in logic and had not previously participated in any experiment on reasoning. One of the participants failed to give any reasons justifying his ratings.

Design and materials

The participants acted as their own controls and were given the four inferences shown in Table 2: two conjunctive inferences and two disjunctive inferences (one of each was affirmative, and one of each was negative). The premises were in Rips' order and were followed by a given conclusion. The participants were not instructed to evaluate the conclusion, but rather to rate the difficulty of each inference on a 5-point scale and to write down the reasons for their evaluations. The inferences were presented in two different random orders, and the participants were assigned at random to one order or the other. The contents of the inferences were the same as those in the previous experiments—that is, they were about people in various cities.

Procedure

The participants were tested individually in a quiet room. The instructions and the problems were assembled into booklets. The instructions on the first page of the booklet were the same as those for the evaluation task in Experiment 1. The four inferences were on the same page, and after each of them there was the following rating scale, shown here with its English translation:

1	2	3	4	5
Muy fácil	Bastante fácil	Fácil	Algo fácil	Nada fácil
(Very easy)	(Rather easy)	(Easy)	(Some-what easy)	(No so easy)

Justify your answer ____ _____
____ ____

As in Experiment 1, there was no time limit on completing the task.

Results and discussion

There were no differences in the ratings of the two groups who carried out the problems in different orders, and so we pooled the results. Table 7 presents the resulting mean ratings of the difficulty of the inferences. The participants rated the disjunctive inferences as more difficult than the conjunctive inferences, and the difference was reliable for both the affirmative and the negative problems (Wilcoxon tests, $n = 40$, $z = 3.06$ and 3.22, $p < .002$ and $p < .001$, respectively). They also rated the negative inferences (2.4) as harder than the affirmative inferences (2.7; Wilcoxon tests: $n = 40$, $z = 3.0$, $p < .003$). The results corroborated the model theory and showed that naive reasoners were aware that multiple-model problems are more difficult than one-model problems, just as they were aware that inferences with negative premises are more difficult than those with affirmative premises—a well-known psycholinguistic phenomenon (see, e.g., Clark & Clark, 1977).

We classified the participants' justifications for their ratings into four main categories:

1. *Connective*: A participant referred to the connective ("and" or "or") in the first premise, for example "Because 'or' throws you out".
2. *Polarity*: A participant referred to the fact that a premise was negative or affirmative, for example "The negative statement gives rise to doubts".
3. *Content*: A participant restated a premise, paraphrased it, or referred to its content.
4. *Miscellaneous*: Any other justification, for example "It's obvious, isn't it?"

TABLE 7
The means of the rated difficulty of the inferences and the percentages of the different categories of justification in Experiment 4

		Rated difficulty[a]	Category of justification			
			Connective	Polarity	Content	Miscellaneous
One-model inferences	Affirmative	2.05	6	22	39	33
(conjunctions)	Negative	2.40	9	24	36	31
Overall		2.25	8	23	37	32
Multiple-model inferences	Affirmative	2.70	51	13	10	26
(disjunctions)	Negative	3.05	42	17	18	23
Overall		2.90	46	15	14	25

[a] 1 = 'very easy' and 5 = 'not so easy'.

Table 7 also presents the percentages of the different justifications for the four categories of inference. Consistent with the model theory, the participants were much more likely to refer to the connective ("or") in a disjunctive premise than to the connective ("and") in a conjunctive premise (Wilcoxon test, $n = 39$, $z = 3.66$, $p < .002$, two-tailed). Conversely, the justifications for the ratings of one-model problems tended to refer either to the specific content of the premises or to the "obviousness" of the inference (the miscellaneous category). In other words, the one-model inferences were rather easy, and their ratings called for little explicit justification, whereas the multiple-model inferences were harder, and the participants typically located the difficulty in the premise containing "or". We examined whether the participants were more likely to refer to the polarity of the premises—whether they were affirmative or negative—in the case of the problems based on the negative premises, but the trend in the data was not reliable (Wilcoxon test, $n = 39$, $z = .44$, $p > .66$, two-tailed).

