THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 1997, *50A* (2), 358–385

Rephrasing between Disjunctives and Conditionals: Mental Models and the Effects of Thematic Content

Juliet Richardson and Thomas C. Ormerod
Lancaster University, Lancaster, U.K.

Two experiments are reported that investigate whether the logical equivalence of conditionals and disjunctives is paralleled by a psychological equivalence. In these experiments, subjects rephrased from one form into the other. Experiment 1 demonstrated strong effects of familiarity and causality of rule content. Similar findings were found in Experiment 2 with a different conditional rule syntax. An account of the experiments is given in terms of mental models theory: In this account, task performance can be seen to depend upon the extent to which the model sets used by subjects to generate rephrasings are complete, task content being the most important factor affecting model set completion. A "Minimal Completion" strategy is proposed to operate in the absence of thematic content. The experiments also falsify the long-held assumption that conditionals with negative antecedents are always interpreted as their disjunctive equivalents. This raises doubts about the mental models explanation for matching bias in conditional reasoning.

A number of experimental paradigms have been used to explore human deductive reasoning, such as evaluation and truth-table construction and completion tasks, as well as more specialized paradigms like the Wason selection task (for a recent review see Evans, Newstead, & Byrne, 1993). A task that has received somewhat less attention is that of rephrasing between logically equivalent linguistic forms (e.g. Cheng & Holyoak, 1985; Fillenbaum, 1975; Ormerod, Manktelow, & Jones, 1993). Rephrasing is a task of theoretical interest, because it offers a vehicle for examining the psychological processes underlying the comprehension and interpretation of premise information independent of factors associated

Requests for reprints should be sent to T.C. Ormerod, Department of Psychology, University of Lancaster, LA1 4YF. Email: T.Ormerod@lancaster.ac.uk

The first author was funded by a research studentship awarded by British Gas plc. The experiments were conducted whilst the authors were at the Department of Human Sciences, Loughborough University. A report of Experiment 1 has previously appeared in the Proceedings of the 17th Cognitive Science Society conference, 1995. We thank Linden Ball and Jeremy Miles for providing assistance with statistics and commenting on drafts and are grateful to Philip Johnson-Laird for drawing to our attention the Stoic view of conditional and disjunctive rule interpretations. We also gratefully acknowledge the helpful comments of the anonymous referees, particularly for the suggested "Translation Table" account of rephrasing and suggestions for further experiments to explore rephrasing.

with the testing of conclusions. Rephrasing is also a task of considerable practical interest that has been explored in domains such as computer programming (Ormerod, 1988) and the specification of design requirements (Shepherd & Ormerod, 1992).

This paper concerns rephrasings between disjunctives and conditionals—a rephrasing not previously reported in the literature. This rephrasing is of particular interest because of a specific claim made by the Stoic logicians for the psychological equivalence of some conditionals and disjunctives. In the second century Galen stated that conditionals with a negated antecedent ". . . are called conditionals by those who pay attention only to the sounds, but a disjunction by those who pay attention to what is meant" (cited in Wason & Johnson-Laird, 1972). In other words, a conditional *If not P then Q* will always be interpreted as the disjunctive *Either P or Q*. For example, the statement "If I do not go out then I will do some work" will be interpreted as "Either I will go out or I will do some work".

This long-held assumption is embodied within a recent and influential theory of deductive reasoning: the mental models theory of conditional reasoning (Johnson-Laird and Byrne, 1991). In their theory, Johnson-Laird and Byrne argue that a mental representation of a conditional contains more information if the conditional has a negative antecedent than if it has an affirmative antecedent. Specifically, for a conditional with a negative antecedent, the affirmative counterpart is represented as well as the negative antecedent itself. As a consequence, the mental representations of conditionals with negative antecedents and their logically equivalent disjunctives are approximately the same. This assumption is perhaps most significant in relation to Johnson-Laird and Byrne's account of matching bias in the selection task, which is discussed later. However, we argue in this paper that there is little empirical support for the assumption, and that a mental models theory of conditionals can be formulated without this assumption being made.

The experiments reported in this paper examine the psychological treatment of conditionals and disjunctives in a rephrasing task. As well as providing an empirical test of the assumed psychological equivalence of conditionals having negative antecedents and their disjunctive rephrasings, the experiments also test a number of other predictions about rephrasing performance derived from mental models theory. In particular, the influence of familiarity and causality of rule content upon the construction of mental models is examined.

Logical and Psychological Equivalence of Conditionals and Disjunctives

There are two main ways in which a conditional of the form *If P then Q* can be interpreted: either as a biconditional, where *If P then Q* implies the converse *If Q then P*, or as an implication, where *Q* can occur in the absence of *P*. Similarly, there are two main ways in which a disjunctive of the form *Either P or Q* can be interpreted: It can be exclusively interpreted, where the statement is taken to mean *Either P or Q but not both*, or inclusively interpreted, where the statement is interpreted as *Either P or Q or both*. If truth tables for the two forms are compared (Shown in Table 1) it can be seen how a conditional that is interpreted as a biconditional is logically equivalent to two forms of an exclusively inter-

TABLE 1
Truth Tables for a Conditional Rule with a Negative Antecedent and Its Logically Equivalent Disjunctive Forms

		Interpretation	
Rule Form	*Premise*	*Biconditional*	*Implication*
If not P then Q	P Q	F	T
	P –Q	T	T
	–P Q	T	T
	–P –Q	F	F
		Exclusive	*Inclusive*
Either P or Q	P Q	F	T
	P –Q	T	T
	–P Q	T	T
	–P –Q	F	F
Either not P or not Q	P Q	F	F
	P –Q	T	T
	–P Q	T	T
	–P –Q	F	T

Note: *If not P then Q* can be rephrased with equal logical equivalence as *Either P or Q* or *Either not P or not Q*, the reversed truth table for an inclusive interpretation of *Either not P or not Q* simply reflecting the inverted rule polarity (T = true, F = false).

preted disjunctive. In contrast, a conditional that is interpreted as an implication is logically equivalent to only one form of an inclusively interpreted disjunctive.

The logical equivalence of conditionals and disjunctives may be demonstrated by truth-table analyses: Whether these forms are psychologically equivalent is an empirical question that has only been partially addressed. As outlined earlier, the Stoics assumed that a psychological equivalence exists between a conditional with negative antecedent and its disjunctive rephrasing. Some support for this assumption was provided by Evans (1972) using a truth-table evaluation task. He found that subjects' classifications of rule instances were more often consistent with a biconditional interpretation for conditionals having negative antecedents than for other conditional forms. However, although this suggests that conditionals with negated antecedents may be treated differently from the other forms, it does not demonstrate that this form of conditional is interpreted as a disjunctive.

Not all studies have shown a psychological equivalence between disjunctives and conditionals. For example, Johnson-Laird and Tagart (1969) investigated subjects' interpretation of the conditional *If P then Q* and its equivalent disjunctive *Not P or Q* using a truth-table evaluation task. They found that instances containing a negation of the antecedent as one component were most likely to be classified as irrelevant for conditionals, whereas for disjunctives these instances were most likely to be classified as true. These results may reflect the operation of a suppositional bias (Evans, 1983) in the interpretation of conditionals: subjects appear simply to ignore cases where the antecedent is denied and

therefore judge them to be irrelevant. This bias does not occur in the interpretation of disjunctives because it is triggered by the word "if" (Evans & Newstead, 1977; Ormerod et al., 1993). So even though the two forms are logically equivalent, they can give rise to different interpretations. However, as Johnson-Laird and Tagart only employed conditionals with affirmative antecedents, their findings do not test the Stoics' claim.

The psychological equivalence of conditionals and disjunctives clearly requires further investigation. A rephrasing task can help us to examine this, as psychological equivalence should give rise to highly accurate rephrasing performance. In particular, it would enable an empirical test of the assumed natural interpretation of conditionals with negative antecedents as disjunctives.

The Rephrasing Task

Although the paradigm is not commonly found in the reasoning literature, there have been some studies that examine rule rephrasing. For example, Cheng and Holyoak (1985) examined the effect of different rule contents on rephrasing between two types of conditional expression: *if p then q* and *p only if q*. They found that subjects' rephrasings were more accurate if the rules embodied a permission context (e.g. "If you are drinking alcohol then you are over 18") rather than arbitrary content (e.g. "If the letter is A then the number is 7"). They interpret this result in terms of their theory of pragmatic reasoning schemas. These are sets of rules that guide the inferences that can be drawn from a particular statement. Statements with a permission context evoke schematic rules that support the process of rephrasing between some conditionals but not others. However, pragmatic reasoning schemas theory cannot account for temporal order biases that occur between rule forms in the absence of contextual information (Ormerod et al., 1993).

