

On the generation and evaluation of inferences from single premises

THOMAS C. ORMEROD

Lancaster University, Lancaster, England

and

JULIET RICHARDSON

Cardiff University, Cardiff, Wales

A theory of how individuals construct mental models to draw inferences from single premises was tested in three experiments. Experiment 1 confirmed a counterintuitive prediction that it is easier to generate inferences between conditionals and disjunctions than it is to evaluate them. Experiment 2 replicated this finding, but an advantage found in the first experiment for conditional-to-disjunction over disjunction-to-conditional inferences was removed with different sentence contents. Experiment 3 showed that disjunction-to-conditional inferences were facilitated when premises expressed familiar indicative relations, whereas conditional-to-disjunction inferences were facilitated when premises expressed causal relations. The results indicate that small changes in task format can have large effects on the strategies that people use to represent and reason about different sentential connectives. We discuss the potential for theories other than mental models to account for these results. We argue that, despite the important role played by single-premise inferences in paraphrasing logical forms during inference, mental logic theories cannot account for the results reported here.

Inferences drawn from single premises are a facet of everyday thought and discourse. When a friend says, "Either you go to the party or I'm not going," one might examine the consequences of her statement by rephrasing it as "If I go to the party then she will go" or, perhaps, "If I don't go to the party then she won't go." These kinds of inference have been shown to be important in natural language understanding (e.g., Gleitman & Gleitman, 1970) and semantic memory (e.g., Baguley & Payne, 1999; Bransford, Barclay, & Franks, 1972). In addition, working out what moves follow legally from the rules of a problem, another example of a single-premise inference, can be a major source of problem difficulty (e.g., Kotovsky, Hayes, & Simon, 1985). Deductive reasoning research has typically focused on inferences whose conclusions express relations between two or more premises, and inferences from single premises have been largely overlooked. Yet single-premise inferences serve important functions in deductive reasoning, providing mental or overt rehearsal of premises and acting as a mechanism for simplifying complex premises (e.g., through the re-

moval of negations). Despite the absence of a requirement to integrate information across multiple premises in the derivation of conclusions, single-premise inferences are frequently nontrivial. For example, Newstead and Griggs (1983) found that participants often endorsed as valid the inference *Some A are B, therefore some A are not B*. Philosophical accounts of the logic of quantifiers tend to treat this inference as invalid, since the premise allows the possibility that *All A are B* (e.g., Tomassi, 1999). It seems likely that the mechanisms individuals use to make single-premise inferences such as this are not driven solely by judgments of logical validity, but these mechanisms are currently unspecified.

This paper is concerned with single-premise inferences drawn between disjunctions and conditionals. For example, given a conditional premise, *If it is a mammal, then it is an animal*, the disjunction *Either it is not a mammal or it is an animal* is a valid inference, whereas *Either it is a mammal or it is an animal* is invalid. We develop and test an account of inferences drawn between conditionals and disjunctions from the perspective of Johnson-Laird and Byrne's (1991, 2002) mental models theory. According to these authors, individuals construct mental models that represent possibilities compatible with the premises. If the models exhaust what is possible given the premises, individuals can use the models to draw conclusions. The connective in an assertion elicits a stored basic representation of its *core meaning*, in the form of a set of mental models that distinguishes its interpretation from that of other connectives. The basic representation of a connective may be modulated by se-

This research was supported mainly by a Study Abroad grant to the first author from the Leverhulme Trust. Experiment 3 was carried out while both authors were at the Department of Human Sciences, Loughborough University of Technology, as part of the second author's Ph.D. research program sponsored by British Gas plc. We gratefully acknowledge the valuable comments made on earlier drafts by Philip Johnson-Laird, Uri Hasson, Nick Chater, Thom Baguley, and an anonymous referee. Correspondence concerning this article should be addressed to T. C. Ormerod, Department of Psychology, Lancaster University, Fylde College, Lancaster LA1 4YF, England (e-mail: t.ormerod@lancaster.ac.uk).

semantic or pragmatic sentence content that elicits prior knowledge, adding models of possibilities not already made explicit in representing the connective. Alternatively, sentence content can block models describing real-world impossibilities that would otherwise be made available in representing the connective.

For the time being, we focus on possibilities made available by the connective alone. For example, the conditional *If the letter is A then the number is 2* makes explicit the possibility that the letter *A* occurs with the number 2. However, it also allows that other possibilities may exist (unlike the connective *and*, which allows only one possibility). The first mental model that an individual forms for this conditional, according to the theory, captures explicitly the structure of the relation between *A* and 2. The second mental model consists of a *mental footnote*, shown as an ellipsis and referred to as an implicit model, to the effect that other possibilities may exist in which the antecedent is false. Thus, in the absence of other semantic or pragmatic information, conditionals are initially represented by two mental models, each shown on a separate line, as in the following example:

$$\begin{array}{l} A \quad 2 \\ \dots \end{array}$$

Note that Johnson-Laird and Byrne's (1991, 2002) view, and the view we adopt in this paper, of the interpretation of conditionals differs from that of some authors. It has been suggested that the connective *if* implies a relation between the antecedent and the consequent that goes beyond co-occurrence, in which the relevance of the consequent is implied by the antecedent (e.g., Barwise, 1986). Johnson-Laird and Byrne (2002) provided a number of counterexamples to the claim that the consequents of conditionals can be inferred from their antecedents. For example, for the deontic example *If a person is drinking beer, then the person is 19 years of age*, given that Fred is drinking beer, it does not follow that Fred is definitely 19 years of age, but only that he ought to be 19 years of age. The precise relation between the antecedent and the consequent in this case comes, they argue, not from the meaning of the connective, *if*, but from semantic and pragmatic modulation.

The basic representation of a disjunction consists, according to the theory, of three mental models. Two models represent explicitly the possibilities given by the first and second clauses of the disjunction, whereas the third model represents implicitly a mental footnote that other possibilities may exist. For example, the disjunction *Either the letter is A or the number is 2* elicits the following initial mental models:

$$\begin{array}{l} A \\ 2 \\ \dots \end{array}$$

One of the central principles of mental models theory is that model construction takes place within a working

memory system of limited capacity. Therefore, reasoning performance is an inverse function of the number of models that must be constructed and fleshed out during reasoning. Reasoning with disjunctions has been shown to be harder than reasoning with conditionals (e.g., Johnson-Laird, Byrne, & Schaeken, 1992), and this greater difficulty is attributed by Johnson-Laird and Byrne (1991) to the larger number of mental models in the basic representation of a disjunction. Ormerod (2000) has extended the notion of a limited capacity for model construction into a broader *principle of minimal completion*, in which individuals endeavor to flesh out or add as few models as possible under the control of their current inferential goal. A consequence of minimal completion is that individuals seek to flesh out partially represented models before they flesh out implicit models, the assumption being that to complete a partially described possibility entails a lower cognitive load than does the discovery of a new possibility. Minimal completion appears to have been absorbed within the current version of mental models theory (e.g., Johnson-Laird, 1999; Johnson-Laird & Byrne, 2002).