GENERAL DISCUSSION

Naive individuals can draw deductively valid conclusions, even when they cannot rely on their general knowledge to help them. A long controversy has concerned the nature of the underlying mental processes. Neither formal rule theories nor the mental model theory are specified down to the last detail, and Evans and Over (1997) have argued that it is difficult to carry out crucial experiments to decide between them. We are sympathetic to this point of view. Certainly, there are many lacunae in the formulation of the model theory, although it has been specified sufficiently to be modelled in a computer program, and moreover the program predicted some surprising phenomena (see, e.g., Johnson-Laird & Savary, 1996). Likewise, Rips (1994) has modelled his formal rule theory in a computer program. Both types of theory offer potential explanations of a major component of human deductive competence—the ability to make inferences that depend on sentential connectives. The three main connectives are "and", "or", and "if", and the two theories make sharply contrasting predictions about certain conjunctive and disjunctive inferences. These inferences are therefore crucial in deciding between the two approaches.

In Rips' original study, the two types of inference—conjunctive and disjunctive—did not differ reliably in difficulty when the participants evaluated given conclusions. Our Experiment 1 corroborated this result. However, when the participants generated their own conclusions, the conjunctive inferences (95% correct) were significantly easier than the disjunctive inferences (72% correct). One possibility is that the evaluation task is not sensitive enough to detect the difference in difficulty, perhaps because it allowed the participants to adopt different strategies.

Reasoners do adopt different strategies. These strategies can depend on the task (see, e.g., Byrne & Handley, 1997). But, even with the same task, as Johnson-Laird, Savary, and Bucciarelli (2000) have observed, naïve reasoners develop a variety of strategies in sentential inferences from multiple premises. Some keep track of all the possibilities consistent with the premises, some combine pairs of premises to draw intermediate conclusions, and some make immediate inferences, where necessary, to convert the premises into a chain of conditionals. These strategies concern the overall plan for an inference, whereas inferential tactics concern the way in which individual steps in a strategy are carried out. Strategies are often consciously accessible and revealed in reasoners' "think aloud" protocols. Tactics, in contrast, appear to be seldom, if ever, accessible to consciousness. Yet the various tactical steps required in the various strategies can all be explained in terms of mental models.

This account may explain some of our results. The task of evaluating given conclusions is easier than the task of constructing one's own conclusions. The difference is likely to reflect the use of different strategies. Similarly, in Experiment 3, individual disjunctive premises took longer to read than individual conjunctive premises. The difference may reflect the tactical distinction between envisaging three models and envisaging one model. These tactical differences, however, did not yield any overall effect on accuracy, because the strategies for evaluating given conclusions are relatively simple to use. Performance was indeed at ceiling. The task of constructing ones own conclusion is more demanding. The relevant strategies call for both constructing models and using them to formulate conclusions. Performance in this case shows differences in accuracy between conjunctive and disjunctive problems.

The role of strategies was also corroborated by Experiment 2 in which the conjunctive and disjunctive premises followed the two conditional premises. In this case, the conjunctive inferences were reliably easier than the disjunctive inferences in both the evaluation task and the generation task. When the reasoners interpret the disjunction, their working memory is already pre-occupied by the models of the conditionals. Thus, in this more demanding condition, the difficulty of manipulating multiple models is again evident.