Ormerod et al., (1993) investigated rephrasing between three types of conditional: *if antecedent then consequent* (IT), *antecedent only if consequent* (OI), and *consequent if antecedent* (IF). They found that the number of correct rephrasings between the OI and IF forms was close to zero, whereas performance on other rephrasings was at or above chance level. Ormerod et al. propose that subjects form a mental representation of the meaning of the original rule that depends upon its linguistic structure and thematic content, and then they generate a rephrasing consistent with this representation. They account for the low level of performance in rephrasing between OI and IF forms as being due to subjects merely dropping or adding the word "only" rather than considering the contingencies embodied in the rule to be rephrased. In other words, subjects did not get as far as forming a mental representation of the original rule's meaning; they simply performed a linguistic transformation of the original rule.

Studies have also investigated rephrasing from conditionals into other rule forms. Fillenbaum (1975) asked subjects to rephrase sentences of the form *if p, q* and *if not p, q* into another form of their choice as accurately as they could without using the word "if". He found that subjects tended to rephrase causal promises using "and". Conversely, they tended to rephrase causal threats using "or", especially if the initial conditional had a negated antecedent (*if not p, q*). Fillenbaum (1976) asked subjects to make equivalence

judgements between different sentences. Subjects were given an initial conditional (*if p, q* or *if not p, q*) and were asked to compare it with a second sentence (*p and q, not p or q* or *p or not q*). He found no difference between the judgements for *if p, q* and *if not p, q*, though subjects were more likely to judge a causal threat as equivalent to *not p or q* and a causal promise as equivalent to *p or not q*. Fillenbaum's findings indicate that, with some thematic contents, disjunctives and conditionals are seen as equivalent. However, he looked only at rephrasing from conditionals into other forms and not vice versa and only at the effect of negating the antecedent and not the consequent.

These findings indicate that thematic content and linguistic form influence rephrasing performance. They also demonstrate the potential of the rephrasing paradigm: The task is within the scope of subjects, as performance is neither at floor nor ceiling and the data obtained can be usefully analysed.

A Mental Models Account of Conditional and Disjunctive Reasoning

In order to derive predictions about rephrasing between conditionals and disjunctives, a comprehensive theory of rule interpretation is required that can account for both abstract reasoning performance and effects of rule content. The mental models theory of Johnson-Laird and Byrne (1991) provides such a theory, although the explanations of rule content effects that are expanded upon further on are not always fully specified by Johnson-Laird and Byrne.

Mental models theory proposes that people construct an initial, possibly incomplete, model set[1] of the given premises. This is then fleshed out, if necessary, to provide a complete model set and is used to formulate and evaluate possible inferences. Johnson-Laird and Byrne provide a notation for describing mental models in which ¬ denotes a negated component, ... denotes the possibility for further models to be fleshed out in a model set, and [ ] denotes the exhaustive representation of one contingency within the model set. For example, given the premise "If the letter is A then the number is 2", an initial model set will be formed as shown in the first model set of Table 2 (adapted from Johnson-Laird & Byrne, 1991, pp. 43–51). If a subject is given the further premise "The letter is not A" and is asked what follows, the initial model set can then be fleshed out as a biconditional or as an implication. The biconditional model set supports the conclusion "The number is not 2", whereas the implication model set supports no unique conclusion. Therefore, the two ways of fleshing out the initial model set give rise to potentially different conclusions.

[1] Throughout this paper the term "model set" is used to describe the set of one or more individual mental models that subjects form as a mental representation of premise information: This avoids the ambiguity of "mental model" referring both to individual models that contribute to a mental representation and to the whole mental representation of premise information. Thus, each line in a model set represents an individual model, and a model set of a conditional or disjunctive may comprise between one and three individual models. Like Johnson-Laird and Byrne (1991), we distinguish between initial model sets and fleshed-out models sets as two stages in model formation. At each stage the models sets may be complete (i.e. having an exhaustive representation of all contingencies) or incomplete.

TABLE 2
Initial and Fleshed-out Mental Model Sets for Disjunctives and Conditionals[a]

Rule	Initial Model	Fleshed out: Biconditional	Fleshed out: Implication
If the letter is A then the number is 2.	A 2 . . .	[A] [2] [¬ A] [¬ 2]	[A] [2] [¬ A] [2] [¬ A] [¬ 2]
If the letter is A then the number is not 2.	A ¬ 2 . . .	[A] [¬ 2] [¬ A] [2]	[A] [¬ 2] [¬ A] [¬ 2] [¬ A] [2]
If the letter is not A then the number is 2.	[¬ A] 2 [A]	[¬ A] [2] [A] [¬ 2]	[¬ A] [2] [A] [2] [A] [¬ 2]
If the letter is not A then the number is not 2.	[¬ A] ¬ 2 [A]	[¬ A] [¬ 2] [A] [2]	[¬ A] [¬ 2] [A] [¬ 2] [A] [2]
Either the letter is A or the number is 2.	A 2	[A] [¬ 2] [¬ A] [2]	[A] [¬ 2] [A] [2] [¬ A] [2]
Either the letter is A or the number is not 2.	A ¬ 2	[A] [2] [¬ A] [¬ 2]	[A] [2] [A] [¬ 2] [¬ A] [¬ 2]
Either the letter is not A or the number is 2.	¬ A 2	[¬ A] [¬ 2] [A] [2]	[¬ A] [¬ 2] [¬ A] [2] [A] [2]
Either the letter is not A or the number is not 2.	¬ A ¬ 2	[¬ A] [2] [A] [¬ 2]	[¬ A] [2] [¬ A] [¬ 2] [A] [¬ 2]

[a] As proposed by Johnson-Laird and Byrne (1991).

A similar account is provided for disjunctive reasoning. For example, given the premise "Either the letter is A or the number is 2", the subject will form a set comprising two initial models, as shown in the fifth initial model set of Table 2. If the subject is then given the further premise "The letter is A" and is asked "What follows?", then the initial model set can be fleshed out in two possible ways corresponding to exclusive or inclusive interpretations. The exclusive model set supports the conclusion "The number is not 2", whereas the inclusive model set supports no unique conclusion.

Johnson-Laird and Byrne incorporate the Stoics' claim that a conditional with a negated antecedent is interpreted as a disjunctive. They state that, in the initial model set formed for these conditionals, "antecedents, unlike consequents, are exhaustively represented" (Johnson-Laird & Byrne, 1991, p. 68; see also Evans, Newstead, & Byrne, 1993, p. 88). Thus, for conditionals with a negated antecedent, the complementary posit-

ive item is represented as well as (or instead of) the negative one. The initial model set formed for *If not P then Q* is:

$$[\neg P] \qquad Q$$
$$[P]$$

This is similar to the initial model set formed for the disjunctive *Either P or Q* (representation of the term [¬P] is, apparently, optional). Johnson-Laird and Byrne claim that this increases the likelihood of interpreting this type of conditional as a disjunctive.

The Stoics' assumption forms an important part of Johnson-Laird and Byrne's explanation of the matching bias that is observed in conditional reasoning performance. Matching bias is perhaps most clearly exemplified by subjects' performance in Wason's (1966) selection task. In this task, subjects are required to test the truth of a rule such as "If there is an A on one side of the card then there is a 7 on the other side" by selecting from four instances (e.g. "A", "T", "7" and "3") taken from a set of cards, each card having a letter on one side and a number on the other side. The most common response is to make the incorrect selection "A" and "7", matching the instances named in the rule. Evans and Lynch (1973) varied the presence of negative components in the rules and found that in all cases subjects' selections matched the cards named in the rules, regardless of the effect that negating rule components might have on the logical status of subjects' selections. Thus, for example, given the rule "If there is not an A on one side of the card then there is a 7 on the other side", then the most common response is, again, incorrectly to choose "A" and "7". This is the same response as that given to the rule "If there is an A on one side of the card then there is a 7 on the other side". Evans (1989) argues that this matching bias arises because the word "if" encourages the subject to consider the situation in which the antecedent is true. Johnson-Laird and Byrne provide a mental models account of this theory (1991, pp. 76–80). They argue that both affirmative and negative conditionals have model sets in which the positive items are explicitly represented. Subjects will use these model sets to inform their selections. Thus the model set for the rule "If there is an A on one side of the card then there is a 7 on the other side" is:

$$[A] \qquad [7]$$
$$\ldots$$

Subjects will choose the two positive items on the basis of this model set. Similarly, for the rule "If there is not an A on one side of the card then there is a 7 on the other side", the model set is:

$$7$$
$$[A]$$

Again, on the basis of this model set two positive items will be selected. Thus, Johnson-Laird and Byrne's explanation of matching bias, one of the most pervasive phenomena in deductive reasoning research, depends upon the truth of the Stoics' assumption.

