Mental Models and Rule Rephrasing

Juliet Richardson
Department of Psychology
Lancaster University
Lancaster, LA1 4YF,
United Kingdom.
J.Richardson@uk.ac.lancaster

Thomas C. Ormerod
Department of Psychology
Lancaster University
Lancaster, LA1 4YF,
United Kingdom.
T.Ormerod@uk.ac.lancaster

Abstract

An experiment is reported which uses a rephrasing task to investigate factors affecting the formation of initial mental models. It was found that both the syntax and the thematic content of the rule affect the initial model set[1] formed: the syntax determines the form of the initial model set and the semantics add to this initial set through the representation of subjects' prior knowledge about the situation in question. Specifically, causal content invokes general knowledge about causal relationships which leads to the addition of models representing counterfactual situations in the initial model set. In comparison, familiar content invokes specific knowledge which leads to the completion of existing models in the initial set. Thus, our experiment enables an extension of mental models to be made that accounts for the differential effects of general and specific prior knowledge.

Introduction

The mental models theory developed by Johnson-Laird (1983) has been used to provide an account of many of the phenomena observed in human deductive reasoning (see Evans, Newstead and Byrne, 1993; Johnson-Laird and Byrne, 1991) and language more generally (see Johnson-Laird, 1983). However a number of authors (e.g. Evans, 1993) suggest that the failure to provide an extensive account of the effects of varying problem content is a weakness in the theory. If mental models theory is to provide a comprehensive account of reasoning and linguistic performance then this problem must be addressed.

According to mental models theory, in deductive reasoning tasks the subject forms models representing each premise. The information contained in these initial models is then combined to form one model from which conclusions can be drawn. In such tasks factors affecting the formation of models and the drawing of conclusions are confounded. In order to study in isolation the effects of rule content on model formation a task is required in which the subject does not have to manipulate premise information to draw a conclusion. Rephrasing between logically equivalent linguistic forms (e.g. Cheng and Holyoak, 1985; Fillenbaum, 1975, 1976; Ormerod, Manktelow and Jones, 1993) is one such task. For example, a subject may be given the rule "If it is raining then the ground is not dry" and asked to produce a rephrasing in the form *Either...or...* such as "Either it is raining or the ground is dry". If subjects use an initial model set representing the first rule to produce a rephrasing of it then their accuracy at rephrasing rules will reflect this initial model formation. By presenting rules with different rule contents to be rephrased the effect of rule content on initial mental model formation can be investigated.

The experiment reported in this paper examines the psychological treatment of conditional (*If...then...*) and disjunctive (*Either...or...*) forms in a rephrasing task. This particular rephrasing has not been previously reported in the literature although there is a large body of research investigating conditional and disjunctive reasoning separately. Mental models theory already provides accounts of disjunctive and conditional reasoning with arbitrary contents; the rephrasing task provides data that can be used to extend these accounts to include reasoning with realistic contents.

Logical Equivalence of Conditionals and Disjunctives

A conditional of the form *If P then Q* can be interpreted in two ways: 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, a disjunctive of the form *Either P or Q* can be interpreted in two ways: it can be exclusively interpreted as *Either P or Q but not both*, or inclusively interpreted as *Either P or Q or both*. If truth tables for the two forms are compared (see Johnson-Laird and Byrne, 1991, pp 7-8) then the logical equivalence of biconditionals and exclusive disjunctives and of implications and inclusive disjunctives can be observed.

[1] We use the term 'model set' to describe the set of one or more individual 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 representation of premise information. Like Johnson-Laird and Byrne (1991) we distinguish between initial model sets and fleshed-out model sets as two stages in model formation. At each stage the model sets may be complete (i.e. having an exhaustive representation of all contingencies) or incomplete.

A Mental Models Account of Conditional and Disjunctive Reasoning

Johnson-Laird and Byrne's (1991) mental models theory provides a plausible psychological account of rephrasing. The theory proposes that subjects construct an initial, possibly incomplete, model set of the given premises. This set is then fleshed out if necessary to provide a complete model set and is used to formulate and evaluate possible inferences.

For example, given the premise "If the letter is A then the number is 2" an initial model set as shown below will be formed[2]:-

A	2
. . .	

