INVITED ADDRESS

Reasoning by rule or model ?

P.N. Johnson-Laird
MRC Applied Psychology Unit, Cambridge, England

INTRODUCTION

Human reasoning depends on three principal skills: comprehension, the generation of conclusions, and the evaluation of conclusions. The critical step in a deduction is the evaluation of a conclusion in order to ensure it is valid. Some theorists propose that this process is akin to a formal derivation in a logical calculus (see e.g. Inhelder and Piaget, 1958; Braine, 1978; Rips, 1983). Other theorists propose that it depends on a search for alternative interpretations, or *models*, of the premises that serve as refutations (e.g. Newell, 1981; Johnson-Laird, 1983). No evidence so far has been universally accepted as counting decisively against one school of thought or the other. My aim in this paper is to try to settle the issue by considering a class of inferences not hitherto investigated experimentally -- multiply-quantified deductions.

An example of a multiply-quantified assertion is: None of the artists is taller than any of the beekeepers. Such assertions contain a relational expression -- here, "taller than" -- and its arguments are quantified using such expressions as "all", "some", "none". These quantifiers behave in ways that are similar to those of the first-order predicate calculus (though there are others that do not). To explain how people reason with multiple quantifiers, it is necessary first to account for how they reason with relational expressions. I will describe a theory of relational reasoning and then some evidence from a crucial experiment carried out to compare this theory and a rule-based theory. Next, I will present a theory of reasoning with multiple quantifiers and evidence from a second experiment designed to decide between this theory and a rule-based theory. Finally, I will outline the essential theoretical difference between the two sorts of theory. All this work was carried out in collaboration with Ruth Byrne.

RELATIONAL REASONING

Most psychological studies of relational reasoning have concerned so-called three-term series problems, such as: Anne is taller than Betty; Carol is shorter than Betty; who is tallest? The evidence from this domain has not sufficed to decide between rule-based and model-based theories (see e.g. Huttenlocher, 1968; Clark, 1969). The validity of these inferences depends on the transitivity of the relation in the premises, and theorists have generally assumed that there is either a general schema for transitivity or content-specific rules, such as:

> if x is taller than y and y is taller than z, then x is taller than z.

The theory based on the manipulation of mental models proposes that the logical properties of relations, such as transitivity, are not explicitly represented at all, but are emergent from the *meanings* of relations (Johnson-Laird, 1983). The meaning of an assertion such as:

The number of artists is greater than the number of beekeepers

enables the interpretative system to construct a model of the situation, e.g.

artist	beekeeper
artist	beekeeper
artist	beekeeper
artist	

and to verify that the relation holds within models. The meaning of "greater than" can be formulated in terms of the concepts of an empty set, a non-empty set, and the addition and subtraction of items from sets:

The number of x's is greater than the number of y's: if there is at least one x and there are no y's or (x - 1) is greater than (y - 1).

There are other ways of giving a recursive definition of the concept, and such definitions can be couched in a form that is suitable for the construction and manipulation of models such as the one above. Hence, given the further assertion:

The number of beekeepers is greater than the number of chemists

the interpretative system can construct the model:

artist	beekeeper	chemist
artist	beekeeper	chemist
artist	beekeeper	
artist		

which supports the conclusion that the number of artists is greater than the number of chemists. To test validity, it is necessary to search for an alternative model of the premises that refutes the conclusion. There are various ways in which the search could be made, and some have been modelled computationally (Johnson-Laird and Bara, 1984). But, here I will not make any strong claims about the procedures that people use, other than that they do not possess any simple deterministic algorithm for searching for refutations (cf. Newell and Simon, 1972). Hence, where a correct response can be made only as a result of considering more than one model, the theory predicts that the task will be reliably harder -- a prediction that has been confirmed in many studies of traditional syllogisms, i.e. arguments that depend on singly-quantified premises (see e.g.

Johnson-Laird and Bara, 1984). We can predict that when people reason about relations, the task should be harder where there is a genuine choice of models of the same premises.

Ruth Byrne and I tested this prediction by examining such descriptions as:

The jug is on the right of the cup
The plate is on the left of the cup
The knife is in front of the plate
The fork is in front of the jug.