With the exception of conditionals having negative antecedents, as described above, the *initial* model sets formed for logically equivalent conditionals and disjunctives differ, according to Johnson-Laird and Byrne's theory. Conditionals have a single model in which both antecedent and consequent components are represented, whereas disjunctives have two models in which each component is represented separately. If biconditional and exclusive interpretations are made of conditionals and disjunctives, respectively, then there are two rephrasings from one form into the other that have identical *fleshed-out* model sets. Compare, for example, the model sets given in Table 2 for "If the letter is A then the number is 2", "Either the letter is not A or the number is 2", and "Either the letter is A or the number is not 2". On the other hand, if implicational and inclusive interpretations are made, then only one rephrasing has an identical fleshed-out model set.

Thus, the psychological equivalence of conditionals and disjunctives will depend upon whether subjects use initial or fleshed-out model sets to reason about each rule form. There are two situations under which fleshed-out mental model sets will be constructed. The first is when subjects are required to test a putative novel conclusion against a model set that represents premise information (a reasoning step that is not undertaken in a rephrasing task, as no novel conclusion is drawn). The second is when the rule to be represented contains meaningful thematic content, in which case a complete initial model set may be constructed that is equivalent to the fleshed-out model used to test putative conclusions.

Mental Models and Thematic Content

If thematic content determines the completeness of the model set used in reasoning, then the presence of thematic content is likely to influence rephrasing performance. Evans (1993) argues that most of the experiments quoted by Johnson-Laird and Byrne use arbitrary contents and that the explanation of pragmatic influences on reasoning is one of the weakest areas of development in the theory. Although this may generally be true, Johnson-Laird and Byrne provide an account of the effects of causal content, and this can be extrapolated to predict and subsequently test the effects of other content types on rephrasing performance.

Marcus and Rips (1979) found that biconditional interpretations of conditionals were more likely for causal statements (e.g. "If the ball rolls left then the red light flashes") than for non-causal statements (e.g. "If the fish is red then it is striped"). An explanation invoking *general* knowledge can be proposed to account for this finding. Subjects have general knowledge about causal relationships, notably that the consequent does not usually occur in the absence of the antecedent in causal events, which directs them to form a biconditional interpretation of conditionals. Johnson-Laird and Byrne state that for conditionals expressing causal relationships "general knowledge informs the choice of what to represent in the models" (1991, p. 70). They propose that the subject forms an initial model set representing the actual and counterfactual situations associated with the causal relationship. For example, if the subject is given a causal assertion such as "If the vase hadn't been dropped then it wouldn't have broken", then the possibility of one event occurring in the absence of the other is not considered, and the initial model set is built accordingly:

[dropped]	[broken]	Actual
[¬ dropped]	[¬ broken]	Counterfactual

Another factor affecting the interpretation of conditionals is familiarity. Markovits (1986) found that unfamiliar content led to more biconditional interpretations of conditionals. He argues that subjects can more easily generate examples of the consequent occurring without the antecedent when the content is familiar. Therefore, they are more likely to interpret a familiar statement as an implication and an unfamiliar statement as a biconditional. Thus it appears that prior knowledge of *specific* instances discourages subjects from forming a biconditional interpretation.

Although Johnson-Laird and Byrne do not describe in detail how familiar content (i.e. *specific* prior knowledge) affects the mental models formed, one might propose an account similar to that given for causality. If subjects are familiar with all of the possible combinations of antecedent and consequent conditions, then they may represent these explicitly in a complete initial model set. For example, given the statement "If it is a mammal then it is an animal", subjects form an initial model set that also represents cases where the consequent occurs in the absence of the antecedent:

[mammal]	[animal]
[¬ mammal]	[animal]
[¬ mammal]	[¬ animal]

If, on the other hand, the subject is given a rule with unfamiliar content, such as "If the quark is blue then the schmidt number is 10", then s/he will have no prior knowledge of occasions when the schmidt number is 10 but the quark is not blue. The subject will be unlikely to consider this contingency, and will therefore form an initial model set reflecting its absence:

[blue]	[10]
...	

The exhaustive representation of affirmative antecedent and consequent components in this initial model set represents a possible inference, following from the subjects' failure to consider all contingencies, that the schmidt number is 10 *only* when the quark is blue. This initial model set can only be fleshed out in a manner that is consistent with a biconditional interpretation:

[blue]	[10]
[¬ blue]	[¬ 10]

Thus, Markovits' findings concerning unfamiliar content and Marcus and Rip's findings concerning causality can both be accounted for by mental models theory. If the content is causal, then subjects incorporate *general* prior knowledge into their initial model sets, whereas if the content is unfamiliar, it is the absence of *specific* prior

knowledge that influences the initial model set formed. In both cases, subjects build initial model sets that can only be fleshed out in a way that is consistent with a biconditional interpretation.

Similar effects of thematic content have also been found with disjunctive reasoning. For example, Newstead, Griggs, and Chrostowski (1984) found that altering the context led to different interpretations of disjunctives. For example, a threat content (e.g. "Either you eat your dinner or you will go straight to bed") led to more exclusive interpretations than a qualification context (e.g. "The man I marry will have to be either rich or handsome"). Newstead and Griggs (1983) suggest that "or" is a fuzzy concept. It can correspond, amongst other things, to inclusive disjunction, exclusive disjunction, and expressions of intention. Also, Fillenbaum (1974) demonstrated that subjects judge disjunctives to be "strange" when they contain unrelated components. Compare, for example, "Either the ball is blue or it is green" with "Either the ball is blue or John is clever".

Johnson-Laird and Byrne do not discuss the effects of rule content for disjunctives. Nevertheless, a similar explanation to that offered for conditionals may explain the findings of Newstead et al. (1984) and Fillenbaum (1974). For example, when the rule has a threat context (e.g. "Either eat your dinner or go to bed"), then subjects will know from experience of threats that a person who does not fulfil the obligation must suffer the consequences (see also Cheng & Holyoak, 1985). The initial model set formed is therefore equivalent to a fleshed-out model set for an exclusive disjunctive:

[eat]	[¬ bed]
[¬ eat]	[bed]

In line with the account offered for conditionals, it might be expected that a subject's general knowledge will contribute to the model set formed for a causal disjunctive, whereas his/her specific knowledge will determine the model set formed for a familiar disjunctive. Thus, if a subject is presented with the familiar disjunctive rule "Either it is not raining or Bill will stay indoors", then s/he may use his/her knowledge about such situations to conclude that if it is raining then Bill will stay indoors and draw up the following model set:

[¬ rain]	[¬ indoors]
[rain]	[indoors]

It should be emphasized that predicted effects of familiarity with both conditionals and disjunctives, and predicted effects of causality with disjunctives, are based simply upon an extrapolation of Johnson-Laird and Byrne's (1991) account of the effects of causal content upon conditionals. Under the proposed account, with both conditionals and disjunctives, the effect of causal content would be to complete initial model sets through the representation of general knowledge about causal relationships, whereas the effect of familiar content would be to complete initial model sets through the representation of specific knowledge about familiar instances. The experiments reported in this paper enable the evaluation of this account using a rephrasing task.

Mental Models and the Rephrasing Task

The preceding discussion of mental models theory provides a basis from which to derive predictions concerning the effects of syntax and thematic content upon rephrasing performance. In order to do this, however, it is first necessary to develop a mental models theory for the generation of rephrasings. Three alternative strategies for the generation of rephrasings from mental models are presented; an "Initial Model Set" strategy, a "Translation Table" strategy,[2] and a "Minimal Completion" strategy. These three alternatives vary according to the extent of model set completion required for the generation of rephrasings, and each leads to different predictions of rephrasing performance. Our description of the three strategies seeks to account, in the first instance, for the rephrasing of rules that have unfamiliar and non-causal content, as this allows us to generate predictions concerning the effects of syntax on rephrasing performance independent of the additional effects on model completeness of familiar and causal content outlined above.

As a rephrasing task does not entail the testing of novel conclusions, it is possible that subjects develop rephrasings solely on the basis of the incomplete initial model sets that are formed through the representation of the original rule's syntax. Under this Initial Model Set strategy, and for rules with unfamiliar non-causal content only, a subject forms the incomplete model set representing the syntax of the initial rule presented to him/her. He or she then uses this initial model set to try to generate a rephrasing. The initial model sets for conditionals and disjunctives are different and do not provide sufficient information to inform the choice of both rule components in the rephrasing. Thus, a floor effect would be predicted if this strategy were applied, with no correct rephrasings for rules with unfamiliar non-causal content regardless of whether the initial rule is a conditional or a disjunctive.