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 as shown below:-

[A]	[2]	Biconditional
[¬ A]	[¬ 2]	

[A]	[2]	Implication
[¬ A]	[2]	
[¬ A]	[¬ 2]	

The biconditional model set supports the conclusion "The number is not 2" whilst the implication model set supports no unique conclusion.

A similar account is provided for disjunctive reasoning. For example, given the premise "Either the letter is A or the number is not 2" a set of two initial models as shown below will be formed:-

A	
	¬ 2

If the subject is 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:-

[A]	[2]	Exclusive
[¬ A]	[¬ 2]	

[A]	[2]	Inclusive
[A]	[¬ 2]	
[¬ A]	[¬ 2]	

The exclusive model set supports the conclusion "The number is not 2" whilst the inclusive model set supports no unique conclusion.

The *initial* model sets formed for logically equivalent conditionals and disjunctives differ. Conditionals have a single model in which both components are represented, whilst disjunctives have two models each representing one of the components (see also Johnson-Laird, Byrne and Schaeken, 1994, Table 1, p424). Thus if a subject forms an initially unfleshed out model then s/he will not be able to use this to produce a rephrasing without fleshing it out.

However the *fleshed out* model sets for logically equivalent conditionals and disjunctives are equivalent. For example, the fleshed out model set for a biconditional interpretation of "If the letter is A then the number is 2" is the same as the model sets for exclusive interpretations of "Either the letter is not A or the number is not 2" and "Either the letter is A or the number is not 2". Whereas if an implicational interpretation is made, then the fleshed out model set formed is only identical to the model set for an inclusive interpretation of "Either the letter is not A or the number is 2" (see Johnson-Laird and Byrne, 1991, pp43-51). Thus only fully fleshed out mental model sets support the production of rephrasings and even then the number of possible correct rephrasings is affected by the particular interpretation made.

Effects of Prior Knowledge on Rule Interpretation

The truth-table analysis and mental models theory show how the logical and psychological equivalence of conditionals and disjunctives depends upon their interpretation. The literature indicates two factors that affect the interpretation of conditionals: familiarity and causality.

Markovits (1986) found that unfamiliar conditionals were more likely to be interpreted as biconditionals. He argued that for familiar conditionals subjects more easily generate examples of the consequent occurring without the antecedent. Therefore, they are more likely to interpret a familiar conditional as an implication and an unfamiliar conditional as a biconditional. It seems that prior knowledge of *specific* instances discourages a biconditional interpretation. Marcus & Rips (1979) found that biconditional interpretations were more likely for causal rather than non-causal conditionals. A similar explanation to that proposed by Markovits can also account for this finding, invoking *general* rather than specific knowledge. People have general knowledge about causal relationships, notably that the consequent does not usually occur in the absence of the antecedent in causal events. Thus, general knowledge about causality directs the subject to form a biconditional interpretation.

Several factors affect the interpretation of disjunctives. For example, Newstead, Griggs, & Chrostowski (1984)

[2] Johnson-Laird and Byrne (1991) provide a notation for describing mental models in which ¬ denotes a negated component, ··· denotes the possibility of further models and [ ] denotes the exhaustive representation of one contingency with respect to another within the model set.

found that altering the context led to different interpretations of disjunctives. For example, a threat context led to more exclusive interpretations than a qualification context. It is likely that knowledge of specific and general cases directs the interpretation of disjunctives in a similar way to conditionals.

Mental Models and Thematic Content

If Johnson-Laird and Byrne's (1991) mental models theory is to provide a comprehensive account then it must explain the effects of thematic content. Although the existing mental models account of the effects of content on reasoning is incomplete (see Evans, 1993) it does provide an account of the effects of causal content.

For conditionals expressing causal relationships, Johnson-Laird and Byrne state that "general knowledge informs the choice of what to represent in the models" (1991, p70). They propose that the subject forms an initial model set representing the actual and counterfactual situations associated with the causal relationship. This is consistent with the proposal that causal content evokes *general* prior knowledge directing the subject to a biconditional interpretation. 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 model set is built accordingly:-

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

Although Johnson-Laird and Byrne do not explicitly describe how familiar content (i.e. *specific* prior knowledge) affects the mental models formed, one might propose a similar account to that for causality. If 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. S/he will be unlikely to consider this contingency, and will 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 an assumption, following from the subject's 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 consistent with a biconditional interpretation:-

[blue]	[10]
[¬ blue]	[¬ 10]

Thus, Markovits' findings concerning unfamiliar content and Marcus & Rips' 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 model sets, whereas if the content is unfamiliar it is the absence of *specific* prior knowledge that influences the model set formed. In both cases, subjects build initial model sets that can only be fleshed out in a way consistent with a biconditional interpretation.