A second cornerstone of mental models theory, the *principle of truth*, states that individuals initially represent a clause given in an assertion only when it is true in the corresponding possibility. Because individuals represent only what is true, the possibilities represented by the implicit models of conditionals and disjunctions are initially unknown. Moreover, the principle of truth limits the initial representation of the two explicit models of the disjunction to what is known to be true in each possibility. Additional processes have to occur in order to recover information that is not represented in initial models. In particular, an initial mental model can be fleshed out into a *fully explicit* model, provided that individuals make a mental footnote about what is false in the model or assume that, where there is no information about a clause in a model, its truth value is opposite to the truth values of the same clause elsewhere in the set of models (see Johnson-Laird & Savary, 1999, for the description of a computer program that implements this mechanism).

The mechanism for fleshing out the initial models of connectives determines the order in which additional possibilities are likely to be envisaged by individuals. For example, the next mental model to be fleshed out from the initial models of the conditional *If A then 2* describes the possibility in which the assertion's clauses have the opposite value to those explicitly represented in the first model, giving rise to the following partially fleshed-out models:

$$\begin{array}{l} A \quad 2 \\ \neg A \quad \neg 2 \\ \dots \end{array}$$

where \neg denotes negation. In the case of a disjunction, minimal completion leads to the prediction that individuals will first flesh out the partially explicit models, by

adding to the explicit clause in each partial model the other sentence clause with an opposite truth value to its representation in other models. Thus, individuals are likely to flesh out initial models of the disjunction *Either A or 2* as follows:

$$\begin{array}{l} A \quad \neg 2 \\ \neg A \quad 2 \\ \dots \end{array}$$

Further fleshing out of the implicit model of the disjunction should then make explicit the possibility of both *A* and *2*. Likewise, further fleshing out of the implicit model of the conditional *If A then 2* should make explicit the possibility of *2* in the absence of *A*. If models of these possibilities are added to the basic representations of each connective, individuals should interpret the disjunction as inclusive and the conditional as a material implication. However, the principle of minimal completion dictates that the fleshing out of initial models ceases when the models allow a conclusion to be drawn that meets the inferential goal. Thus, if an inferential goal can be met without making explicit the third model of each connective, the logical interpretation of each connective remains ambiguous.

In addition to proposing a process for the construction of models, Johnson-Laird and Byrne (1991) proposed a process for the drawing of conclusions from models. The basis of their approach is to seek a maximally parsimonious description of the models: Connectives are sought to connect clauses in such a way that each clause is mentioned once only in the resulting conclusion. Where there are more than two clauses, models must be simplified by *score tables*. The details of the latter process need not concern us here (for further information, see Johnson-Laird & Byrne, 1991, p. 185), since the single-premise inferences studied in this paper involve only two clauses. To derive conclusions with models that represent only two clauses, the process cycles through stored representations of connectives, matching a template of models in the basic representation of each connective against models of the premises. When a match is found, a conclusion is generated that joins the clauses, using the successful connective.

In a final, *validation* phase of reasoning, individuals search for alternative models, or counterexamples, in which their putative conclusion is false. If none is found, their conclusion is valid; otherwise, they must attempt to draw a further conclusion that is consistent with all models of the premises. Ormerod (2000) argued that a corollary of the principle of minimal completion is that individuals engage in a search for counterexamples only if the inferential goal makes the exhaustive testing of all possibilities in which a clause might appear an explicit requirement (e.g., in tasks, such as Wason's, 1966, selection task, that require participants to test the truth or falsity of an assertion). One situation, however, in which counterexamples are likely to be available without requiring explicit search is when a putative conclusion is incon-

sistent with an individual's prior knowledge. Counterexamples cued from mismatches between prior knowledge and putative conclusions make available new models, and individuals must then seek an alternative conclusion that is true in all the models (Oakhill & Johnson-Laird, 1985).

The experiments reported in this paper employed a paraphrase task to explore single-premise inferences between conditionals and disjunctions. In this task, the inferential goal is to draw a conclusion from a single premise, using a connective nominated by the experimenter. For example, given the disjunction *Either the letter is A or the number is 2* and the requirement to paraphrase it in the form "If . . . then . . .," an individual might draw the valid conclusion *If the letter is not A then the number is 2*. If the interpretation of the disjunction is exclusive, another valid conclusion is *If the letter is A then the number is not 2*. Valid conclusions are also possible in which the order of the terms is swapped (e.g., *If the number is 2 then the letter is not A*). However, the conclusions *If the letter is A then the number is 2* and *If the letter is not A then the number is not 2*, or equivalent assertions in which the clause order is reversed, would be invalid under either interpretation of the disjunction.

How might mental models theory account for drawing a conditional inference from a disjunction? First, individuals construct initial models of the disjunction. They then have to flesh out the initial models, in the way described above. The template of the conditional connective can then be mapped onto either of the two fully explicit models that result from the fleshing out, and a conditional conclusion can be derived. Either of the two explicit models of a partially fleshed-out representation of a disjunction is sufficient to draw a putative conclusion, and the principle of minimal completion holds that the implicit model will not be fleshed out if the individual's inferential goal can be met without doing so. The absence of familiar thematic content in assertions makes it unlikely that individuals untutored in logic will identify counterexamples that might challenge the validity of either putative conclusion. Thus, in the absence of any other bias, individuals should be equally likely to generate or endorse the inferences *Either the letter is A or the number is 2, therefore if the letter is not A then the number is 2* and *Either the letter is A or the number is 2, therefore if the letter is A then the number is not 2*. Inferences from conditionals to disjunctions are generated by a similar process, except that they require the fleshing out of the implicit model in the initial models of the conditional in order to identify a new possibility that has a different truth value from the explicit model in the initial models of the conditional.

There are, according to the account of the model described here, a number of opportunities for error in generating a single-premise inference. For example, fleshing out an initial model is a complex process, requiring the identification of a missing clause in the initial model, the detection of the truth value of the missing clause in other

models, the addition of the missing clause with an inverted truth value to make the model fully explicit, and a check that the fully explicit model does not contradict any other models in the representation. Fleshing out an implicit model is similarly error prone. In effect, fleshing out initial or implicit models (in the absence of semantic or pragmatic information) is akin to creating hypotheses about new possibilities that must be checked for consistency with known possibilities. The more fleshing out that must occur to satisfy an inferential goal, the greater the likelihood of error will be.