An alternative account of rephrasing is the Translation Table strategy. Under this account, subjects search for a match between the model set of the rule to be rephrased and the model sets of all the possible rephrasings. If a match is found, then the syntax that gives the matching model set is the one chosen for the rephrasing. The Translation Table strategy is based upon two assumptions: first, that subjects construct and compare model sets for both the original rule and the possible rephrasing(s), and, second, that the initial model sets of these rules be fleshed out to enable such comparisons to be made (as a comparison of the incomplete initial model sets that result through representation of rule syntax alone would reveal no matches). As a consequence, such a strategy is more cognitively demanding than is the Initial Model Set strategy, but it allows for some degree of competence at rephrasing.

If a Translation Table strategy is adopted, then the syntax of the rule to be rephrased may affect subsequent rephrasing performance, leading to an asymmetry in rephrasing from conditionals and from disjunctives when the rule content is unfamiliar and non-causal. The initial model sets for disjunctives contain two incomplete models, compared with one complete model for conditionals. It may be argued that the establishment of two (albeit incomplete) models makes the initial model sets of disjunctives closer to the

[2] Suggested by an anonymous referee of an earlier draft of this paper.

fleshed-out model sets of conditionals than the initial model sets of conditionals are to the fleshed-out model sets of disjunctives. Therefore, if disjunctive model sets are easier to complete than conditional model sets, then rephrasings from disjunctives should be easier than rephrasings from conditionals when the rule content is unfamiliar and non-causal.

Under a Minimal Completion strategy, subjects undertake a series of steps in which they represent the original rule's syntax in an initial model set, add one component to the rephrasing, partially complete the initial model set to represent the second component, and then use this partially completed model set to generate the rephrasing. The completion is minimal in that the only component added to the initial model set is the as yet unrepresented term from the original rule. The idea that subjects attempt rephrasings from partial rather than completely fleshed-out model sets is consistent with the claim that subjects leave as much information implicit in their models as possible rather than spelling it out explicitly (Johnson-Laird, Byrne, & Schaeken, 1992). The Minimal Completion strategy is illustrated in Table 3.

The Minimal Completion strategy also leads to a predicted asymmetry in rule rephrasings for rules with unfamiliar and non-causal content, though one in exactly the opposite direction to that predicted by the Translation Table strategy. One of the principles of mental models theory is that the more models needed to represent premise information, the greater the load imposed upon a limited-capacity working memory system. Given that one initial model is formed for a conditional and two initial models are formed for a disjunctive when the rule content is unfamiliar and non-causal, then rephrasing should be harder from a disjunctive into a conditional than when the rephrasing is in the opposite direction.

There is one important consideration that makes the adoption of a Minimal Completion strategy more likely than a Translation Table strategy on a priori grounds: Con-

TABLE 3
The Minimal Completion Strategy for Generating Rephrasings

	Rephrasing from	
Step	"*If Circle then Triangle*"	"*Either Circle or Triangle*"
1. Create initial model set from rule syntax	circle triangle . . .	circle triangle
2. Generate a part rephrasing with one component	"Either there is a circle . . ."	"If there is a circle . . ."
3. Flesh out the second component in the original model set	circle triangle ¬ triangle triangle (?)	circle ¬ triangle triangle ¬ triangle(?)
4. Complete the rephrasing	"Either there is a circle or there is not a triangle"	"If there is a circle then there is not a triangle"

Note: (a) Step 2 could involve the addition of either first or second components. (b) The representation of both contingencies of the second component in Step 3 is optional. (c) Steps 3 and 4 can, but need not necessarily, lead to the generation of a correct rephrasing.

structing model sets for both the original rule *and* the putative rephrasing would greatly increase the task demands. A Translation Table strategy not only requires the construction of a model set for the putative rephrasing, it requires the construction (or at least elicitation from memory) and search through all *four* possible model sets for the alternative syntax polarities. It *may* be that some subjects carry out additional steps after they have generated a rephrasing, in which they represent the model set entailed in the rephrasing's syntax and then compare this against the model set for the original rule. Such additional steps might improve rephrasing accuracy but would also complicate the task greatly. Thus, we believe a Translation Table approach to be unlikely on grounds of cognitive economy. Nonetheless, the differing predictions concerning a rephrasing asymmetry enable a choice between the accounts to be made on the basis of empirical evidence.

The three strategies lead to different predictions concerning rephrasing performance from each syntax when the rule content is unfamiliar and non-causal. It was argued above that the presence of familiar or causal content enables the formation of a complete initial model set. Thus, performance should therefore be more accurate for rephrasings with familiar and/or causal content. Furthermore, as the complete model set of a rule is equivalent to the model set for its logically equivalent rephrasing, an asymmetry in rephrasing from conditionals or from disjunctives is not expected when rules contain familiar and/or causal content.

EXPERIMENT 1

Experiment 1 investigated the influence of rule syntax, polarity, and thematic content upon performance at a task in which subjects were required to rephrase conditionals into disjunctives, and vice versa.

The experiment enabled the test of a number of hypotheses deriving from mental models theory. First, it enabled an empirical test of the rephrasing strategies outlined above, depending upon the presence of a floor effect in rephrasing performance (evidence favouring the Initial Model Set strategy), a rephrasing asymmetry that confers advantage on rephrasings from disjunctives (evidence for the Translation Table strategy), or a rephrasing asymmetry that confers advantage on rephrasings from conditionals (evidence for the Minimal Completion Strategy). It should be noted that these predicted asymmetries apply only to rules with unfamiliar and non-causal content.

Second, the Stoics' claim for psychological equivalence of negative antecedent conditionals and disjunctives was evaluated. Based upon the assumption of Johnson-Laird and Byrne (1991) that negated antecedents in conditionals are represented exhaustively in initial model sets, one can predict that performance will be better when subjects are rephrasing from conditionals with negated antecedents than from other conditionals. Furthermore, if familiar and causal content lead to completed initial model sets, then fewer errors should be made when rephrasing rules with familiar and/or causal content. Finally, if causal and familiar contents give rise to complete initial model sets, then working memory loads are equivalent for both rule forms. Thus the assymetry in rephrasing from each rule predicted by the Minimal Completion strategy should only be observed when the content is non-causal and unfamiliar.

Method

Subjects

Seventy-one students from Loughborough University participated in the study as part of a first year course in Experimental Psychology.

Materials and Design

Four factors were manipulated in this experiment. The first was the between-subjects factor of causality: Subjects were randomly assigned to one of two groups, one rephrasing causal rules, the other rephrasing non-causal rules. The other three factors were all within-subjects. The first of these was familiarity: subjects received both familiar and unfamiliar rules to rephrase. The unfamiliar rules concerned chemical processes, and the familiar rules concerned everyday situations. The rules were rated for familiarity and causality by two independent judges, whose judgements agreed 100% with our own. Examples of the four content types are:

causal familiar	If it is raining then the ground will be wet.
causal unfamiliar	If the isopropanol rises then the valve will open.
non-causal familiar	If it is a flamingo then it is pink.
non-causal unfamiliar	If it is iodide then its schmidt number is 36.

The second within-subjects factor was the original rule: Subjects generated disjunctive rephrasings from an original conditional and vice versa. The final factor was polarity: for each type (e.g. familiar causal disjunctives), subjects were presented with the four possible combinations of negated elements. Therefore, each subject rephrased 16 rules (shown in Appendix 1).

Procedure

The experiment was presented in a booklet containing instructions (shown in Appendix 2), followed by the 16 trials, one per page, in a different randomized order for each subject. Subjects wrote their rephrasings in the booklet. They completed each rephrasing at their own pace but were unable to change previous answers in the light of subsequent rephrasings. The duration of the experiment was 20 minutes.

Results

Number of Correct Rephrasings. Subjects' rephrasings were judged correct if they could be judged as being logically equivalent to the original rule according to any possible interpretation of that rule. Thus for each rule two possible forms of rephrasing were allowed. Also, implicit as well as explicit negatives were allowed, for example "dry" was allowed in place of "not wet". Table 4 shows the average percentage of correct rephrasings for each original rule syntax and content.