This analysis suggests that familiarity and causality affect the formation of initial model sets rather than their subsequent fleshing-out. It also seems probable that the initial model sets determine performance in a rephrasing task. Thus, manipulating the familiarity and causality of rule content is likely to have a large effect on rephrasing performance.

The Experiment

This experiment investigated effects of familiarity and causality on rephrasing between conditionals and disjunctives. Subjects received a task in which they were required to rephrase a given conditional into a disjunctive or vice versa.

The initial model sets for disjunctives and conditionals are not equivalent whereas the fleshed out sets are. Thus if an incomplete initial model set is formed it will need to be fleshed out in order to produce a rephrasing. Fleshing out of mental model sets places a cognitive load on the subject and can lead to errors. Thus rephrasing performance should be worse when the rule content is unfamiliar or non-causal and the initial model sets need to be fleshed out.

Another prediction about rephrasing performance also emerges from the effect of content on initial model set formation. One of the principles of mental models theory is that the more models that are formed the greater the load on working memory, with consequent reductions in reasoning performance. If one initial model is formed for a conditional and two initial models are formed for a disjunctive, then one can predict an asymmetry in rephrasing performance: rephrasing will be harder from a disjunctive into a conditional than when the rephrasing is in the opposite direction. This asymmetry should be observed when the initial model set is not fleshed out, in other words 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 rephrase either causal rules or non-causal rules. The other factors were all within-subjects. The first of these was familiarity: subjects received both familiar (everyday situations) and unfamiliar (chemical processes) rules to rephrase. The rules were rated for familiarity and causality by two independent judges, whose judgements agreed 100% with our own. 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 components. Therefore, in total, each subject rephrased sixteen rules.

Procedure. The experiment was presented in a booklet containing instructions followed by the sixteen trials, one per page, in a different randomised order for each subject.

Subjects wrote their rephrasings in the booklet at their own pace, but were unable to change answers in the light of subsequent rephrasings. The duration of the experiment was 20 minutes.

Results and Discussion

Subjects' attempted rephrasings were judged correct if they could be judged logically equivalent to the original rule according to any possible interpretation of that original rule. Thus for each rule two possible forms of rephrasing were allowed. Also, implicit as well as explicit negatives were allowed, for example "in" was allowed in place of "not out". Table 1 summarises the data obtained from this experiment in terms of average percentage of correct rephrasings for each original rule and content. These data were subjected to an analysis of variance to test for main effects of and interactions between original rule, familiarity and causality.

Original Rule There was a significant main effect of original rule, $F(1, 69)=4.22$, $p<0.05$. Performance was better when rephrasing from a conditional (76.7% correct) than from a disjunctive (72.1% correct). Rephrasing from a disjunctive is harder than from a conditional because the two models represented in the initial model set for a disjunctive place a greater load on working memory than the single model for a conditional. As predicted this effect was greatest for non-causal and unfamiliar content, in other words, when the initial model was predicted to be incomplete.

Familiarity. There was a significant main effect of familiarity, $F(1, 69)=34.37$, $p<0.001$. Rephrasing from familiar rules (81.5% correct) was significantly better than rephrasing from unfamiliar rules (67.2% correct). The presence of unfamiliar content leads to the formation of incomplete initial model sets, which makes subsequent rephrasing harder because the subject must attempt to flesh out the model set. Fleshing out model sets increases the processing load faced by subjects, thereby increasing the likelihood of errors. Thus rephrasing performance is worse in the presence of unfamiliar content when the initial mental model is incomplete and must be fleshed out.

There was also a significant two-way interaction between original rule and familiarity, $F(1, 69)=65.44$, $p<0.001$. 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 (74.7% correct - familiar, 78.8% correct - unfamiliar).

Familiar rule content should invoke specific knowledge which causes secondary additions to the initial models set

Table 1: Mean percentages of correct rephrasings obtained in the experiment.