The aim of this study was to test this model theory of drawing inferences from single premises. The first experiment compared the generation and the evaluation of paraphrases. Individuals, we argue, adopt different strategies for generation and evaluation of paraphrases because it is more cognitively economical to do so. Strategic differences in the ways that individuals approach each response format give rise to the surprising prediction that participants will be better at generating than at evaluating paraphrases. This prediction was tested in Experiment 1. Experiment 2 replicated this test with different materials and also tested a hypothesis that pragmatic modulation will increase the number of inferences drawn that are consistent only with inclusive and material implication interpretations of disjunctions and conditionals, respectively. Experiment 3 tested a prediction that familiar indicative content should facilitate only disjunction-to-conditional paraphrases, whereas causal content should facilitate only conditional-to-disjunction paraphrases. We will expand upon the rationale for these predictions below.

EXPERIMENT 1

In this experiment, participants generated and evaluated paraphrases between conditionals and disjunctions. Intuitively, one might expect evaluation to be easier than generation, in the same way that recognition generally outstrips recall in memory retrieval. An evaluation format may present responses that participants fail to discover for themselves (Dugan & Revlin, 1990). Evaluation might also entail a lower processing load, since participants have only to check whether a conclusion is valid, rather than searching for a possible conclusion and then testing its validity. For example, Hardman and Payne (1995) found that participants were better at evaluating than at generating conclusions to syllogisms.

The opposite result is predicted by a model theory of single-premise inferences. When participants generate a single-premise inference, they need only flesh out models of the presented premise, which can then be matched against a stored representation of the target connective. When participants evaluate a single-premise inference, we predict that they will adopt a *comparison* strategy. Using this strategy, the participants construct and flesh out models of the premise *and* of the presented conclusion. Once both sets of models are fleshed out, they can be compared. If they match, the conclusion can be ac-

cepted as valid. Otherwise the conclusion should be rejected as invalid. Both generation and evaluation require two sets of models to be matched. However, evaluation requires the *fleshing out* of two sets of models, rather than one, so an evaluation task should be more difficult than a generation task.

It is possible, on the other hand, that participants might adopt a *generate-and-test* strategy to evaluate single-premise inference. Using this strategy, participants would flesh out initial models of the premise, generate an inference from these models, and then check it against the presented conclusion. However, if the generated inference and the presented conclusion did not match, the participants would have to check whether another inference could be generated that matched the conclusion, before being certain that the presented conclusion was invalid. Thus, if participants adopt a generate-and-test strategy, response times should be longer on invalid inferences. In contrast, if participants use a comparison strategy, they should be no faster at evaluating valid inferences, since evaluations of both valid and invalid inferences would require the construction and comparison of two sets of models. Indeed, evaluating invalid inferences might even be faster, since *invalid* conclusions could be drawn as soon as a mismatch between any part of the model sets was detected, whereas valid conclusions would require the construction and comparison of complete model sets for each assertion.

Our theory of single-premise inferences predicts that, in the absence of semantic or pragmatic information, the logical interpretation of conditionals and disjunctions remains ambiguous in a paraphrase task. Thus, we scored the generation or endorsement of conclusions that are valid under either interpretation as correct responses. It is possible that some participants might adopt a stricter material conditional/inclusive interpretation of the presented premises. They would then reject as invalid some conclusions that we are scoring as valid, inflating the number of evaluation trials scored as errors. Thus, we also inspected whether there was a difference in performances on trials in which the paraphrase to be evaluated was valid only under a biconditional/exclusive interpretation and on trials in which the paraphrase was valid under both interpretations.

Method

Participants. Thirty-two undergraduate psychology students from Lancaster University were paid £3 each to take part. None had previously studied logic.

Design and Materials. The participants were assigned at random to one of two groups, one receiving conditional-to-disjunction inferences, the other receiving disjunction-to-conditional inferences. Each participant generated inferences and, in other conditions, evaluated valid or invalid inferences. In each of these conditions, the participants received eight trials, consisting of two examples each of four possible premise polarities to counterbalance the presence of negation across conditions. The materials consisted of 24 sentences describing arbitrary and unfamiliar indicative relations between clauses, counterbalanced across conditions. Exemplars were constructed from four types of unfamiliar relations, counterbalanced across conditions (the examples shown here are conditional-to-

disjunction conditions with affirmative premises). The relations concerned chemical properties (e.g., *If it is iodide then its viscosity is 4*), position and color (e.g., *If the light is blue then the switch is pointing left*), horses and riders (e.g., *If the rider is Evans then the horse is Silver*), and letters and numbers (e.g., *If the card is F then the weight is 3*). For evaluation trials, each disjunctive or conditional premise was paired with a disjunctive or conditional conclusion, respectively—one a valid paraphrase, the other an invalid paraphrase. For example, a conditional premise, *If it is a rectangle then it is red*, might be paired with a disjunction valid under any logical interpretation (*Either it is not a rectangle or it is red*), one valid only under an exclusive interpretation (*Either it is a rectangle or it is not red*), or one valid under no logical interpretation (*Either it is not a rectangle or it is not red or Either it is a rectangle or it is red*).

Procedure. The participants were instructed to generate paraphrases according to a given template (“... if then ...” or “Either ... or ...”) that had precisely the same meaning as presented premises or to judge the validity of paraphrases of premises on the same grounds. They were shown examples of a valid and an invalid paraphrase. After reading the instructions, each participant received 3 practice trials in which they generated and evaluated conditional paraphrases (one valid and one invalid) of universally quantified statements, receiving feedback after each practice trial. Each participant then received 24 experimental trials, presented in a randomized order, using a computer program written in Hypercard 2.3. For generation trials, the participants were presented with a sentence and were instructed to study it until they felt that they understood it and then to state their paraphrase verbally. The time from presentation of the premise to the start of the verbal statement of the conclusion was recorded, and then the participants typed their conclusion as confirmation of their verbal response. For each evaluation trial, as soon as the participants had read and understood the premise, they selected a screen button marked *Evaluate paraphrase* with the cursor. The participant pressed *yes* or *no* labels covering the F and L keys, respectively, on the computer keyboard to indicate whether they judged the paraphrase to be correct or incorrect. The time from presentation of the premise to evaluation of the conclusion was recorded, along with the response made. The participants worked through the trials at their own rate. The duration of the experiment was approximately 20 min.

Results and Discussion

No effects were found of the relations used to construct exemplars in each condition, so the data were collapsed across the four types in subsequent analyses. The mean numbers of correct responses and response times in each condition are shown in Table 1. The participants were reliably more accurate in generating paraphrases (64% correct) than they were in evaluating either valid paraphrases (48%) or invalid paraphrases [46%; $F(2,60) = 7.13, MS_e = 0.318, p = .0017$]. Single sample *t* tests (one-

tailed) were conducted on each of the tasks to see whether performance differed significantly from chance. Of these, only two were significant: conditional-to-disjunction generation [$t(15) = 4.06, p < .001$] and disjunction-to-conditional generation [$t(15) = 1.95, p < .05$]. Response times did not differ significantly [generation = 21.2 sec, valid paraphrase evaluation = 20.6 sec, and invalid paraphrase evaluation = 19.6 sec; $F(2,60) < 1$], suggesting that the participants used a comparison, rather than a generate-and-test strategy, to evaluate paraphrases. There was no evidence of a speed-accuracy tradeoff to account for the difference between generation and evaluation. Furthermore, the participants were not significantly more accurate on evaluation trials valid under both logical conditional interpretations (47%) than they were on trials valid under only a biconditional interpretation [45%; $F(1,30) < 1$]. Thus, the key prediction, that it is easier to generate than it is to evaluate single-premise inferences, was confirmed by the results of this experiment.