Analysis of variance showed a significant main effect of original rule, $F(1, 69) = 4.22$, $p < .05$. Performance was better when rephrasing from a conditional (76.7% correct) than from a disjunctive (72.1% correct). There was a significant main effect of familiarity, $F(1, 69) = 34.37$, $p < .01$. Rephrasing from familiar rules (81.5% correct) was significantly better than rephrasing from unfamiliar rules (67.2% correct). There was a significant main effect of polarity, $F(3, 207) = 7.01$, $p < .05$. Post-hoc Tukey's tests show

TABLE 4
Percentages of Correct Rephrasings Obtained in Experiment 1

			Causal		*Non-causal*	
Rephrasing	*Familiarity*	*Polarity*	*%*	*Mean*	*%*	*Mean*
conditional to disjunctive	familiar	P/Q	92.5		58.1	
		P/– Q	87.5		64.5	
		– P/Q	82.5		67.7	
		– P/– Q	80.0	85.6	64.5	63.7
	unfamiliar	P/Q	85.0		67.7	
		P/– Q	90.0		80.6	
		– P/Q	80.0		61.3	
		– P/– Q	85.0	85.0	80.6	72.6
disjunctive to conditional	familiar	P/Q	97.5		87.1	
		P/– Q	97.5		87.1	
		– P/Q	90.0		77.4	
		– P/– Q	90.0	93.8	80.6	83.1
	unfamiliar	P/Q	85.0		64.5	
		P/– Q	80.0		51.6	
		– P/Q	52.5		29.0	
		– P/– Q	50.0	66.9	32.3	44.4

that rephrasings from rules with positive first components (*p/q* = 79.7%; *p/not q* = 79.9%) were correct significantly more often than rephrasings from rules with negative first components (*not p/q* = 67.6%; *not p/not q* = 70.4%). There was also a significant main effect of causality, $F(1, 69)=16.02$, $p < .01$. Rephrasing from causal rules (82.8% correct) was significantly better than rephrasing from non-causal rules (66.0% correct).

The two-way Original Rule × Causality, $F(1, 69) = 0.02$, and Familiarity × Causality, $F(1, 69) = 0.06$, interactions were not significant. There was a significant two-way Original Rule × Familiarity interaction, $F(1, 69) = 65.44$, $p < .01$. Rephrasing from disjunctives was better when the content was familiar (88.5% correct) than when it was unfamiliar (55.7% correct). This effect of familiarity was not seen in rephrasing from conditionals. There was also a significant three-way Original Rule × Familiarity × Causality interaction, $F(1, 69) = 5.46$, $p < .05$. The facilitatory effect of familiarity on rephrasing from disjunctives was greater when the content was non-causal than when it was causal. The three-way Original Rule × Causality × Familiarity interaction was also significant, $F(1, 69) = 5.462$, $p < .05$. The effect of rule form was greatest when the content was unfamiliar and non-causal.

The two-way Original Rule × Polarity, $F(3, 207) = 4.71$, $p < .01$, and Familiarity × Polarity, $F(3, 207) = 3.14$, $p < .05$, interactions were significant, though the two-way Causality × Polarity interaction, $F(3, 207) = 0.61$, was not significant. The three-way Original Rule × Familiarity × Polarity interaction was also significant, $F(3, 207) = 3.36$, $p < .05$. Post-hoc Tukey's tests show that rephrasings from unfamiliar disjunctive rules with a negative first component were significantly less accurate than all other rephrasings, at $p < .01$.

Error Types. Incorrect rephrasings can be divided into three classes: logically incorrect rephrasings in which subjects used the correct rule syntax but with an incorrect combination of polarities, syntactically incorrect rephrasings in which subjects used an incorrect rule syntax, and rephrasings to which subjects responded "none" because they believed that no rephrasing was possible. Logically incorrect rephrasings can be further subdivided into rephrasings in which subjects changed the polarity of neither of the elements from the original rule (labelled "neither") and rephrasings in which subjects changed the polarity of both the elements from the original rule (labelled "both"). The percentages of rephrasings for which each type of error was observed are given in Table 5. The most common type of incorrect rephrasing was one in which the subject changed the polarity of neither of the elements from the original rule (57.4% of all incorrect rephrasings).

Discussion

There are several interesting findings from this experiment. First, rephrasing performance is not at ceiling in any condition and is especially poor in the absence of familiar or causal content. Thus, a simple assumption of psychological equivalence to match the logical equivalence of conditionals and disjunctives can be rejected. On the other hand, the absence of a floor effect for unfamiliar non-causal rules undermines the Initial Model Set strategy, suggesting that subjects must use more than simply the initial model set entailed by the syntax of the original rule in generating rephrasings.

Second, no advantage was found for rephrasings from conditionals with negative antecedents over rephrasings from conditionals with positive antecedents. Thus the assumed psychological equivalence of disjunctives and conditionals with negated antecedents is not supported. Subjects do not appear to interpret conditionals of the form *If not p then q* differently to other conditionals.

The rephrasing of unfamiliar non-causal rules was more accurate when the original rule was a conditional rather than a disjunctive. This rephrasing asymmetry is consistent

TABLE 5
Percentages of Rephrasings for which Each Type of Error was Observed in Experiment 1.

	Rephrasing	*Familiarity*	*Neither*	*Both*	*None*	*Syntax*	*Total*
Causal	conditional to disjunctive	familiar	8.1	5.0	0.6	0.7	14.4
		unfamiliar	5.6	6.3	3.1	0.0	15.0
	disjunctive to conditional	familiar	2.5	0.6	3.1	0.0	6.2
		unfamiliar	20.6	11.9	0.6	0.0	33.1
Non-causal	conditional to disjunctive	familiar	17.7	8.1	8.1	2.4	36.3
		unfamiliar	11.3	9.7	6.5	0.0	27.4
	disjunctive to conditional	familiar	11.3	0.8	4.8	0.0	16.9
		unfamiliar	41.1	9.7	4.8	0.0	55.6

Note: Neither = Polarity of neither element altered. Both = Polarity of both elements altered. None = Subject believes no rephrasing is possible. Syntax = Syntax of rephrasing is incorrect.

with the use of a Minimal Completion strategy rather than a Translation Table strategy, and it demonstrates the influence that the *number* of mental models represented in a model set has on rephrasing in the absence of other factors (such as familiar or causal content) that affect model set formation. The asymmetry was also evident in rephrasings from conditionals that had negative antecedents, a finding that is inconsistent with Johnson-Laird and Byrne's (1991) claim that a negative antecedent of a conditional is always represented exhaustively (leading to two models rather than one in the initial model set). Instead, conditionals with negative antecedents appear to be represented in the initial model set, like other conditionals, by a single model.

Various effects of rule content were found. Causal content gave rise to better rephrasing performance than non-causal content. Similarly, familiar content gave rise to better rephrasing performance than unfamiliar content. An interaction was found between familiarity and original rule: The facilitatory effect of familiar content was confined to rephrasings from disjunctive rules. There was also a three-way interaction: the enhanced facilitatory effect of familiar content with disjunctives was greater when the rule was non-causal rather than causal.

These results suggest that, in rephrasing from conditionals, causality dominates performance. Causal content appears to lead subjects to form a complete initial model set, thus making the task of rephrasing easier. However, no facilitatory effects were found with conditionals arising through the presence of familiar content. This suggests, contrary to the account offered in the Introduction, that familiarity does not give rise to a complete initial model set with conditionals. In rephrasing from disjunctives, on the other hand, familiarity dominates performance. Familiar content appears to lead subjects to complete initial model sets for disjunctives. Causal rule content may affect rephrasing from original disjunctives only in aiding the subsequent interpretation of the initial model set as a conditional.

An unexpected finding was that performance was better in rephrasing from unfamiliar disjunctives with positive rather than negative first components. Post-hoc inspection of the subjects' rephrasings reveals that a positive antecedent was maintained in rephrasing from unfamiliar disjunctives with positive first components in 80.4% of correct rephrasings and 82.5% of incorrect rephrasings, whereas a negative antecedent was maintained in rephrasing from unfamiliar disjunctives with negative first components in 45.0% of correct rephrasings and 66.7% of incorrect rephrasings. Thus there appears to be a tendency for subjects to generate or maintain a positive first component in their rephrasings, which is reduced where rephrasings produced from unfamiliar disjunctives with negative first components are incorrect. A tendency to produce positive antecedents in rephrasing from disjunctives into conditionals may be related to the operation of a suppositional bias on the comprehension of conditionals, in which the word "if" invites the supposition of the antecedent that it modifies (Evans, 1983). The operation of suppositional bias on conditionals may increase the likelihood of error in subsequent completion of the rephrasing in cases where subjects fail to make this conversion.

The most common error was a rephrasing that maintained the same polarity for both elements from the original rule. The absence of familiar or causal content may make it more difficult to construct an initial model set that is sufficiently complete to implement a Minimal Completion strategy. As a consequence, subjects may bypass the construction of

a model set entirely, instead of relying on a linguistic transformation of the rule that simply substitutes the words "either" and "or" for the words "if" and "then", or vice versa. A similar linguistic transformation was proposed by Ormerod, Manktelow, and Jones (1993) to underlie the floor effect they found in subjects' rephrasings between *P only if Q* and *Q if P* rule forms. Alternatively, the failure to implement a Minimal Completion strategy successfully may leave subjects undertaking what is in effect the Initial Model Set strategy. If, in forming a rephrasing, subjects represent only those elements that are explicit in the initial model set entailed by the original rule's syntax, then they will end up generating a rephrasing that does not change any of the elements from the initial rule. This is equivalent to the "focusing" effect proposed by Legrenzi, Girotto, and Johnson-Laird (1993) to explain performance on Wason's (1966) selection task.