In Experiment 3, the premises were presented in two different orders and one at a time under the participants' control. Accuracy with the two types of inference did not differ, probably because of the change in procedure. When reasoners control how long they examine each premise, there is likely to be a trade-off between reading times and accuracy. In particular, they can take longer to read disjunctive premises in order to cope with their greater difficulty. Indeed, the reading times and the response latencies showed that the conjunctive problems were easier than the disjunctive problems. When the conjunctive premise (e.g., Ana is in Granada and Pablo is in Barcelona) occurred first, the reasoners presumably constructed one model:

Ana   Pablo

and when the disjunctive premise (e.g., Ana is in Granada or Pablo is in Barcelona) occurred first, they presumably constructed multiple models:

Ana	
	Pablo
Ana	Pablo

The reading times showed that it took them longer to construct the multiple models for the disjunction, and that they tended to slow down in processing subsequent premises. They were indeed faster overall in processing the conjunctive problems than the disjunctive problems. In contrast, when the two conditionals occurred first, it took the participants longer to interpret the second conditional because it elicits multiple models. Thus, given the initial premises

If Ana is in Granada then Teresa is in Toledo.
If Pablo is in Barcelona then Teresa is in Toledo.

reasoners had to construct the multiple models:

Ana		Teresa
	Pablo	Teresa
	. . .	

Afterwards, the difference between a conjunction or a disjunction was not in the time to read the premise, but in the time to determine whether the conclusion follows from the premises. The conjunction maps onto the first model to yield the conclusion. (It also maps onto the second or third model to yield it.) But the disjunction calls for both explicit models to be checked to ensure that they support the conclusion. In sum, there was a trade-off between time and accuracy. Reasoning is accordingly highly sensitive to minimal changes in the format and presentation of a task (García-Madruga, Moreno, Carriedo, & Gutiérrez, 2000). Once again, these changes bring out the flexible nature of inferential strategies, which can adapt to the exigencies of the task.

In Experiment 4, the participants rated the difficulty of conjunctive and disjunctive inferences, and gave written justifications for their ratings. They rated the conjunctive inferences as easier than the disjunctive ones. Their written justifications showed that they tended to refer specifically to the difficulty of "or" in accounting for the disjunctive inferences, whereas they were often at a loss to explain why the conjunctive inferences were easier.

Could the phenomena be explained by formal rule theories? As we showed in the Introduction, these theories postulate that reasoners apply the same formal rules to the conjunctive and disjunctive inferences whether they have to evaluate given conclusions or to construct their own conclusions. Likewise, Rips (1990) has argued that human reasoning is based on a single deterministic strategy. It is therefore not easy to see how formal rule theories could in principle account for our results. On the one hand, they have no strategic component to explain the difference between evaluating and constructing conclusions for the present inferences. On the other hand, the length of derivations predicts that the disjunctive inferences should be easier than the conjunctive inferences—a difference that never occurred in any of our experiments.

An advocate of formal rules might argue post hoc that the dilemma rule is much harder to apply than the rules for the conjunctive inference (see also Rips, 1994). Likewise, rule theorists might propose that simple rules for conjunctions, disjunctions, and conditionals are more basic than the rule for dilemmas. From a developmental perspective, rule theorists could argue that the dilemma rule is constructed from the components of disjunctive and conditional rules. However, the introduction of a distinction between "basic" rules and "composite" rules has drawbacks. At first sight, for instance, it contravenes the difficulty weights for the relevant rules: 0.41, 0.47, and 0.16, respectively, for detaching a conjunction, modus ponens and dilemma (Braine & O'Brien, 1998b, p. 80). However, even this way to salvage formal rules does not appear to account for all the phenomena. It fails to explain why the difference between the two types of inference is enhanced when reasoners have to draw their own conclusions and when the conjunctive and disjunctive premises are presented after the conditional premises. Even if rule theories were modified to allow a variety of strategies to be applied to these inferences—a quite substantial modification, as we have seen—it still is not clear how the theories could explain the effects of the order of premises in either the present studies or those carried out by Girotto et al. (1997). In contrast, the model theory motivated the choice of experimental manipulations and predicted their results.