The participants were significantly more accurate in generating conditional-to-disjunction paraphrases (71%) than they were in generating disjunction-to-conditional paraphrases [56%; $F(1,30) = 5.33, MS_e = 0.264, p = .028$]. There were no significant differences between conditional-to-disjunction and disjunction-to-conditional paraphrases in evaluating either valid paraphrases (47% vs. 46%) or invalid paraphrases (47% vs. 50%). This result replicated a study by Richardson and Ormerod (1997, Experiment 1) that used the same materials as the present experiment. In that study, participants generated more correct disjunctive paraphrases of conditionals (73%) than they did conditional paraphrases of disjunctions (44%).

Richardson and Ormerod (1997) argued that conditional-to-disjunction inferences were easier to draw than disjunction-to-conditional inferences because of the lower cognitive load entailed in representing the single initial model of a conditional, as compared with two initial models of a disjunction. Although consistent with their data and those of the present experiment, their account contradicts the principle of minimal completion. If individuals flesh out partially represented models before they flesh out implicit models, and given that a conditional can be drawn from fleshing out just one of the initial models of a disjunction, then according to this principle, it ought, if anything, to be easier to draw conditional conclusions from disjunctions than vice versa. We have re-

Table 1
Mean Numbers of Correct Responses (Out of Eight per Participant) and Mean Response Times (RTs, in Seconds) in Each Condition of Experiment 1

Response Format	Conditional-to-Disjunction				Disjunction-to-Conditional			
	No. Correct		Mean RT		No. Correct		Mean RT	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Generation	5.7	1.7	19.7	11.9	4.5	1.2	22.7	11.7
Evaluation								
Invalid	3.8	1.8	19.0	10.7	3.9	1.5	20.2	15.2
Valid	3.9	1.8	20.6	14.1	3.5	1.3	20.7	13.3

cently conducted a study of paraphrase generation (Ormerod & Johnson-Laird, 2002), in which an advantage for conditional-to-disjunction paraphrases over disjunction-to-conditional paraphrases was absent and in some cases reversed, a result that is consistent with the principle of minimal completion.

How might these contradictory findings be resolved? The answer may lie in differences between the sentence contents used here and by Richardson and Ormerod (1997) and those used by Ormerod and Johnson-Laird (2002). In the present experiment, the sentences described arbitrary relations between clauses, such as *Either it is iodide or its viscosity is not 14*. In contrast, the participants in Ormerod and Johnson-Laird's studies generated or evaluated paraphrases of such sentences as *Either James is in Scotland or he is not in Glasgow*. The relations were nonarbitrary because the content described mutually exclusive geographical locations: An individual cannot be in two different places at the same time. It may be, then, that the source of difficulty in disjunction-to-conditional inferences in Experiment 1 and in Richardson and Ormerod's study was not the greater number of models in the initial representation of a disjunction than in a conditional. Instead, it may have occurred because it is more difficult to form an initial representation of a disjunction than of a conditional when the relation between clauses in the assertion is arbitrary. This contention is consistent with Fillenbaum's (1974) observation that individuals often spontaneously paraphrase disjunctions in which clauses describe unrelated events by replacing *or* with *while*.

EXPERIMENT 2

A thorough test of the explanation offered above for differences between the results of Experiment 1 and those of Ormerod and Johnson-Laird (2002) lies outside the scope of this paper. However, it is important to investigate whether similar changes in content might also affect the primary phenomenon under investigation—namely, the advantage for generation over evaluation. According to our model theory of single-premise inferences, the advantage should occur irrespective of sentence content, since the number of model sets required for each task should be unaffected by the content of assertions. Thus, Experiment 2 was conducted to test whether the advantage for generation over evaluation is also found with materials of the sort used by Ormerod and Johnson-Laird. Thematic content was included as a factor, comparing performance in both generation and evaluation response formats with assertions containing either arbitrary or nonarbitrary relations. If arbitrary relations impair the construction of initial models of a disjunction, we predict an advantage for conditional-to-disjunction inferences only when the sentence content is arbitrary.

The comparison of arbitrary and nonarbitrary relational contents also allowed a test of the validation phase of the model theory of single-premise inferences. For ex-

ample, given a premise containing arbitrary content, such as *Either the letter is A or the number is 2*, we predicted that individuals would be equally likely to generate either of the following conclusions: *If the letter is not A then the number is 2* or *If the letter is not A then the number is 2*. If a premise contains familiar thematic content, on the other hand, then individuals might generate a putative conclusion that is inconsistent with their prior knowledge. For example, given the disjunction *Either John is not in London or he is in Europe*, by fleshing out partially explicit models alone, individuals might generate either of the following conclusions: *If John is not in London then he is not in Europe* or *If John is in London then he is in Europe*. The first of these putative conclusions is likely to cue prior knowledge of counterexamples—that is, European destinations other than London that might be John's location. If this occurs, individuals should add a further model to the representation of the disjunction and seek a conclusion that is true in all models. Thus, we predicted that assertions containing nonarbitrary relations concerning location would elicit more conclusions on generation trials that were consistent with inclusive and material implication interpretations of disjunctions and conditionals, respectively.

Method

Participants. Twenty-six undergraduate and postgraduate students from Lancaster University were paid £3 each to take part. None had previously studied logic.

Design and Materials. Each participant generated and, in other conditions, evaluated valid and invalid conditional-to-disjunction and disjunction-to-conditional inferences. In each of these conditions, the participants received four trials, two each for premises with arbitrary and nonarbitrary relations. One of the two trials for each content contained affirmative clauses, whereas the other contained a negative second clause. Premises with nonarbitrary relations described geographical locations of a named individual, whereas the materials from Experiment 1 served as the premises with arbitrary relations. For evaluation trials, each sentence was paired with a valid or an invalid paraphrase. All the valid evaluation trials consisted of inferences that were valid under material implication/inclusive interpretations, whereas the invalid evaluation trials were not valid under any logical interpretation.

Procedure. The procedure was identical to that of Experiment 1, with one exception. For all the trials, once participants were ready to give a response, they selected a screen button marked *Ready to state answer*, and the time from presentation of premise(s) to response was recorded. The participants were then required to verbally state their responses (either yes/no or a generated paraphrase), and their responses were recorded on audiotape.