EXPERIMENT 2

Some of the predictions derived from mental models theory were substantiated in Experiment 1. However, the expected facilitation of negating the antecedent of a conditional prior to rephrasing into a disjunctive was not obtained. This appears to undermine the assumption of special status for this particular polarity of conditional, but there are alternative explanations. It may be, for example, that the relative ease of rephrasing between conditionals with negative antecedents and their disjunctive equivalents is countered by a general increase in difficulty in rephrasing from any logical rule, regardless of its syntax, that contains a negative first component. Indeed, this is exactly the result that was obtained in Experiment 1 with unfamiliar disjunctives, where rephrasing from disjunctives with negative antecedents was shown to be more difficult than from disjunctives with affirmative antecedents.

In order to choose between these two possibilities, Experiment 2 was conducted, in which subjects carried out essentially the same task as in the first experiment, except that the rephrasings were between disjunctives and the ... *only if* ... form of conditional. This conditional form was chosen because, although it is logically equivalent to the *If P then Q* form, the initial model sets of the two forms are proposed to differ. Specifically, in reasoning with the *P only if Q* rule form, a subject will form "two explicit models right from the start ... one represents the positive contingency ... the other represents the negative contingency" (Johnson-Laird & Byrne, 1991, p. 50; see also Ormerod, Manktelow, & Jones, 1993 for a mental models account of ... *only if* ... and *If* ... *then* ... rule forms). This can be compared with the initial model set formed for an expression of the form *If P then Q*, which contains only one explicit model. The initial model sets for each form (assuming they contain unfamiliar and non-causal content) are (from Johnson-Laird & Byrne, 1991, p. 51):

If P then Q	P	Q
	...	
P only if Q	[P]	Q
	[¬ P]	[¬ Q]
	...	

Two experimental findings support this distinction between the two forms in mental models theory. First, Evans (1977) found that subjects made more Modus Ponens than Modus Tollens inferences with the *If P then Q* form, whereas the two inferences were drawn equally often with the *P only if Q* form. The mental models explanation of this finding is that the initial model set of the *P only if Q* form can support both deductions without further fleshing out, whereas the initial model set for the *If P then Q* form supports Modus Ponens but must be fleshed out before the Modus Tollens deduction can be made.

Second, Girotto, Mazzocco, and Cherubini (1992) found that when the expressions had a deontic content, performance on Wason's selection task was better for ... *only if* ... than for *If* ... *then* ... expressions. Legrenzi, Girotto, and Johnson-Laird (1993) argue that people have a tendency to consider only items that are explicitly represented in the mental model set that they form. Thus when asked to test a rule of the form *If P then Q*, subjects only consider the positive items *P* and *Q* represented in the model set and not the correct items *P* and *not Q*. When the rule is of the form *P only if Q*, then the correct items are explicitly represented in the model set and so are more likely to be considered.

The initial model set formed for *P only if Q* is equivalent to the initial model sets formed for its disjunctive equivalents (see Table 2). Thus, in the absence of causal or familiar content, rephrasing from conditionals of the form *P only if Q* into disjunctives should be more accurate than from conditionals of the form *If P then Q*. Furthermore, in Experiment 1 it was found that rephrasing was easier from conditionals into their disjunctive equivalents when the content was unfamiliar and non-causal. This asymmetry was explained as resulting from the fact that fewer models are required to represent a conditional of the form *If P then Q* in an initial model set than are required to represent an equivalent disjunctive. The use of the *P only if Q* form allows a test of a different prediction concerning the influence of the *number* of models within a model set upon rephrasing performance. As the initial model set of the *P only if Q* form contains two models, the load on working memory should be equal in representing both forms prior to rephrasing. Consequently, in Experiment 2 there should be no asymmetry in rephrasing from disjunctives and conditionals.

Method

Subjects

Fifty-seven subjects from Loughborough University participated in this study as part of a first year course in Experimental Psychology. None of these subjects had taken part in Experiment 1.

Materials and Design

The materials and design were the same as for Experiment 1 except for two points. First, the conditional rules used were of the form ... *only if* ... rather than *If* ... *then* ... . Second, for each rule type (e.g. causal, familiar disjunctive), subjects received rules of the form P/Q and – P/Q rather than rules of all four combinations of polarity. Therefore each subject rephrased eight rules in total.

Procedure

The experimental presentation was the same as for Experiment 1 except that the instructions given were adapted for rephrasing between ... *only if* ... conditionals and disjunctives.

Results

Table 6 summarizes the data obtained from this experiment in terms of average percentage of correct rephrasings for each original rule and content. The correctness of the rephrasings was judged by the same criteria as were used in Experiment 1, and a similar analysis of variance was conducted.

There was no effect of original rule on rephrasing performance, $F(1, 55) = 0.1$. A significant main effect of familiarity was found, $F(1, 55) = 26.86$, $p < .01$, with rephrasing from familiar rules (92.5%) significantly better than rephrasing from unfamiliar rules (76.8%). There was no main effect of causality on rephrasing performance, $F(1, 55) = 1.3$. There was a significant main effect of polarity, $F(1, 55) = 9.48$, $p < .01$. Rephrasing from rules with affirmative first components (90.2%) was better than rephrasing from rules with negative first components (79.1%).

A significant Original Rule $\times$ Familiarity interaction, $F(1, 55) = 4.11$, $p < .05$, was found. The facilitatory effects of familiarity were greater in rephrasing from disjunctives (95.1% familiar, 73.5% unfamiliar) than from conditionals (89.9% familiar, 80.0% unfamiliar). There was a significant Familiarity $\times$ Polarity interaction, $F(1, 55) = 14.34$, $p < .01$. The effect of polarity was greater for rules with an unfamiliar content than for those with a familiar content. There was a significant three-way Original Rule $\times$ Causality $\times$ Polarity interaction, $F(1, 55) = 5.96$, $p < .05$. The effect of rule polarity on rephrasing from initial disjunctives was greater when the rule content was non-causal rather than causal.

TABLE 6
Percentages of Correct Rephrasings Obtained in Experiment 2

			Causal		*Non-causal*	
Rephrasing	*Familiarity*	*Polarity*	*%*	*Mean*	*%*	*Mean*
conditional to disjunctive	familiar	P/Q	96.7		87.5	
		– P/Q	87.9	92.3	87.5	87.5
	unfamiliar	P/Q	90.9		83.3	
		– P/Q	66.7	78.8	79.2	81.25
disjunctive to conditional	familiar	P/Q	96.7		91.7	
		– P/Q	100.0	98.4	91.7	91.7
	unfamiliar	P/Q	90.9		83.3	
		– P/Q	69.7	80.3	50.0	66.7

Discussion

Inspection of Tables 4 and 6 reveals that, as predicted, performance at rephrasing between *P only if Q* and disjunctives in Experiment 2 was better than performance at rephrasing between *If P then Q* and disjunctives in Experiment 1. This is true for all rule contents and polarities, with two exceptions out of 16 cross-experiment comparisons. These results are consistent with mental models theory, which states that whereas a single model is represented in the initial model set of *If P then Q* rules, two models are represented in the initial model set of *P only if Q* rules, thus facilitating rephrasings between this form and its disjunctive equivalents. The two exceptions do not detract from this finding, as they also involve causal and/or familiar content, where the initial model sets of *If P then Q* and *P only if Q* forms are not determined solely by rule syntax.

Also as predicted, the asymmetry between rule forms in rephrasing performance found in Experiment 1 was not found in Experiment 2. This supports the mental models account of the differential representation in initial model sets of *If P then Q* having only one model and *P only if Q*, which, like disjunctives, contains two models and thus places the same load on working memory as disjunctives.

The effects of familiarity that were found in Experiment 1 were replicated in Experiment 2: Familiar content had a greater facilitatory effect in rephrasing from disjunctives than from conditionals. However, the effects of causality upon rephrasing from initial conditionals that were found in Experiment 1 were not present in Experiment 2. The mental models account of causality suggests that the effect of causal content is to add a model of the counterfactual to the initial model set of a conditional. As illustrated above, the initial model set for an abstract ... *only if* ... conditional contains what is effectively a model representing the counterfactual case. Thus the initial model sets formed for causal and non-causal ... *only if* ... conditionals will be the same. Consequently, causality has no effect on rephrasings between ... *only if* ... conditionals and disjunctives.

