65 Mental Models, Deductive Reasoning, and the Brain

PHILIP N. JOHNSON-LAIRD

ABSTRACT This chapter considers the two main approaches to deductive thinking: theories based on formal rules of inference postulate that deduction is a syntactic process akin to a logical proof; the mental model theory postulates that it is a semantic process akin to the search for counterexamples. Experimental evidence bears out the predictions of the model theory: the more models needed for a deduction, the harder it is; erroneous conclusions are consistent with the premises; and general knowledge affects the process of search. Recent neurological evidence bears out, as the model theory predicts, a significant involvement of the right hemisphere in reasoning.

> If deduction is a purely verbal process then
> it will not be affected by damage to the
> right hemisphere.
> It *is* affected by such damage.
> It is not a purely verbal process.

This argument is an example of a valid deduction: Its conclusion must be true if its premises are true. (They may not be, of course.) Deductive reasoning is under intensive investigation by cognitive scientists, and more is known about it than about any other variety of thinking. The aim of this chapter is to explain its nature and to relate it to the brain. "The cerebral organization of thinking has no history whatsoever," Luria remarked (1973, 323); and Fodor (1983, 119) suggested that nothing can be known about the topic, because thinking does not depend on separate "informationally encapsulated" modules (but cf. Shallice, 1988, 271). Many regions of the brain are likely to underlie it, but as we shall see, a start has been made on the neuropsychology of reasoning.

Many cognitive scientists have argued that deductive reasoning depends on formal rules of inference like those of a logical calculus, and that these unconscious rules are used to derive conclusions from the representations of premises. These "propositional" representations are syntactically structured strings of symbols in a mental language, and the chain of deductive steps is supposedly analogous to a logical proof (see, e.g., the theories of Braine, Reiser, and Rumain, 1984; Osherson, 1974-1976; Rips, 1983). An alternative account postulates a central role for mental models. This account does not reject propositional representations, but it treats them as the input to a process that constructs a mental model corresponding to the situation described by the verbal discourse. The process of deduction—as well as induction and creation (Johnson-Laird, 1993)—is carried out on such models rather than on propositional representations. Models are the natural way in which the human mind constructs reality, conceives alternatives to it, and searches out the consequences of assumptions. They are, as Craik (1943) proposed, the medium of thought. But what *is* a mental model?

The underlying idea is that the understanding of discourse leads to a model of the relevant situation akin to one created by perceiving or imagining events instead of merely being told about them (Johnson-Laird, 1970). Experimental studies have indeed found evidence for both initial propositional representations and mental models (see e.g., Johnson-Laird, 1983; van Dijk and Kintsch, 1983; Garnham, 1987). The same idea has led to the model theory of deductive reasoning. The theory was not cut from whole cloth, but was gradually extended from one domain to another. From a logical standpoint, there are at least four main domains of deduction:

1. Relational inferences based on the logical properties of such relations as *greater than*, *on the right of*, and *after*.
2. Propositional inferences based on negation and on such connectives as *if*, *or*, and *and*.

PHILIP N. JOHNSON-LAIRD Department of Psychology, Princeton University, Princeton, N.J.

3. Syllogisms based on pairs of premises that each contain a single quantifier, such as *all* or *some*.
4. Multiply quantified inferences based on premises containing more than one quantifier, such as *Some pictures by Turner are more valuable than any by any other English painter.*

Logicians have formalized a predicate calculus that covers all four domains and includes the propositional calculus, which deals with inferences based on connectives. The model theory was developed first for relational inferences and syllogisms, and recently for propositional and multiply quantified inferences. In contrast, psychological theories based on formal rules exist for relational and propositional inferences, but not for syllogisms or for multiply quantified inferences.

Theories and evidence have been reviewed in detail elsewhere (Johnson-Laird and Byrne, 1991, 1993; Holyoak and Spellman, 1993). In this chapter, we will stand back from the details and present an integrated account of mental models based on all of this work. We will also bring the story up to date and relate it to the neuropsychology of thinking. The chapter begins with relational inferences and establishes that a model-based system does not require postulates specifying the logical properties of relations. It then shows how models can underlie reasoning with sentential connectives, such as *or,* and quantifiers, such as *all.* Next, it shows how certain sorts of diagrams inspired by the model theory can help reasoners to cope with disjunctions. Finally, it considers the neuropsychological findings, and draws some conclusions about the assumptions underlying mental models.