Results and Discussion

The data for 2 participants were omitted from analysis, since 1 participant spoke English only as a second language and another failed to understand the task requirement to produce a yes/no response on evaluation trials. The mean numbers of correct responses and response times in each condition for the remaining 24 participants are shown in Table 2.

As in the first experiment, the participants were more accurate in generating paraphrases (78% correct) than they were in evaluating either valid paraphrases (56%)

Table 2
Percentages of Correct Responses and Mean Response Times (RTs, in Seconds)
in Each Condition of Experiment 2, With Overall Means (and Standard Deviations)
for Each Response Format and Content (Four Trials per Participant)

Response Format	Content	Conditional-to-Disjunction		Disjunction-to-Conditional		No. Correct	
		% Correct	Mean RT	% Correct	Mean RT	<i>M</i>	<i>SD</i>
Generation	Arbitrary	83	19.4	73	22.5	3.13	0.95
	Location	77	24.3	79	24.6	3.13	1.12
Evaluation Invalid	Arbitrary	65	14.4	71	18.0	2.71	0.95
	Location	56	21.0	64	22.0	2.42	1.41
Valid	Arbitrary	54	22.0	71	20.8	2.50	1.18
	Location	40	22.6	60	23.4	2.00	1.10

or invalid paraphrases [64%; $F(2,46) = 17.7, MS_e = 0.13, p < .001$]. Planned comparisons showed the differences between generation and evaluation trials [$F(1,23) = 31.5, MS_e = 0.13, p < .001$] and between invalid and valid evaluation trials [$F(1,23) = 4.4, MS_e = 0.13, p = .048$] to be reliable. Thus, the predicted advantage for generation over evaluation was confirmed once more. Unlike Experiment 1, response times differed significantly [generation = 22.7 sec, valid paraphrase evaluation = 22.2 sec, invalid paraphrase evaluation = 18.7 sec; $F(2,46) = 6.8, MS_e = 6,101, p = .002$]. On this measure, the difference between generation trials and evaluation trials was not significant [$F(1,23) = 3.5, MS_e = 178, p = .07$]. However, response times on invalid evaluation trials were significantly shorter than response times on valid evaluation trials [$F(1,23) = 13.3, MS_e = 87, p < .001$]. The quicker and more accurate evaluations of invalid than of valid inferences are consistent with an early detection of mismatching model components on invalid trials and with an exhaustive fleshing out to detect a complete match on valid trials. These results provide further evidence that the participants used a comparison, rather than a generate-and-test, strategy to evaluate paraphrases.

As was predicted, the percentages of correct conclusions generated by the participants that were consistent only with inclusive and material implication interpretations of disjunctions and conditionals, respectively, were greater on trials with nonarbitrary relations (78%) than on trials with arbitrary relations (39%). Interestingly, the participants were also slower on trials with nonarbitrary relations (22.9 sec) than on trials with arbitrary relations [19.5 sec; $F(1,23) = 5.8, MS_e = 289, p = .03$]. This is suggestive of a further phase in which putative conclusions are validated against prior knowledge and in which counterexamples are cued by conclusions, further models are added to the premise representation, and an alternative conclusion is sought that is consistent in all models.

Only one other effect was significant in the analyses, an unexpected significant interaction in the accuracy data between task and rule [$F(2,46) = 4.5, MS_e = 0.14, p = .016$]. The only sizable difference in means was between evaluations of valid disjunction-to-conditional paraphrases (60%) and valid conditional-to-disjunction paraphrases (47%). This difference is consistent with the principle of minimal completion, assuming that partici-

pants construct and flesh out models of each premise in the order of presentation. To evaluate a disjunctive paraphrase of a conditional, participants must flesh out both sets of models completely in order to return a *valid* response. In contrast, if the first model of a conditional conclusion matches a fleshed-out model of the disjunctive premise, participants can return the response *valid* without fleshing out the implicit model of the conditional.

EXPERIMENT 3

The results of Experiment 2 confirmed the predicted advantage for generation over evaluation, and nonarbitrary relational content was found to influence the conclusions that the participants drew. So far, for the model theory of single-premise inferences we are advancing, we have considered only the effects of thematic content on a final reasoning phase in which putative conclusions are validated. Yet it seems likely that thematic content elicits semantic or pragmatic knowledge that facilitates reasoning performance from the outset, improving accuracy and generally reducing response latencies (for a review of such effects, see Manktelow, 1999). For example, Cheng and Holyoak (1985) found that assertions describing permissions and obligations facilitated certain inferences between *If P then Q* and *P only if Q* conditionals. However, the pragmatic reasoning schemas theory of Cheng and Holyoak makes no predictions regarding other content types, such as familiar indicative relations of the form *If a bird is a flamingo, then it is pink*.

According to mental models theory, the elicitation of semantic or pragmatic knowledge by thematic content operates more generally, adding known possibilities to the representation of assertions or blocking known impossibilities. The principle of minimal completion holds that possibilities, whether elicited from prior knowledge or from the fleshing out mechanism, are added to models under the control of the individual's current inferential goal. As a result, individuals may possess relevant knowledge, but it will affect their reasoning performance only when the possibilities it makes available are drawn to the individual's attention during the course of the task. A change in the method of Experiments 1 and 2 was intro-

duced in Experiment 3, whereby the premise and the conclusion were presented individually and in sequence, rather than in parallel. The change allowed us to control the point at which different possibilities were drawn to the attention of the participants. When the thematic content of the premise elicits prior knowledge, it draws to the attention of individuals possibilities that remain available for consideration when the conclusion is presented. When possibilities are drawn to the attention of the individual only by the conclusion, the models of the premise must be fleshed out to include possibilities cued by the conclusion before the models for the premise and the conclusion can be matched. The experiment tested two specific predictions: first, that content describing familiar semantic relations will facilitate only the evaluation of disjunction-to-conditional inferences and, second, that content describing general pragmatic relations of causality will facilitate only the evaluation of conditional-to-disjunction inferences. These predictions are explained below.

The basic representation of the disjunction *Either an animal is a snail or it does not have a shell* consists of two incomplete mental models and an implicit model:

snail
 ¬ shell
 ...

Semantic knowledge, that snails do indeed have shells, allows the immediate inclusion of the clause *has a shell* in the first model and of the clause *it is not a snail* in the second model:

snail shell
 ¬snail ¬ shell
 ...

Each explicit model contains both clauses, so it is possible to draw a conditional inference. The implicit model may also be fleshed out if individuals consider the possibility that there are animals other than snails that have shells. However, according to the principle of minimal completion, individuals are less likely to do so prior to the generation or presentation of a putative conclusion, since they would have to add a new model, rather than simply completing partial models. Moreover, in a paraphrase task, the individual's inferential goal can be met without fleshing out the implicit model. In essence, unless or until a putative conclusion cues a counterexample from prior knowledge, nothing in the task at hand draws to the individual's attention the relevance of considering any animals with shells other than snails.