The description corresponds to a single determinate model:

plate	cup	jug
knife		fork

and so it should be relatively easy to answer a question about the relation between the knife and the fork: The knife is on the left of the fork. But, when the second premise is instead:

The plate is on the left of the jug

the description is consistent with at least two distinct models:

plate	cup	jug		cup	plate	jug
knife		fork		knife	fork	

The same relation holds between the knife and the fork in either layout, but the model-based theory predicts that the task should be harder because both models ought to be constructed in testing the validity of the answer. The task should be still harder where the correct response can be made only by constructing both models. The following description:

The jug is on the right of the cup
The plate is on the left of the jug
The knife is in front of the plate
The fork is in front of the cup

is consistent with two distinct models:

plate	cup	jug		cup	plate	jug
knife	fork			fork	knife	

that have no relation in common between the knife and fork, and so there is no valid conclusion. Granted that working memory has a limited processing capacity, the model-based theory predicts the following rank order of increasing difficulty: one-model problems, multiple-model problems with valid conclusions, and multiple-model problems with no valid conclusions.

A formal rule for making such inferences needs to be of the following sort: If x is related to y on one dimension, and w is related to x on an orthogonal dimension, and z has the same orthogonal relation to y, then w is related to z in the same way as x is related to y. The rule can be applied to the configuration:

```
x ------- y
|         |
w         z
```

in any orientation. Moreover, it can be directly applied to the premises yielding a multiple-model problem with a valid conclusion (see above) because the relation between x and y is directly asserted by the premise: The plate is on the left of the jug. But there is no such premise in the one-model problem: the relevant relation has to be inferred (using a further rule of inference) before the present rule can be applied. Hence, the rule-based theory predicts that one-model problems should be harder than multiple-model problems with valid concludions -- exactly the opposite prediction to the model-based one.

We have carried out a series of experiments in order to compare the predictions of the two theories. In our latest experiment, 18 adult subjects acted as their own controls and carried out four inferences of each of the three sorts. The percentages of their correct responses were: 70% for one-model problems, 46% for the multiple-model problems with valid conclusions, and 8% for the multiple-model problems with no valid conclusion. This reliable trend corroborates the model-based theory but runs counter to the theory based on rules.

MULTIPLY-QUANTIFIED REASONING

Granted the following definition of simple consanguineal relationships:

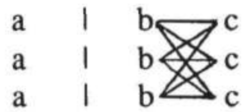
> x is related to y if x is a parent of y (or x is a child of y) or x is a parent of z (or x is a child of z) and z is related to y

it is an emergent property that the relation is symmetric and transitive. Thus, the following multiply-quantified premises:

> None of the artists is related to any of the beekeepers.
> All of the beekeepers are related to all of the chemists.

yield the valid conclusion: None of the artists is related to any of the chemists.

The model-based theory assumes that reasoners construct a model of the state of affairs described by the premises:

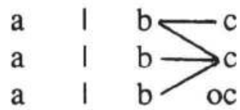


where the vertical barrier represents negation and prevents any relation between the artists and beekeepers, and each link corresponds to a relation between two individuals. It is impossible to construct any model of the premises that refutes the conclusion.

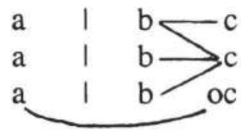
The following premises, which differ only in a single quantifier:

None of the artists is related to any of the beekeepers.
All of the beekeepers are related to some of the chemists.

yield a similar model:



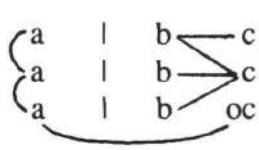
from which the same conclusion as before can be drawn. But in this case a search for a counterexample will be successful:



The two models support only the conclusion:

Some of the artists are not related to any of the chemists.

but again this conclusion can be refuted by a further model:



The only valid conclusions are:

Some of the chemists are not related to any of the artists.
None of the artists is related to *some* of the chemists.

or weaker ones, e.g. the ambiguous assertion, None of the artists are related to all of the chemists. Plainly, the model-based theory predicts that the one-model inference should be easier than the multiple-model inference.