Sentential reasoning is at the heart of human deductive competence, and the contrast between conjunctive and disjunctive inferences is a crux in the difference between formal rule theories and the mental model theory. For formal rules, the conjunctive inferences should be either harder or no different from the disjunctive inferences. For mental models, conjunctions call for only one model, whereas disjunctions call for multiple models. Hence, the conjunctive inferences should be easier than the disjunctive inferences, unless reasoners are able to use a compensating strategy. Our results corroborated the model theory and revealed the importance of inferential strategies.

REFERENCES

Bauer, M.I., & Johnson-Laird, P.N. (1993). How diagrams can improve reasoning. *Psychological Science, 4,* 372–378.

Braine, M.D.S. (1990). The "natural logic" approach to reasoning. In W.F. Overton (Ed.), *Reasoning, necessity and logic: Developmental perspectives*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Braine, M.D.S., & O'Brien, D. (1998a). *Mental logic*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Braine, M.D.S., & O'Brien, D. (1998b). The theory of mental-propositional logic: Description and illustration. In M.D.S. Braine & D. O'Brien (Eds.), *Mental logic*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Braine, M.D.S., O'Brien, D., Noveck, I.A., Samuels, M.C., Lea, R.B., Fisch, S.M., & Yang, Y. (1998). Further evidence for the theory: Predicting intermediate and multiple conclusions in propositional logic inference. In M.D.S. Braine & D. O'Brien (Eds.), *Mental logic*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Braine, M.D.S., Reiser, B.J., & Rumain, B. (1984). Some empirical justification for a theory of natural propositional reasoning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 1). New York: Academic Press.

Braine, M.D.S., Reiser, B.J., & Rumain, B. (1998). Evidence for the theory: Predicting the difficulty of propositional logic inference problems. In M.D.S. Braine & D. O'Brien (Eds.), *Mental logic*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Byrne, R.M.J., & Handley, S.J. (1997). Reasoning strategies for suppositional deductions. *Cognition, 62,* 1–49.

Clark, H.H., & Clark, E.V. (1977). *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.

Evans, St.B.T., Newstead S.E., & Byrne R.M.J. (1993). *Human reasoning*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J.St.B.T., & Over, D. (1997). Rationality in reasoning: The problem of deductive competence. *Cahiers de Psychologie Cognitive / Current Psychology of Cognition, 16,* 3–38.

García-Madruga, J.A., Moreno, S., Carriedo, N., & Gutiérrez, F. (2000). Time measures in Rips' problems. In J.A. García-Madruga, N. Carriedo, & M.J. González Labra (Eds.), *Mental models in reasoning* (pp. 213–226). Madrid: UNED.

Girotto, V., Mazzoco, A., & Tasso, A. (1997). The effect of premise order in conditional reasoning: A test of the mental model theory. *Cognition, 63,* 1–28.

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

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

Johnson-Laird, P.N., Legrenzi, P., Girotto, V., Legrenzi, M.S., & Caverni, J-P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review, 106,* 62–88.

Johnson-Laird, P.N., & Savary, F. (1996). Illusory inferences about probabilities. *Acta Psychologica, 93,* 69–90.

Johnson-Laird, P.N., Savary, F., & Bucciarelli, M. (2000). Strategies and tactics in reasoning. In W.S. Schaeken, G. de Vooght, A. Vandierendonck, & G. d'Ydewalle (Eds.), *Deductive reasoning and strategies* (pp. 209–240). Makwah, NJ: Lawrence Erlbaum Associates, Inc.

Polk, T.A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review, 102,* 533–566.

Rips, L.J. (1990). Paralogical reasoning: Evans, Johnson-Laird, and Byrne on liar and truth-teller puzzles. *Cognition, 36,* 291–314.

Rips, L.J. (1994). *The psychology of proof: Deductive reasoning in human reasoning.* Cambridge, MA: MIT Press.

Yang, Y., Braine, M.D.S., & O'Brien, D.P. (1998). Some empirical justifications of the mental predicate logic model. In M.D.S. Braine & D.P. O'Brien (Eds.), *Mental logic*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Original manuscript received 23 April 1999
Accepted revision received 10 January 2000