THE FORM AND FUNCTION OF MENTAL MODELS

P. N. Johnson-Laird

Centre for Research on Perception and Cognition
Laboratory of Experimental Psychology
University of Sussex
Brighton BN1 9QG England

You are lost in the maze at Hampton Court Palace. You come to a turning and for a moment you are not sure which way to go. You recognize that you have been at this point before, and, in your imagination, you turn right, proceed down an alley, and are then confronted by a dead end. And so this time around, you decide to turn left. What you did was to reconstruct a route through the maze on the basis of a mental model of it. You may hardly have experienced any imagery at all; or you may have had a succession of vivid images like a snippet from an imaginary movie that culminates in a leafy cul-de-sac. In either case, there was nothing verbal about your reasoning: you navigated your way through your model of the maze much as a rat in a psychological laboratory might have done (O'Keefe and Nadel, 1978). Yet, there is another method that you could use to make your decision. You recall instead that the way to get out of the maze is to keep turning left at every available opportunity, and, since you are presented with such an opportunity, you accordingly decide to turn left. This method makes use of a mental representation of verbal propositions.

The two alternatives illustrate the contrast between exploiting a mental model (perhaps with accompanying imagery) and making use of a propositional representation. My aim in this paper is to show that the contrast is real -- that there are both forms of mental representation -- and to offer an account of the purpose that they serve. Indeed, if there are mental models, then the two most important questions about them are: what form do they take? what function do they serve? I will try to answer both questions.

Direct empirical evidence for the contrast between propositional representations and mental models comes from a series of experiments that Kannan Mani and I have carried out (Mani and Johnson-Laird, in press). In the most recent of our studies, the subjects heard a verbal description of a spatial layout, such as:

The spoon is to the left of the knife
The plate is to the right of the knife
The fork is in front of the spoon
The cup is in front of the knife.

They were then shown a diagram, such as:

spoon knife plate

fork cup

and they had to decide whether or not the diagram was consistent with the description. (If you think of the diagram as depicting the arrangement of the objects on a table top, then obviously it is consistent with the description.) Half the descriptions that the subjects received were determinate like the example above, and the other half were indeterminate. The indeterminate descriptions were constructed merely by changing the last word in the second sentence:

The spoon is to the left of the knife
The plate is to the right of the spoon
The fork is in front of the spoon
The cup is in front of the knife.

This description is consistent with two radically different diagrams:

(1)			(2)		
spoon	knife	plate	spoon	plate	knife
fork	cup		fork		cup

The materials were counterbalanced so that for each set of five objects, a subject received either the determinate or else the indeterminate description. After the subjects had judged a series of eight descriptions and diagrams, they were given an unexpected test of their memory for the descriptions. On each trial, they had to rank four alternatives in terms of their resemblance to the original description: the original description, an inferrable description, and two 'foils' with a different meaning. The inferrable description for the example above contained the sentence:

The fork is to the left of the cup

in place of the sentence interrelating the spoon and the knife. The description can therefore be inferred from the layout corresponding to the original description in the case of both the determinate and the indeterminate descriptions.

The subjects remembered the layouts of the determinate descriptions very much better than those of the indeterminate descriptions. The percentages of trials on which they ranked the original and the inferrable descriptions prior to the foils was 88% for the determinate descriptions, but only 58% for the indeterminate descriptions. All twenty of the subjects conformed to the trend, and there was no effect of whether or not a diagram had been consistent with a description. However, the percentages of trials on which the original description was ranked higher than the inferrable description was 68% for the determinate descriptions, but 88% for the indeterminate descriptions. This difference was highly reliable, too.

Evidently, subjects tend to remember the layout of determinate descriptions better than that of

indeterminate descriptions, but they tend to remember verbatim detail of indeterminate descriptions better than that of determinate descriptions. This 'cross-over' effect is impossible to explain without postulating at least two sorts of mental representation. A plausible account of the results is indeed that subjects construct a mental model of the determinate descriptions but abandon such a representation in favour of a superficial propositional one as soon as they encounter an indeterminacy in a description. Mental models are easier to remember than propositional representations, perhaps because they are more structured and elaborated (cf. Craik and Tulving, 1975) and require a greater amount of processing to construct (cf. Johnson-Laird and Bethell-Fox, 1978). But, models encode little or nothing of the linguistic form of the sentences on which they are based, and subjects accordingly confuse inferrable descriptions with the originals. Propositional representations are relatively hard to remember, but they do encode the linguistic form of sentences. Hence, when they are remembered, the subjects are likely to make a better than chance recognition of verbatim content.

It is natural to suppose that propositional representations are produced as part of the normal process of comprehending discourse (cf. Kintsch, 1974; Fodor, Fodor, and Garrett, 1975), but they can also serve the useful purpose of providing an economical representation of radically indeterminate discourse. The function of mental models is profoundly semantic: a propositional representation is true or false with respect to a mental model of the world. This relation is established by mapping propositional representations onto mental models, and I have argued elsewhere for a procedural semantics that carries out this task (see, e.g. Johnson-Laird, 1980, for a description of a program that builds up spatial arrays from verbal descriptions). Truth or falsity with respect to reality ultimately depends on the construction of mental models on the basis of perceptual experience.