The basic representation of a conditional, *If it is a snail then it has a shell*, elicits a single explicit model:

snail shell
 ...

Unlike its disjunctive counterpart, we suggest that familiar content adds nothing further to the initial repre-

sentation of this conditional. In essence, knowing that snails have shells does not immediately prompt individuals to consider things that are not snails or things that do not have shells. Thus, we predicted that evaluating conditional paraphrases of disjunctions would be facilitated by familiar content because it elicits prior knowledge that completes the partial models of the disjunctive premise prior to presentation of the conclusion to be evaluated. Familiar content would not facilitate the evaluation of disjunctive paraphrases of conditionals, because there are no partial models in the initial representation of the conditional that draw an individual's attention to relevant possibilities, even though the individual may have knowledge of such possibilities. Thus, the effects of familiar semantic content in a single-premise inference task are limited to the completion of initial models.

In contrast, causal content may add new possibilities (i.e., make implicit models explicit). Johnson-Laird and Byrne (1991, p. 70) suggested that causal content provides both the actual and the counterfactual situations associated with a relation. The counterfactual situation is provided because individuals possess relevant *pragmatic* knowledge regarding causal relations—namely, that in the absence of a cause, the effect will not arise. As a consequence, the representation of a causal conditional consists of two models, one for the possibility given in the premise, the other for its counterfactual possibility. For example, the causal conditional *If the conveyor stops then the pulper will slow down* elicits two explicit models:

Conveyor stops Pulper slows
 ¬ Conveyor stops ¬ Pulper slows

All the clauses necessary for matching the models of a disjunctive conclusion are represented in the models. This contrasts with the representation of a noncausal conditional that has a single model. Thus, causal content should facilitate the evaluation of disjunctive conclusions drawn from conditional premises.

The counterfactual possibility comes as part of the basic representation of a disjunction, such as *Either the conveyor stops or the pulper will not slow down*:

Conveyor stops
 ¬ Pulper slows

This premise is represented by two partial initial models, one for the antecedent possibility and one for its counterfactual consequent. The models remain incomplete because of the absence of familiar content. Thus, causal content will not facilitate evaluation of conditional paraphrases of disjunctions.

Method

Participants. Thirty-two first-year undergraduate psychology students from Loughborough University were paid £2 to take part. None had previous experience of logic or philosophy or of mathematics or computing courses in which logic had been taught.

Design and Materials. The participants were assigned at random to one of two groups, one receiving causal content, the other re-

Table 3
Mean Numbers of Correct Responses (With Standard Deviations) and Mean Comprehension and Decision Times (in Seconds) in Each Condition of Experiment 3

Content	Conditional → Disjunction						Disjunction → Conditional					
	No. Correct		Comprehension Times		Decision Times		No. Correct		Comprehension Times		Decision Times	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Familiar												
Causal	6.0	1.8	4.5	1.4	7.0	2.6	6.6	1.1	7.0	2.3	5.2	1.6
Noncausal	3.3	1.8	4.4	1.5	7.6	3.0	5.3	1.4	6.4	2.4	5.7	2.6
Unfamiliar												
Causal	5.5	1.8	8.5	2.5	8.4	2.8	4.2	1.5	10.5	3.0	6.7	3.0
Noncausal	4.1	1.0	6.6	2.0	6.5	2.3	4.1	1.4	8.5	3.3	6.0	2.2

ceiving noncausal content. Each participant evaluated conditional-to-disjunction and disjunction-to-conditional inferences, with both familiar and unfamiliar content. In each condition, the participants received four trials, consisting of valid and invalid paraphrases with both affirmative and negative antecedents.

The materials were those of Richardson and Ormerod (1997, Experiment 1), consisting of 16 disjunctions and 16 conditionals. The materials were rated for familiarity and causality by two independent judges in addition to the two authors, with 100% agreement. Examples of the four content types are (1) causal familiar, *If you cut your finger then it will bleed*; (2) causal unfamiliar, *If the redler conveyor stops then the pulper will slow down*; (3) noncausal familiar, *If the animal is a snail then it has a shell*; and (4) noncausal unfamiliar, *If the solid contains chloride then it does not absorb water*.

Procedure. The procedure was identical to the evaluation trials of Experiment 1, except that the premise was removed on presentation of the conclusion to be judged. This allowed us to separate the time taken by the participant to comprehend the premise (the comprehension time) from the time taken to judge the conclusion (decision time).

Results and Discussion

Analyses of variance were carried out on response frequencies, comprehension times, and decision times. The participants were more accurate in evaluating familiar paraphrases (68%) than unfamiliar paraphrases [56%; $F(1,30) = 22.5, MS_e = 0.042, p < .001$]. They were also quicker to comprehend familiar premises (5.6 sec) than unfamiliar premises [8.5 sec; $F(1,30) = 132.9, MS_e = 4.2, p < .001$], although decision times did not differ significantly [7.4 sec vs. 7.3 sec; $F(1,30) < 1$]. As predicted, familiar content facilitated disjunction-to-conditional paraphrases (familiar = 75%, unfamiliar = 52%), but not conditional-to-disjunction paraphrases [familiar = 61%, unfamiliar = 60%; $F(1,30) = 20.9, MS_e = 0.037, p < .001$].

The participants were also more accurate at evaluating causal paraphrases (70%) than noncausal paraphrases [54%; $F(1,30) = 10.0, MS_e = 0.156, p < .001$]. As was predicted, causality had more effect on conditional-to-disjunction paraphrases (72% vs. 50%) than on disjunction-to-conditional paraphrases [68% vs. 59%; $F(1,30) = 5.7, MS_e = 0.05, p = .02$]. The presence of both causal and familiar content gave the greatest degree of facilitation [familiar causal = 79%, familiar noncausal = 51%, unfamiliar causal = 61%, and unfamiliar noncausal = 51%; $F(1,30) = 6.0, MS_e = 0.04, p = .02$]. Unexpectedly, al-

though causal content led to greater accuracy, in the absence of familiar content, it lengthened the time taken to comprehend premises [5.7, 5.4, 9.5, and 7.5 sec, for familiar causal, familiar noncausal, unfamiliar causal, and unfamiliar noncausal content, respectively; $F(1,30) = 9.9, MS_e = 4.23, p < .001$] and to judge conclusions [6.1, 6.6, 7.6, and 6.2 sec, respectively; $F(1,30) = 12.0, MS_e = 4.52, p = .002$]. It appears that if the premise content was unfamiliar, the participants were slower to recognize and capitalize upon the causal relation.