Although there are rule-based theories of reasoning with quantifiers (e.g. Braine and Rumain, 1983), there is no current theory that is powerful enough to derive multiply-quantified conclusions. However, the two sorts of inference can be derived in the first-order predicate calculus. Both proofs are long -- at least fifteen lines of derivation -- but they are remarkably similar: they require the same number of steps, and the rules that are used at each step are identical except in two cases, and even here there are only minor differences. Since the number of steps in a derivation is taken to predict the psychological difficulty of an inference (see e.g. Rips, 1983), the rule-based theory does not predict any difference between the two sorts of problem.

Ruth Byrne, Patrizia Tabossi, and I have carried out a series of experiments to investigate multiply-quantified reasoning in which the subjects drew their own conclusions in their own words. In one of these experiments, we tested 11 adults with a series of problems including the two described above and other similar ones. They made 77% correct responses for the one-model problems, but only 23% correct responses for the multiple-model problems -- a difference that was highly significant. The only divergence between the rule-based derivations of the two sorts of problem concerns the elimination and subsequent re-introduction of quantifiers: the one-model problem has these steps for the quantifier "all" where the multiple-model problem has them for "some". Could there be an intrinsic difference in the difficulty of the two classes of rules? Another sort of one-model problem in the experiment called for the use of the rules for "some", e.g.: Some of the beekeepers are related to all the artists; None of the chemists are related to any of the beekeepers. This problem was relatively easy with 67% correct conclusions. Hence, there is no intrinsic difficulty in dealing with "some".

CONCLUSIONS

The model-based theory extends naturally to reasoning with premises containing multiply-quantified relations. Rule-based theories, however, appear neither to predict the relative difficulty of inferences or the typical errors that reasoners make. A fundamental difference between the two approaches concerns the treatment of variables. Theories based on formal rules propose that reasoners, having gone to the trouble of understanding the premises, base their inferences on something other than a full semantic interpretation. They are supposed to abstract the underlying logical skeleton -- so-called "logical form" -- and then exploit formal rules of instantiation in order to replace quantified variables by *single* hypothetical individuals that stand in for them; after a stage of formal reasoning about sentential connectives, quantifiers can then be restored by formal rules of generalization operating on these hypothetical individuals. According to the model-based theory, however, the work of instantiation is merely part of the normal process of comprehension: universally quantified phrases are instantiated by sets of mental tokens that are treated as exhaustively representing the relevant set; existential quantified phrases are similarly instantiated by sets of mental tokens, except that there are optional items that do not satisfy the conditions of the assertion. One of the consequences of this semantic distinction is that the choice of quantifier in one premise can affect the number of possible models of the premises as a whole.

Of course it may be possible to devise a theory based on formal rules that will account for our experimental results -- rule-based theories in general have universal Turing machine power. Nevertheless, the lack of any currently feasible rule-based theory may not be accidental. The theory of mental models is, by comparison, relatively simple to refute. It predicts that whenever the meaning of premises supports more than one conclusion about what is possible though not necessary, there are multiple models of the premises, and so the inferential task will be harder. This prediction has now withstood empirical testing for three domains of inference: syllogisms, relational reasoning, and reasoning with multiple quantifiers.

REFERENCES

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

Braine, M.D.S., and Rumain, B. (1983) Logical reasoning. In Flavell, J.H., and Markman, E.M. (Eds.) *Handbook of Child Psychology, Vol. III, Cognitive Development.* 4th Ed. New York: Wiley.

Clark, H.H. (1969) Linguistic processes in deductive reasoning. *Psychological Review*, 76, 387-404.

Huttenlocher, J. (1968) Constructing spatial images: A strategy in reasoning. *Psychological Review*, 75, 550-560.

Inhelder, B., and Piaget, J. (1958) *The Growth of Logical Thinking from Childhood to Adolescence.* London: Routledge & Kegan Paul.

Johnson-Laird, P.N. (1983) *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness.* Cambridge: Cambridge University Press. Cambridge, MA: Harvard University Press.

Johnson-Laird, P.N., and Bara, B.G. (1984) Syllogistic inference. *Cognition*, 16, 1-61.

Newell, A. (1981) Reasoning, problem solving and decision processes: the problem space as a fundamental category. In Nickerson, R. (Ed.), *Attention and Performance, Vol. 8.* Hillsdale, NJ: Erlbaum.

Newell, A. and Simon, H. (1972) *Human Problem Solving.* Englewood Cliffs, NJ: Prentice-Hall.

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