There is one other crucial function served by mental models. The fundamental semantic principle of truth is that an assertion is true provided that there is no counterexample to it. The assertion, "Socrates is dead," has only one possible counterexample, namely, that Socrates is not dead; the assertion, "All men are mortal," has a large number of potential counterexamples. Likewise, given the truth of a set of premises, a conclusion is necessarily true only if there is no counterexample to it, that is, no way of interpreting the premises that renders the conclusion false. If human beings have grasped this principle, then they can reason validly without possessing any mental logic, rules of inference, or inferential schemata. It is a straightforward matter to write computer programs that make inferences without recourse to rules of inference: they construct models of the premises, draw a putative conclusion on the basis of a simple heuristic, and then search for counterexamples to the conclusion. On the previous occasion that I presented this idea (Johnson-Laird, 1980), it was viewed as on a par with the Pelagian heresy in some quarters. Yet cognitive scientists should be prepared to accept that the doctrine of mental logic may be just as mistaken as the idea of original sin. Abandoning the doctrine certainly solves the otherwise intractable mystery of how children could acquire logic without being able to reason validly. The thesis that logic is innate is the only plausible solution but it has no more explanatory value or empirical content than an appeal to divine intervention. If there is no mental logic, then the question of of its origins does not even arise. The theory of mental models also reveals the major cause of inferential error: the greater the number of mental models that have to be constructed in order to make a valid deduction, the greater load on working memory, and the more likely an error is to be made. My colleagues and I have checked this prediction in a variety of inferential tasks. Table 1 presents the relevant data from four of our experiments in which the subjects had to draw their own conclusions from syllogistic premises: it gives the percentages of valid conclusions that were drawn depending on the number of mental models that had to be

Table 1: The percentages of correct valid conclusions drawn from syllogistic premises in four experiments. The percentages are shown as a function of the number of mental models that have to be constructed in order to draw a valid conclusion.

	One model problems	Two model problems	Three model problems
Experiment 1	92	46	28
Experiment 2	80	20	9
Experiment 3	62	20	3
Experiment 4	58	0	0

constructed. In Experiment 1, 20 students at Teachers College, Columbia University, were asked to state what followed from premises in each of the 64 logically distinct varieties (see Johnson-Laird and Steedman, 1978). Experiment 2 was a replication with 20 students at Milan University, and Experiment 3 was a further replication in which 20 Italian subjects were given just 10 seconds in which to make each of their responses. These experiments were carried out in collaboration with Bruno Bara. Finally, in Experiment 4, which Debbie Bull and I designed, 19 children between 11 and 12 years of age were asked to draw conclusions from 20 out of the 64 possible pairs of syllogistic premises. The trend in each experiment was remarkable: not a single subject that we have tested has ever failed to perform best on those syllogisms that require only a single model to be constructed. It is difficult to resist the conclusion that inferential ability is based on the manipulation of mental models.

Let me finish with one final conjectural flourish. The psychological core of understanding any phenomenon consists in your having a 'working model' of it in your mind. If you understand inflation, a mathematical proof, the way a computer works, DNA or a divorce, then you have a mental representation of it that serves as a model in much the same way as, say, a clock functions as a model of the solar system. Like a clock, a mental model need not be wholly accurate to be useful, which is just as well because, of course, there are no complete models of any empirical phenomena. If a television set is mentally represented as containing a beam of electrons that are magnetically deflected across the screen, then this component of the model serves an explanatory function. It accounts, for example, for the distortion of the picture that occurs when a magnet is held near to the screen. Other components of the model may serve no such function. One might imagine, say, each electron as deflected by the magnetic field much as a ball-bearing is diverted from its course by a magnet, but without having any representation of the nature of magnetism: the 'picture' is just a picture, which simulates reality rather than models its underlying principles. At least one other component of every dynamic model is

neither modelled nor simulated. This element is time. Time is not represented in a dynamic model, but rather the model unwinds in real time in much the same way as do the events that are modelled, though perhaps at a different rate. In models that are not dynamic, of course, time can be represented by a spatial axis. What one should expect in examining the growth of expertise in a particular domain is the gradual transition from mere propositional principles to a fully articulated mental model, and the gradual replacement of simulated elements by their modelled counterparts.

REFERENCES

Craik, F.I.M., & Tulving, E. Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General, 1975, 104, 268-294.

Fodor, J.D., Fodor, J.A., & Garrett, M.F. The psychological unreality of semantic representations. Linguistic Inquiry, 1975, 4, 515-531.

Johnson-Laird, P.N. Mental models in cognitive science. Cognitive Science, 1980, 4, 71-115.

Johnson-Laird, P.N., & Bethell-Fox, C.E. Memory for questions and amount of processing. Memory and Cognition, 1978, 6, 496-501.

Johnson-Laird, P.N., & Steedman, M.J. The psychology of syllogisms. Cognitive Psychology, 1978, 10, 64-99.

Kintsch, W. The representation of meaning in memory. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Mani, K., & Johnson-Laird, P.N. The mental representation of spatial descriptions. Memory and Cognition, In press.

O'Keefe, J., & Nadel, L. The hippocampus as a cognitive map. London: Oxford University Press, 1978.