Relational inferences and emergent logical properties

Consider the following simple inference:

> The Turner painting is on the right of the Daumier.
> The Corot sketch is on the left of the Daumier.
> What follows?

A valid answer is that the Turner painting is on the right of the Corot sketch. Psychological theories based on formal rules of inference (e.g., Hagert, 1984; Ohlsson, 1984) explain the derivation of the answer in terms of a formal proof. It depends on the logical properties of the relations: *on the left of* is the converse of *on the right of,* and both are transitive relations. These properties have to be added to the premises by stating them in so-called meaning postulates, that is, postulates that depend on the meanings of these relations:

> For any x, y, if x is on the left of y, then y is on the right of x.
> For any x, y, z, if x is on the right of y, and y is on the right of z, then x is on the right of z.

With these postulates the conclusion can be derived:

> The Turner painting is on the right of the Corot,

using various rules of inference, including *modus ponens*:

$$\begin{array}{l} \text{if } p \text{ then } q \\ p \\ \therefore q \end{array}$$

where p and q denote any propositions whatsoever. The formal derivation for this simple inference is surprisingly long: It calls for eight steps, but that is the price to be paid for using formal rules.

The theory of mental models takes a different approach. It treats propositional representations as instructions for the construction of models. The meaning of, say, *on the right of* consists in the appropriate increments to the Cartesian coordinates of one object y, in order to locate another object, x, so that: x is on the right of y. Hence, the propositional representation of the assertion:

> The Turner painting is on the right of the Daumier

can be used to construct a spatial model:

$$d \qquad t$$

where d denotes the Daumier and t denotes the Turner. The information in the second premise:

> The Corot sketch is on the left of the Daumier

can be added to yield:

$$c \qquad d \qquad t$$

This model supports the conclusion:

> The Turner painting is on the right of the Corot sketch.

The conclusion is true in the model, but does it follow validly from the premises? The crucial manipulation to test validity is to search for alternative models of the premises that refute the conclusion. In fact, there are no alternative models of the premises in which the conclusion is false, and so it is valid. The model-based method of reasoning accordingly has no need of meaning postulates or formal rules of inference. The logical properties of a relation, such as its transitivity, are not explicitly represented at, all, but emerge from the meaning of the relation when it is put to use in the construction of models. The general procedure of searching for alternative models is used to test validity.

The evidence from three-term series problems, such as the example above, does not suffice to decide between formal rules and mental models. However, studies of two-dimensional spatial reasoning have produced more decisive data (Byrne and Johnson-Laird, 1989). We examined problems of the the following sort:

The cup is on the right of the saucer.
The plate is on the left of the saucer.
The fork is in front of plate.
The spoon is in front of the cup.
What is the relation between the fork and the spoon?

Subjects tend to imagine symmetrical arrangements, and so the description corresponds to a single model:

plate	saucer	cup
fork		spoon

It should be relatively easy to answer that the fork is on the left of the spoon. When the second premise of the problem is changed to

The plate is on the left of the cup

the resulting premises are consistent with at least two distinct models:

plate	saucer	cup		saucer	plate	cup
fork		spoon			fork	spoon

The same relation holds between the fork and the spoon in both models, but the theory predicts that the task should be harder because both models must be constructed in order to test the validity of the answer. The task should be still harder where the correct response can be made only by constructing both models. The description

The cup is on the right of the saucer
The plate is on the left of the cup
The fork is in front of plate
The spoon is in front of saucer

is consistent with two distinct models:

plate	saucer	cup		saucer	plate	cup
fork	spoon			spoon	fork	

that have no relation in common between the fork and the spoon, and so there is no valid answer to the question. Granted that the mind has a limited processing capacity, the model theory predicts the following rank order of increasing difficulty: one-model problems, multiple-model problems with valid answers, and multiple-model problems with no valid answers.