Initial Rule	Familiarity	Causality	Example Rule	Mean % Correct	Model Set
Conditional	Familiar	Causal	*If the milk is left out of the fridge then it will go off*	85.6	[out] [off] [¬ out] [¬ off]
		Non-causal	*If it is a satsuma then it is orange*	63.7	satsuma orange . . .
	Unfamiliar	Causal	*If the ethanol passes through a separator then chlorate will be removed*	85.0	[pass] [remove] [¬ pass] [¬ remove]
		Non-causal	*If a solid contains chloride then it absorbs water*	72.6	[chloride] [absorbs] . . .
Disjunctive	Familiar	Causal	*Either you drink a bottle of whisky or you will stay sober*	93.8	[drink] [¬ sober] [¬ drink] [sober]
		Non-causal	*Either it is a frog or it walks*	83.1	[frog] [¬ walks] [¬ frog] [walks]
	Unfamiliar	Causal	*Either the hypersorber contains hydrogen or the channel will emit benzene*	66.9	[contain] [¬ emit] [¬ contain] [emit]
		Non-causal	*Either the gas is ammonia or its schmidt number is 0.8*	44.4	ammonia 0.8

formed by the representation of the rule syntax. Two, albeit incomplete, models emerge through the representation of a disjunctive's syntax. For example, given the rule "Either it is a flamingo or it is not pink", representation of the syntax alone gives the following incomplete initial 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 have specific knowledge 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 models to the initial model set, but only completes partial models that are already represented.

Causality. There was a significant main effect of causality, $F(1, 69)=16.02$, $p<0.001$. Rephrasing from causal rules (82.8% correct) was significantly better than rephrasing from non-causal rules (66.0% correct). The presence of non-causal content leads to the formation of incomplete initial model sets making subsequent rephrasing harder because of the need to flesh out the model set. This effect of causality was strongest when the original rule was a conditional, with a smaller effect when it was a disjunctive. Johnson-Laird and Byrne (1991) argue that the presence of causal content leads to the explicit representation of counterfactual contingencies in the initial model set of conditionals. Thus, the representation of causal content creates new models in the initial model set for a conditional, enabling its completion. The smaller effect of causality on disjunctives may simply be because a partial representation of a counterfactual model already exists in the representation of the disjunctive syntax.

Summary

Johnson-Laird and Byrne's (1991) mental models theory appears to provide a parsimonious and coherent account of rephrasing performance. Two factors contribute to an initial model set: representing the syntactic form of the original rule, and representing the thematic content embodied in the rule. Effects of familiarity and causality can be explained as secondary additions to the initial model set created by representing the rule syntax. Table 1 shows the mental models formed for different combinations of rule form and rule content. The effect of causal content is, as suggested by Johnson-Laird and Byrne (1991), to add an additional model representing the counterfactual contingency to the initial model set of a conditional. When the initial rule is a disjunctive a partial counterfactual model is included in the representation of the rule syntax. Therefore causal content does not add extra models to the initial model set for a disjunctive. The effect of familiar content is to complete the partial models within the initial model set representing the rule syntax. Familiar content does not, in itself, add new models to the initial model set.

Acknowledgements

The first author was funded by a research studentship awarded by British Gas plc. The experiment was conducted whilst the authors were at the Department of Human Sciences, Loughborough University. We thank Linden Ball and Jeremy Miles for providing assistance with statistics and commenting on drafts.

References

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

Evans, J. S. B. T. (1993). On rules, models and understanding. *Behavioral and Brain Sciences,* 16, 345-346.

Evans, J.S.B.T., Newstead, S.E. and Byrne, R.M.J. (1993). *Human reasoning: the psychology of deduction.* Hove: Lawrence Erlbaum Associates.

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.

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

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

Johnson-Laird, P.N., Byrne, R.M.J. and Schaeken, W. (1994). Why models rather than rules give a better account of propositional reasoning: a reply to Bonatti and to O'Brien, Braine and Yang. *Psychological Review,* 101(4), 734-739.

Marcus, S.L. and 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. and Chrostowski, J.J. (1984). Reasoning with realistic disjunctives. *Quarterly Journal of Experimental Psychology,* 36A, 611-627.

Ormerod, T.C., Manktelow, K.I. and Jones, G.V. (1993). Reasoning with three types of conditional: biases and mental models. *Quarterly Journal of Experimental Psychology,* 46A(4), 653-677.