GENERAL DISCUSSION

The following results were obtained in Experiments 1 and 2:

- In both experiments, performance was below ceiling but above floor—a result that suggests there is no simple psychological equivalence that parallels the logical equivalence of conditionals and disjunctives.
- In Experiment 1, rephrasings from *If ... then* conditionals into disjunctives were correct more often than rephrasings in the opposite direction. However, no asymmetry was observed in Experiment 2 in rephrasings between ... *only if* ... conditionals and disjunctives. These results support a Minimal Completion strategy for generating rephrasings from partially completed initial model sets.
- Familiar content was shown to facilitate rephrasing performance in both experiments, though the interactions show that this effect was significant only when rephrasing from disjunctives into conditionals, and not vice versa.

- Causal content was shown to facilitate rephrasings between *If ... then* conditionals and disjunctives in Experiment 1, particularly when the initial rule was an *If ... then* conditional. However, causality did not facilitate rephrasings between ... *only if* ... conditionals and disjunctives in Experiment 2.
- More correct rephrasings were made from rules with positive first components than from rules with negative first components. This effect was particularly strong with rephrasings from disjunctives that had unfamiliar content.
- There was no evidence in the results of Experiment 1 for an advantage in rephrasing from *If ... then* rules with negative antecedents. Similar effects of polarity were found with rephrasings from ... *only if* ... rules in Experiment 2.

The asymmetry in rephrasing from each rule form that was observed in Experiment 1 but was absent in Experiment 2 is consistent with the predictions deriving from mental models theory about the number of models in the initial model set for each rule syntax. The *If ... then* conditional has only one model in the initial set, whereas the disjunctive has two models. Thus, it appears that the lower working memory load of the *If ... then* conditional facilitates rephrasings.

The effects of familiarity and causality can be seen as secondary additions to the initial model set created in representing the rule syntax. As a consequence, the effects that rule syntax have on model set size are masked when thematic content leads to the creation of complete initial model sets.

The effects of causality were greatest when rephrasing from *If ... then* conditionals. Like Johnson-Laird and Byrne (1991), we argue that the presence of causal content leads to the explicit representation of the counterfactual in the initial model set. Thus, causal content adds a new model to the initial model set created in representing the rule syntax. The fact that causality has less effect on rephrasing from disjunctives or from ... *only if* ... conditionals can be explained simply because a partial representation of the counterfactual model already exists in the initial model set of these rule syntaxes. Thus, our data appear to confirm the account of causality effects for disjunctives derived from Johnson-Laird and Byrne's (1991) account of conditional reasoning.

Turning to the facilitatory effect of familiarity, this was seen only for rephrasing from disjunctives, and not for rephrasing from *If ... then* conditionals. As a consequence, the account of familiarity effects offered in the introduction (i.e. the suggestion that familiarity completes the initial model sets for both conditionals and disjunctives) is not entirely substantiated by our data. It may be that familiarity operates only to fill in missing terms in existing models and cannot add further models that are not only already in the initial model set through representation of the rule syntax. Two, albeit incomplete, models emerge through the representation of a disjunctive's syntax. These can then be completed in the presence of familiar information to enable an exclusive interpretation of the rule. For example, given the rule "Either it is a flamingo or it is not pink", representation of the syntax alone gives the following incomplete model set:

flamingo

¬ pink

The effect of familiar content will be to add the missing components of the existing models in the initial model set (i.e. subjects know that flamingos are pink and that things that are not pink cannot be flamingos):

[flamingo]	[pink]
[¬ flamingo]	[¬ pink]

On the other hand, given the conditional "If it is a flamingo then it is pink", representation of the syntax alone gives only a single model in the initial model set:

flamingo	pink
...	

The negative contingency is not represented in this initial model set, and the presence of familiar content about pink flamingos does not encourage the subject to consider things that are not flamingos or things that are not pink. Thus, familiar content does not add new models to the initial model set, but only completes partial models that are already represented.

The theory that causality adds models to an initial model set, whereas familiarity can only complete existing partial models, provides a revised account of how general and specific knowledge are used in the construction of mental models. Note that this account differs from the one offered in the introduction, in which causality and familiarity both added and completed models in the initial model sets of both conditionals and disjunctives. This extrapolation from Johnson-Laird and Byrne's (1991) account of effects of causality on conditionals now appears insufficient to explain syntactic effects on rephrasing in the presence of thematic content.

In the current experiments, the causality and familiarity of the stimuli were assessed only by category (causal versus non-causal; familiar versus unfamiliar). However, it may be that effects occur only with specific *degrees* of causality and familiarity. For example, Johnson-Laird and Byrne (1991) suggest that one can distinguish three different degrees of causality, each giving rise to a different initial model set, as shown in Table 7. One might predict that only strongly causal relations will affect rephrasing as only they have the effect of adding an additional model to the initial model set. We are currently conducting an experiment to test this prediction.

An alternative characterization is offered by Cummins (1995; see also Cummins, Lubart, Alksnis, & Rist, 1991), who argues that causal deduction is sensitive to the existence of possible alternative causes and disabling conditions. For example, the rule "If the vase hadn't been dropped then it wouldn't have broken" does not preclude alternative causes of the vase being broken, such as the vase being shot, overheated, and so on. Similarly, it is possible to conceive of dropping the vase (e.g. onto a soft cushion) without it breaking, which represents a disabling condition in which the effect is prevented from occurring despite the given cause. Cummins et al. (1991) found that subjects were more inclined to accept arguments based upon causal conditions when there were few alternative causes and disabling conditions. Cummins (1995) has also shown that these effects are enhanced by familiar content, in that the more familiar the causal

TABLE 7
Examples of Varying Degrees of Causality[a]

Causality	*Example*	*Initial Model Set*	
strong	If the vase hadn't been dropped then it wouldn't have broken	d	b
		¬ d	¬ b
enabling	If the vase hadn't been fragile then it wouldn't have broken	f	b
		¬ f	¬ b
		f	¬ b
weak	If the vase hadn't been touched then it might not have broken	t	b
		¬ t	¬ b
		¬ t	b

[a] Adapted from Johnson-Laird & Byrne, 1991, pp. 70–71.

relationship (and hence the more alternative causes and disabling conditions that were available in a subject's experience), the less willing were subjects to accept conclusions based upon it.

Cummins' theory of causal deduction appears, on the face of it, to contradict the account of causality and familiarity offered in this paper. To recast her observations in mental models terms, the presence of *familiar* alternative causes or disabling conditions appears to add new models to the model set used in evaluating conclusions, whereas we have argued that the role of familiarity is only to complete partially represented models. However, it should be noted that the inference evaluation task used by Cummins (1995) and Cummins et al. (1991) entails the second stage of fleshing-out a model set, whereas we have argued that the rephrasing task entails only the construction of an initial model set. An examination of rephrasing performance with causal rules that vary in terms of available alternative causes or disabling conditions would provide a test of these competing accounts. If causal rules that permit alternative causes or disabling conditions are rephrased differently from those that do not, then this may undermine our proposed account of familiarity and causality effects. If, on the other hand, there is no effect, then this would suggest that the effects of alternative causes or disabling conditions are reserved for the subsequent fleshing-out of model sets and that causality is initially processed as a general relationship whose initial representation is determined by rule syntax.

Perhaps the most challenging result from our experiments is the absence of support for the prediction that the antecedent of the conditional *If not P then Q* is represented exhaustively, unlike other conditionals. There seems to be no a priori reason arising from the procedural semantics of mental models theory (Johnson-Laird, 1983) for representing negative antecedents exhaustively in an initial model set. Furthermore, in the account offered here, the exhaustive representation of negated rule components plays no part in rephrasing performance and is not needed to explain any effects of rule syntax, polarity, or thematic content. Thus, we believe it is important to formulate a version of mental models theory that does not include this assumption. Furthermore, as outlined earlier, Johnson-Laird and Byrne's (1991) explanation of matching bias depends upon the

truth of this assumption. If mental models theory is to be retained, an alternative account of matching bias must be found.

To summarize, on the basis of two studies reported in this paper we offer a mental models account of rephrasing performance that is based upon a Minimal Completion strategy. Under this account, a subject forms an initial model set representing the rule to be rephrased; he or she then fleshes this out only as far as is necessary to represent both rule components that are necessary for a rephrasing, and finally generates a rephrasing from this partially completed model set. The formation of model sets and their subsequent effect upon the generation of rephrasings is shown to be determined by the following factors:

- Rephrasing performance is poor with unfamiliar and non-causal stimuli, because subjects must base their rephrasings upon an incomplete initial model set.
- Where a rephrasing is based on an incomplete initial model set, the number of models in that set affects performance, more models leading to poorer performance.
- Causal content acts to add a counterfactual model to the initial model set. A disjunctive model set already contains a partially represented counterfactual, and so the effect of causal content is greater for conditionals than for disjunctives.
- Familiar content acts to complete the models that are partially represented in an initial model set but does not add new models. The initial model set for a disjunctive contains two incomplete models, whereas the initial model set for a conditional contains one complete model. Familiar content therefore has a greater effect on rephrasing from disjunctives.
- No evidence was found for the Stoics' assumption concerning conditionals with negative antecedents, which undermines Johnson-Laird and Byrne's (1991) account of matching bias.