Formal-rule theories need complex meaning postulates to support two-dimensional deductions (Hagert, 1984; Ohlsson, 1984). Whatever rules a theory uses, however, the one-model problem calls for a longer derivation than the multiple-model problem with a valid answer. It is necessary to infer the relation between the plate and the cup for the one-model problem, but there is no need for such a derivation with the multiple-model problem because the relation is directly asserted by the second premise:

The plate is on the left of the cup.

Hence, formal rule theories predict that the one-model problems should be harder than the multiple-model valid problems, which is exactly the opposite prediction to the one made by the model theory.

Our experiments compared the predictions of the two theories (Byrne and Johnson-Laird, 1989). In one experiment, 18 adults carried out four inferences of each of the three sorts, and the percentages of their correct responses were as follows: 70% for the one-model problems, 46% for the multiple-model valid problems, and 8% for the multiple-model problems with no valid conclusion. This robust trend corroborates the model theory but runs counter to the formal-rule theories. The same results have been obtained from analogous problems concerning temporal relations (Schaeken and Johnson-Laird, 1993). Subjects also drew correct conclusions to one-model problems reliably faster than to multiple-model problems.

Models for connectives and quantifiers

What remains to be accounted for are the logical constants—sentential connectives and quantifiers. Some psychological theories postulate formal rules of inference for connectives, but no such theories exist for quantifiers. The model theory, however, proposes an account for both. Connectives call for models of alternative possibilities. A conjunction of the form:

p and q

requires only a single model:

where p and q respectively denote the situations described by the two propositions. But an exclusive disjunction such as:

p or else q, but not both

requires two alternative models, which are shown here on separate lines:

p
$\qquad q$

A conditional of the form:

If p, then q

calls—at least initially—for one explicit model (of the antecedent and consequent) and one implicit model of an alternative situation:

$p \quad q$
. . .

The implicit model symbolized by the three dots may subsequently be rendered explicit, but for many inferences the implicit model suffices. We have implemented a computer program (Propsych) that computes the numbers of explicit models required by inferences (see Johnson-Laird, Byrne, and Schaeken, 1992). Consider, for example, the following argument:

> Studies have shown that children of people who smoke more than two packs per day have a greater exposure than others to secondhand smoke or a lowered resistance to viral infection. Children exposed to secondhand smoke have an increased risk of lung cancer. Children with lowered resistance to viral infection are harder to treat with chemotherapy. These two factors make for intractable cases of lung cancer. Thus, these children risk contracting untreatable lung cancer.

The first step is to represent the underlying propositional connectives in the premises:

If child of smoker *then* (exposed to smoke *or* lowered resistance).
If exposed to smoke *then* greater risk.
If lowered resistance *then* chemotherapy harder.
If greater risk or chemotherapy harder *then* risk of untreatable cancer.
If child of smoker *then* risk of untreatable cancer.

We can then use the Propsych program to work out the total number of explicit models that have to be constructed to carry out the inference. Thus, the first premise calls for two explicit models and one implicit model:

$c \qquad s$
$c \quad l$
. . .

where c denotes a child of a smoker, s denotes exposure to smoke, and l denotes lowered resistance. The second premise calls for the following models:

$s \quad r$
. . .

where r denotes a greater risk. The principles for combining sets of models are simple: A new model is made, if possible, from each pairwise combination of a model from one set with a model from the other set, according to the principles in Johnson-Laird, Byrne, and Schaeken (1992, 425):

1. If the model in one set is implicit and the model in the other set is implicit, then the result is an implicit model.
2. If the model in one set is implicit but the model in the other set is not, then no new model is formed from them.
3. If the pair of models is inconsistent, that is, one contains the representation of a proposition and the other contains a representation of its negation, then no new model is formed from them.
4. Otherwise the two models are joined together, eliminating any redundancies.

The result of combining the sets of models for the first two premises is, accordingly:

$c \qquad s \quad r$
$c \quad l \quad s \quad r$

Hence, so far, the process of inference has called for the construction of five explicit models. The set of premises as a whole calls for the construction of nine explicit models. Initial models of this sort suffice for all the 61 direct inferences used in a study by Braine, Reiser, and Rumain (1984), and the program was used to count them: They predicted the difficulty of the problems as well as these authors' rule-based theory.