Again, as in Experiment 2, there was no difference in the accuracy of evaluating conditional-to-disjunction paraphrases (61%) and disjunction-to-conditional paraphrases [63%; $F(1,30) = 0.7, MS_e = 0.05$]. However, conditional premises were comprehended more quickly (6.0 sec) than were disjunctive premises [8.0 sec; $F(1,30) = 57.9, MS_e = 4.8, p < .001$]. This result is consistent with the greater difficulty in reasoning with disjunctions than with conditionals found by Johnson-Laird et al. (1992) and attributed to the larger number of mental models that must be fleshed out in the basic representation of a disjunction. Decision times for conditional conclusions drawn from disjunctive premises were shorter (5.9 sec) than those for disjunctive conclusions drawn from conditionals [7.4 sec; $F(1,30) = 31.7, MS_e = 4.24, p < .001$]. As in Experiment 2, this result may reflect the fact that, in evaluating a conditional paraphrase of a disjunction, if the first model of the conditional matches a fleshed-out model of the disjunction, participants can return the *valid* response without needing to flesh out the implicit model of the conditional.

GENERAL DISCUSSION

In three experiments, we investigated how individuals paraphrase conditionals and disjunctions. The experiments tested a proposal that individuals draw inferences from single premises by constructing and fleshing out mental models of assertions under principles of truth (i.e., representing a clause given in an assertion only when it is true in the corresponding possibility) and minimal completion (i.e., fleshing out models only when necessary for the current inferential goal). To generate a conditional inference from a disjunction or vice versa, we pro-

pose that individuals first construct initial models of the premise. They flesh out the models in this representation until it can be matched against a basic representation of the target conclusion's connective. To evaluate the same kinds of inferences, we propose that individuals must construct and flesh out models of both the premise and the conclusion before checking whether the models match. The prediction that participants should be less accurate at evaluating single-premise inferences than they are at generating them, because evaluation requires the fleshing out of two sets of models rather than one, was confirmed in Experiments 1 and 2. To our knowledge, this is the first demonstration of an advantage for generation over evaluation in studies of reasoning.

We also tested proposals concerning the effects of semantic and pragmatic content on the evaluation of single-premise inferences. Under the principle of minimal completion, individuals add possibilities drawn from prior knowledge of semantic content to a model representation only if the possibilities appear relevant to the task at hand. The basic representation of a disjunction draws two partially explicit possibilities to the attention of an individual, and these possibilities are made fully explicit if semantic content elicits knowledge of the possibility. In contrast, the basic representation of a conditional contains a single fully explicit possibility. Unless cued by a putative conclusion to consider a counterexample, individuals are not alerted to the relevance of other possibilities, even if they may have prior knowledge of them. A prediction that nonarbitrary relational content would cue counterexamples that would encourage the generation of alternative conclusions consistent only with exclusive/material-implication interpretations of assertions was confirmed in Experiment 2.

According to our proposals, unfamiliar causal relations elicit the pragmatic knowledge that effects typically do not occur in the absence of their causes. In the case of a conditional assertion, this makes available the counterfactual possibility to that given in the explicit model of the conditional and, so, adds a new model to the representation of a conditional assertion. In the case of a disjunctive assertion, the counterfactual possibility is already partially represented in one of the two incomplete models of the disjunction, so causal content does not add anything to the representation of a disjunction. Two predictions that follow from these proposals—that semantic content concerning familiar indicative relations would facilitate the evaluation only of disjunction-to-conditional inferences and that pragmatic content concerning relations would facilitate the evaluation only of disjunction-to-conditional inferences—were confirmed in Experiment 3.

To what extent can other theories of human reasoning account for these findings? The answer is that, in their present form, no other theory can deal with them. We do not suggest that our results falsify other theories: Proponents can reasonably argue that their theories were not developed to account for single-premise inferences of

the kind investigated here. However, theories that ought, in principle, to be applicable to the single-premise inferences explored in this paper are the mental logic accounts of Rips (1994) and of Braine and O'Brien (1998). According to these theories, reasoning proceeds by applying logical inference rules in order to transform premises into a conclusion via a series of paraphrases. In other words, drawing inferences in *all* reasoning activities is characterized as a string of single-premise inferences.

Rips's (1994) *Psycop* would test the validity of the inference *Either P or Q, therefore if not P then Q* by applying a rule for *backward IF introduction* to the conclusion, as follows:

1. Given the conclusion *If not P then Q*, suppose *not P* and add a subgoal to prove *Q*.

The subgoal to prove *Q* is then satisfied by applying a *forward disjunctive syllogism* rule to the initial premise and the supposition; thus,

2. Given the premise *P or Q* and the supposition *Not P*, then *Q* follows.

The logic schemas of Braine and O'Brien (1998) allow approximately the same proof, invoking Schema 3 for OR elimination, which is similar to Rips's forward disjunctive syllogism. However, these theories encounter an immediate difficulty in accounting for the results reported here, since neither theory offers a way to derive the inference *If P then Q, therefore either not P or Q*. Rips introduces a conditional transformation rule in the same form as this inference to create a version of his program, *Psycop+*, that stops in a finite number of steps given any list of premises, thereby demonstrating a halting theorem. However, he claims "these rules seem to be exactly the sort that one would not want to include in a cognitively realistic deductive system" (Rips, 1994, p. 128).¹

The absence of a mechanism for proving *If P then Q, therefore not P or Q* might be justified if individuals showed some competence in drawing disjunction-to-conditional inferences but were unable to draw the opposite inference. A similar justification is made for including modus ponens and excluding modus tollens in mental logic theories. However, the present results show that generation and evaluation of conditional-to-disjunction inferences was above chance in a number of experimental conditions. Moreover, in Experiment 1, conditional-to-disjunction inferences were generated more accurately than disjunction-to-conditional inferences, the reverse of what might be predicted by mental logic theories. These results are not fatal for mental logic theories, since their authors acknowledge that the systems may be incomplete. Indeed, Rips (1994) suggests that "exactly which inferences are immediate or primitive is an empirical matter" (p. 112). However, adding a rule that allows the inference directly (in the same way that modus ponens follows from forward IF elimination in Rips's theory) will not suffice, since performance with the inference, although above chance, was not at ceiling.

Causing more problems for mental logic theories are the effects of response format found in Experiments 1 and 2 and the effects of thematic content found in Experiments 2 and 3. Rips (1994) states that “if we present [Psychop] with an argument to evaluate, the system will use [forward and backward] rules in an attempt to construct an internal proof of the conclusion from the premises. If we present the system with a group of premises and ask for entailments of those premises, the system will use the rules to generate proofs of possible conclusions” (pp. 104–105). In other words, generation of conclusions is achieved through a generate-and-test mechanism and, so, should always be at least as difficult, and often more difficult, than the evaluation of conclusions. Yet the opposite result was found in Experiments 1 and 2. Furthermore, logical inference rules are content independent and should apply equally regardless of the semantic content of an argument. Yet familiar and causal content facilitated the participants’ ability to evaluate paraphrases in Experiment 3. Mental logic theories have frequently been criticized for their inability to account for effects of thematic content on multiple-premise inferences, such as Wason’s (1966) selection task (e.g., Cheng & Holyoak, 1985; Evans & Over, 1996). The results reported here extend the demonstration of content effects to inferences from single premises, challenging the very cornerstone of mental logic theories. It may be possible for a theory that combines mental logic with pragmatic principles to account for the data presented in this paper. For example, Braine and O’Brien (1998) offer pragmatic principles that influence the interpretation of conditionals according to semantic content, expectations of relevance, and the existence of invited inferences. Precisely how such pragmatic principles are implemented and how they might, therefore, account for the present data remain unspecified. One possible implementation of these pragmatic principles is the mental models account of single-premise inferences offered here.