It should be pointed out that a degree of caution is needed in accepting these conclusions. In particular, in both experiments only one item was used per condition. This problem has been addressed by Clark (1973), who refers to it as the "language-as-fixed-effect fallacy". Clark's basic argument is that if only a few items are used to investigate a linguistic hypothesis, then there is no guarantee that the findings obtained will generalize to a new set of materials. Thus, until the two experiments reported in the paper are replicated with new materials, there is always a chance that the large effects of rule form and content seen may simply be due to the items used rather than to their form or content. Nonetheless, we believe that our account of rephrasing performance with conditionals and disjunctives, and also the integration of familiar and causal content into initial model sets, provide a plausible and testable extension of mental models theory.

REFERENCES

Cheng, P.W., & Holyoak, K.J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*, 391–416.

Clark, H.H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359.

Cummins, D.D. (1995). Naive theories and causal deduction. *Memory and Cognition*, *23*, 646–658.
Cummins, D.D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory and Cognition*, *19*, 274–282.
Evans, J.St.B.T. (1972). Interpretation and matching bias in a reasoning task. *Quarterly Journal of Experimental Psychology*, *24*, 193–199.
Evans, J.St.B.T. (1977). Linguistic factors in reasoning. *Quarterly Journal of Experimental Psychology*, *29A*, 297–306.
Evans, J.St.B.T. (1983). Linguistic determinants of bias in conditional reasoning. *Quarterly Journal of Experimental Psychology*, *35A*, 635–644.
Evans, J.St.B.T. (1989). *Bias in human reasoning: Causes and consequences*. Hove, UK: Lawrence Erlbaum Associates Ltd.
Evans, J.St.B.T. (1993). On rules, models and understanding. *Behavioral and Brain Sciences*, *16*, 345–346.
Evans, J.St.B.T., & Lynch, J.S. (1973). Matching bias in the selection task. *British Journal of Psychology*, *64*, 391–397.
Evans, J.St.B.T., & Newstead, S.E. (1977). Language and reasoning: A study of temporal factors. *Cognition*, *5*, 265–283.
Evans, J.St.B.T., Newstead, S.E., & Byrne, R.M.J. (1993). *Human reasoning: The psychology of deduction*. Hove: Lawrence Erlbaum Associates Ltd.
Fillenbaum, S. (1974). Or: Some uses. *Journal of Experimental Psychology*, *103(5)*, 913–921.
Fillenbaum, S. (1975). If: Some uses. *Psychological Research*, *37*, 245–260.
Fillenbaum, S. (1976). Inducements: On the phrasing and logic of conditional promises, threats and warnings. *Psychological Research*, *38*, 231–250.
Girotto, V., Mazzocco, A., & Cherubini, P. (1992). Judgements of deontic relevance in reasoning: A reply to Jackson and Griggs. *Quarterly Journal of Experimental Psychology*, *45A*, 547–574.
Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge: Cambridge University Press.
Johnson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction*. Hove: 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., & Tagart, J. (1969). How implication is understood. *American Journal of Psychology*, *82*, 367–373.
Legrenzi, P., Girotto, V., & Johnson-Laird, P.N. (1993). Focussing in reasoning and decision making. *Cognition*, *49*, 37–66.
Marcus, S.L., & Rips, L.J. (1979). Conditional reasoning. *Journal of Verbal Learning and Verbal Behaviour*, *18*, 199–223.
Markovits, H. (1986). Familiarity effects in conditional reasoning. *Journal of Educational Psychology*, *78(6)*, 492–494.
Newstead, S.E., & Griggs, R.A. (1983). The language and thought of disjunction. In J.St.B.T. Evans (Ed.), *Thinking and reasoning: Psychological approaches*. London: Routledge & Kegan Paul.
Newstead, S.E., Griggs, R.A., & Chrostowski, J.J. (1984). Reasoning with realistic disjunctives. *Quarterly Journal of Experimental Psychology*, *36A*, 611–627.
Ormerod, T.C. (1988). *Cognitive processes in programming*. Unpublished Ph.D. Thesis, Sunderland.
Ormerod, T.C., Manktelow, K.I., & Jones, G.V. (1993). Reasoning with three types of conditional: Biases and mental models. *Quarterly Journal of Experimental Psychology*, *46A(4)*, 653–677.
Shepherd, A., & Ormerod, T.C. (1992). *Development of a formal method of user requirements specification for process plant displays*. Report. Loughborough University.
Wason, P.C. (1966). Reasoning. In B.M. Foss (Ed.), *New horizons in psychology*, *Vol 1*. Harmondsworth: Penguin.
Wason, P.C., & Johnson-Laird, P.N. (1972). *Psychology of reasoning*. London: Batsford.

Original manuscript received 2 September 1994
Accepted revision received 18 August 1996

APPENDIX 1

Rules Used in Experiment 1

Causal familiar rules

If the milk is left out of the fridge then it will go off.
If it is raining then the ground will not be dry.
If there is not a hole in the hull then the boat will float.
If it is not winter then the squirrels will not be hibernating.

Either you drink a bottle of whisky or you will stay sober.
Either the bomb explodes or the people will not die.
Either you do not wash your clothes or they will be clean.
Either a person does not have a fever or their temperature will not be normal.

Causal unfamiliar rules

If the ethanol passes through a separator then chlorate will be removed.
If you add Butene then the density will not fall.
If the hopper is not faulty then the viewing window will be black.
If the Redler conveyor does not stop then the pulper will not slow down.

Either the hypersorber contains hydrogen or the channel will emit benzene.
Either the isopropanol rises or the diaphragm will not open.
Either the temperature is not above – 20C or the viscosity will be too low.
Either the fan is not broken or the vapour pressure will not rise.

Non-causal familiar rules

If it is a satsuma then it is orange.
If it is a flamingo then it is not grey.
If it is not a plant then it moves.
If it is not a horse then it does not have hooves.

Either it is a frog or it walks.
Either it is a foodstuff or it cannot be eaten.
Either it is not a mother or it is female.
Either it is not a bird or it does not have scales.

Non-causal unfamiliar rules

If a solid contains chloride then it absorbs water.
If vinyl chloride is compressed then it does not remain gaseous.
If the solid is not barium then its atomic number is 40.
If a pipe is not made of alkathene then it does not carry acid.

Either the gas is ammonia or its schmidt number is 0.8.
Either the solid is calcium or it does not thermally dissociate.
Either the elevator is not spiral-type or it needs water.
Either it is not iodide or its viscosity is not 14.

APPENDIX 2

Subject Instructions in Experiment 1

The following experiment is an investigation into the ways in which people rephrase sentences. I am particularly interested in rephrasing between sentences that take "disjunctive" and "conditional" forms. A disjunctive sentence has the form *Either P or Q*, where *P* and *Q* are two statements. The sentence Either *we will go swimming* or *we will go shopping* is an example of a disjunctive sentence. A conditional sentence takes the form *If P then Q*. The sentence If *you do your homework* then *you can go out* is an example of a conditional sentence. The main point to remember is that a disjunctive links two statements with *Either ... or ...* and a conditional links them with *If ... then ...* .

You will be required to work though 16 sentences. Each sentence that you will be given will be either a disjunctive or a conditional. Your task is to rephrase the sentence from the form in which it is given into the alternative form. So if you are given a disjunctive sentence then you must rewrite it as a conditional and vice versa. You should write the rephrasing that you produce in the space provided below the original sentence.

The most important point is that you should rewrite the sentence so that it still has the **same meaning** even though it is expressed in a different form. The rephrasing that you produce should not violate the meaning of the original sentence: the two should be consistent.

You should work through one sentence at a time. Once you have completed a sentence you may not go back and alter your answer. You should work through the sentences at your own pace, attempting as much as possible in the time available. Do not spend too long on any one sentence (approx. 1 minute maximum on each).

If there are any sentences that you feel it is impossible to rephrase then write "none" below them to signify that there are no possible rephrasings in your opinion.

Please note that while many of the terms in the sentences will be familiar to you there will be some that you will not know. You should not let this worry you but try and answer the questions in the way that you believe a person familiar with the terms would.

If you have any questions at any stage during the completion of this experiment please indicate this to the experimenter and s/he will come and assist you.