Although many inferences in daily life can be made with such models, sometimes one has to think more carefully and flesh out the models completely. Given the conditional

If there is a triangle then there is a circle,

could there be a circle without a triangle? Presumably so, given one interpretation of the conditional. Could there be a triangle without a circle? Of course not. That would contravene the meaning of the conditional. Hence, as soon as individuals begin to think more closely about the meaning of the conditional, they realize that the explicit model in the following set:

△ ○
. . .

represents the only possible situation in which a triangle can occur. That is, it must occur with a circle given the truth of the conditional. One way to represent this information is to use a special annotation:

[A] o
. . .

where the square brackets indicate that triangles have been exhaustively represented in relation to circles. The procedure for fleshing out models works as follows: When a proposition has been exhaustively represented, its negation is added to any other models; when a proposition has not been exhaustively represented, it and its negation form separate models that replace the implicit model (denoted by three dots). Triangles cannot occur in fleshing out the implicit model above, because they are already exhausted, but their negations can occur with either a circle or its negation. Hence, the result is:

A o
¬△ o
¬△ ¬○

where ¬ is an annotation representing negation. Because there is no longer any implicit model, there is no need for symbols representing exhaustive representations. Exhaustion is thus a device that allows the inferential system to represent certain information implicitly—it can be made explicit, but at the cost of fleshing out the models.

The same principles suffice for the representation of quantifiers. The interpretation of an assertion, such as

All the Frenchmen in the restaurant are gourmets

calls for a model of the following sort, in which each line no longer represents a separate model, but rather a separate individual in one and the same model:

[/] *g*
[*f*] *g*
. . .

where *f* denotes a Frenchman, *g* denotes a gourmet, and the three dots represent implicit individuals. As before, the square brackets indicate an exhaustive representation: The tokens denoting Frenchmen exhaust the set in relation to the set of gourmets. The set of gourmets, however, is not exhaustively represented. Hence, if the implicit individuals are fleshed out explicitly, some of them may be gourmets, but none of them can be Frenchmen unless they are also gourmets.

The information from a second premise, say

Some of the gourmets are wine drinkers

can be added to the model

[/] *g* *w*
[*f*] *g*
w
. . .

This model supports the believable conclusion:

Some of the Frenchmen in the restaurant are wine drinkers

This conclusion is erroneous, though it is drawn by most subjects (see Oakhill, Johnson-Laird, and Garnham, 1989). It is refuted by an alternative model of the premises:

[*f*] *g*
[*f*] *g*
g *w*
w
. . .

When the second premise is instead:

Some of the gourmets are Italians

the initial model supports the unbelievable conclusion:

Some of the Frenchmen are Italians

and hardly any subjects err now. In other words, reasoners tend to "satisfice" (see Simon, 1959): If they reach a congenial conclusion they tend not to search for alternative models. Satisficing is a frequent cause of everyday disasters, both major and minor. It seems an obvious danger, yet it cannot be predicted by rule theories, which contain no elements corresponding to models of situations.

The model theory generalizes to multiply quantified assertions. For example, the premises:

None of the Avon letters is in the same place as any of the Bury letters
All of the Bury letters are in the same place as all of the Caton letters

yield the valid conclusion:

> None of the Avon letters is in the same place
> as any of the Caton letters.

Granted the following definition of *in the same place as:*

> x is in the same place as y = x is in a place that
> has the same spatial coordinates as those for y

the premises support a model of the state of affairs:

| [*a*] [*a*] [*a*] | [*b*] [*b*] [*b*] [*c*] [*c*] [*c*] |

where the vertical barriers demarcate separate places, and there are arbitrary numbers of individuals of each sort (*a*s denote Avon letters, *b*s denote Bury letters, and *c*s denote Caton letters). This model yields the conclusion:

> None of the Avon letters is in the same place
> as any of the Caton letters.

No alternative model of the premises refutes the conclusion. As the theory predicts, one-model deductions are easier than multiple-model deductions (see Johnson-Laird, Byrne, and Tabossi, 1989).