Specifying the earliest stages of premise representation may reveal much about reasoning, not only with inferences from single premises, but also with multiple-premise inferences. For example, Ormerod (1997) argued that the higher rate of modus tollens acceptance in a conditional inference task than in Wason’s (1966) selection task occurs because the inferential goal of each task is different: A conditional inference task requires simply the completion of a single possibility, whereas the selection task requires the discovery of every possibility in which a clause might occur. As a consequence, the sets of models required to accomplish each task differ. To generate a modus tollens inference requires a single additional model to represent the possibility that the antecedent might be false if the consequent is false. In the selection task, on the other hand, the reasoner must construct models for every possibility in which the consequent might be false before knowing whether to select the *not Q* card. Hardman and Payne (1995) also explained their finding of an advantage for evaluation over

generation in syllogistic reasoning in terms of the number of sets of models that must be constructed for each task. Generating a conclusion entails the search for counterexamples in alternative models of the syllogistic premises, whereas evaluating a conclusion requires only that a single set of models consistent with this conclusion can be found.

In each of these tasks, we argue, people adopt task-specific strategies, determined by their current inferential goal, that minimize the cognitive load entailed in representing and manipulating premise information. Where a task requires more than one set of models to be fleshed out, as is the case with the evaluation of single-premise inferences, the generation of syllogistic inferences, and Wason’s (1966) selection task, performance suffers in comparison with tasks that require the fleshing out of only a single set of models, such as the generation of single-premise inferences, the evaluation of syllogistic inferences, and the production or evaluation of conditional inferences.

REFERENCES

BAGULEY, T., & PAYNE, S. J. (1999). Recognition memory for sentences from spatial descriptions: A test of the episodic construction trace hypothesis. *Memory & Cognition*, *27*, 962-973.

BARWISE, J. (1986). Conditionals and conditional information. In E. C. Traugott, A. ter Meulen, J. S. Reilly, & C. A. Ferguson (Eds.), *On conditionals* (pp. 21-54). Cambridge: Cambridge University Press.

BRAINE, M. D. S., & O'BRIEN, D. P. (Eds.) (1998). *Mental logic*. Mahwah, NJ: Erlbaum.

BRANSFORD, J. D., BARCLAY, J. R., & FRANKS, J. J. (1972). Sentence memory: A constructive versus interpretative approach. *Cognitive Psychology*, *3*, 193-209.

CHENG, P. W., & HOLYOAK, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*, 391-416.

DUGAN, C. M., & REVLIN, R. (1990). Response options and presentation format as contributors to conditional reasoning. *Quarterly Journal of Experimental Psychology*, *42A*, 829-848.

EVANS, J. S. B., & OVER, D. (1996). *Rationality and reasoning*. Hove, U.K.: Psychology Press.

FILLENBAUM, S. (1974). Or: Some uses. *Journal of Experimental Psychology*, *102*, 913-921.

GLEITMAN, L. R., & GLEITMAN, H. (1970). *Phrase and paraphrase: Some innovative uses of language*. New York: Norton.

HARDMAN, D. K., & PAYNE, S. J. (1995). Problem difficulty and response format in syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, *48A*, 945-975.

JOHNSON-LAIRD, P. N. (1999). Deductive reasoning. *Annual Review of Psychology*, *50*, 109-135.

JOHNSON-LAIRD, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences*, *5*, 407-455.

JOHNSON-LAIRD, P. N., & BYRNE, R. M. J. (1991). *Deduction*. London: Erlbaum.

JOHNSON-LAIRD, P. N., & BYRNE, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*, 646-678.

JOHNSON-LAIRD, P. N., BYRNE, R. M. J., & SCHAEKEN, W. (1992). Propositional reasoning by model. *Psychological Review*, *99*, 418-439.

JOHNSON-LAIRD, P. N., & SAVARY, F. (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition*, *71*, 191-229.

KOTOVSKY, K., HAYES, J. R., & SIMON, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, *17*, 248-294.

MANKTELOW, K. I. (1999). *Reasoning and thinking*. Hove, U.K.: Psychology Press.

NEWSTEAD, S. E., & GRIGGS, R. A. (1983). Drawing inferences from

- quantified statements: A study of the square of opposition. *Journal of Verbal Learning & Verbal Behavior*, **22**, 535-546.
- OAKHILL, J. V., & JOHNSON-LAIRD, P. N. (1985). Rationality, memory, and the search for counter-examples. *Cognition*, **20**, 79-94.
- ORMEROD, T. C. (1997). Rationalities 1 and 2: Dual processes or different task demands? *Cahiers de Psychologie Cognitive*, **16**, 181-189.
- ORMEROD, T. C. (2000). Mechanisms and strategies for rephrasing. In W. Schaeken, G. De Vooght, A. Vandierendonck, & G. d'Ydewalle (Eds.), *Deductive reasoning and strategies* (pp. 131-152). Hove, U.K.: Erlbaum.
- ORMEROD, T. C., & JOHNSON-LAIRD, P. N. (2002). *How pragmatics modulates the interpretation of sentential connectives*. Manuscript submitted for publication.
- RICHARDSON, J., & ORMEROD, T. C. (1997). Rephrasing between disjunctives and conditionals: Mental models and the effects of thematic content. *Quarterly Journal of Experimental Psychology*, **50A**, 358-385.
- RIPS, L. J. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.
- TOMASSI, P. (1999). *Logic*. New York: Routledge.
- WASON, P. C. (1966). *Reasoning*. Harmondsworth, U.K.: Penguin.

NOTE

1. Yingrui Yang (personal communication, June 2001) offered an ingenious suggestion for proving the argument, using Braine and O'Brien's (1998, p. 307) Schema 7 (p OR q; IF p THEN r; IF q THEN s, THEREFORE, r OR s):

- | | |
|------------------------|--------------------------|
| 1. If P then Q | Premise |
| 2. If Not P then Not P | Truism |
| 3. P or not P | Truism |
| 4. not P or Q | By 1, 2, 3, and Schema 7 |

Braine and O'Brien provide truisms for predicate logic (pp. 309-310), but not for propositional logic. In any case, the psychological plausibility of the proof is open to question.

(Manuscript received August 17, 2001;
revision accepted for publication December 27, 2002.)