Diagrams and disjunctions

Formal-rule theories predict that the difficulty of a deduction depends on the length of its derivation; the model theory predicts that it depends on the number of models that have to be constructed. Although some of these predictions run in parallel, there are interesting divergencies between them. According to the model theory, inferences based on exclusive disjunctions (two models) should be easier than inferences based on inclusive disjunctions (three models). Rule theories can accommodate this result by assuming that the rule for exclusive disjunction is easier to use than the rule for inclusive disjunction, but they cannot *predict* the phenomenon. The simplest prediction of the model theory, however, does not require any detailed account of numbers of models: Erroneous conclusions should tend to be consistent with the truth of the premises rather than inconsistent with them, because reasoners will often base their conclusions on only some of the possible models of the premises. Current theories based on formal rules of inference make no predictions about the nature of systematically erroneous conclusions.

Experiments in all the main domains of deduction have corroborated these two predictions of the model theory. Deductions that call for only a single model are reliably easier than those that call for multiple models; and erroneous conclusions tend to be consistent with the premises rather than inconsistent with them. We will illustrate this evidence with some recent studies of so-called double disjunctions (see Johnson-Laird, Byrne, and Schaeken, 1992; Bauer and Johnson-Laird, 1993).

If you wish to experience the phenomenon, then ask yourself what, if anything, follows from these double disjunctive premises:

> Raphael is in Tacoma or Jane is in Seattle, or both.
> Jane is in Seattle or Paul is in Philadelphia, or both.

Each premise supports three explicit models, and when the information from both premises is combined the result is five distinct models:

[*t*]	[*s*]	[*p*]
[*t*]	[*s*]	
	[*s*]	[*p*]
	[*s*]	
[*t*]		[*p*]

where *t* denotes Raphael in Tacoma, *s* denotes Jane in Seattle, and *p* denotes Paul in Philadelphia, though the actual models that people construct will probably represent particular individuals in particular cities. The models support the conclusion:

> Jane is in Seattle, or Raphael is in Tacoma and
> Paul is in Philadelphia.

As the model theory predicts, a double disjunction is reliably easier when the disjunctions are exclusive, e.g.:

> Raphael is in Tacoma or Jane is in Seattle, but not both.
> Jane is in Seattle or Paul is in Philadelphia, but not both.
> What follows?

because there are now only two possible models:

[*t*]		[*p*]
	[*s*]	

which support the conclusion:

> Jane is in Seattle, or Raphael is in Tacoma and
> Paul is in Philadelphia.

The problems are difficult, and most errors are consistent with the premises, that is, based on only some of their possible models, and typically on only a single model.

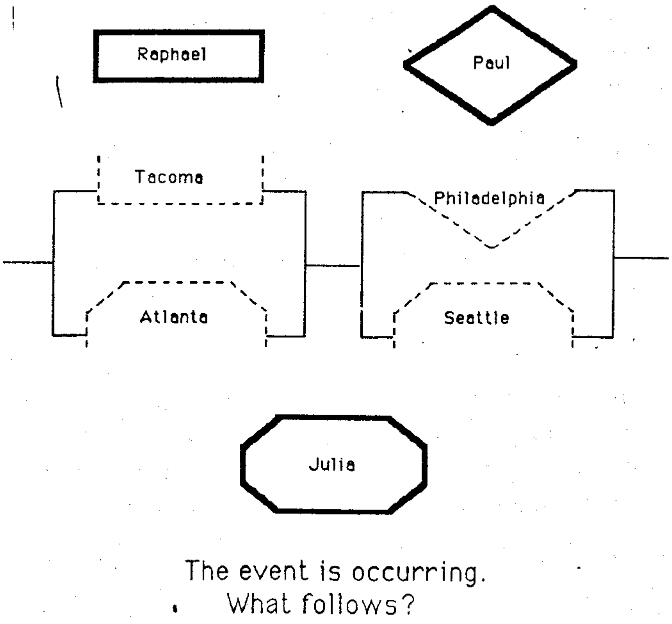

FIGURE 65.1 A diagram presenting a double disjunction problem: An event occurs only if the circuit diagram is completed from left to right by moving the shapes into their congruent positions in the circuit.

The psychological problem of deduction is to keep track of alternative possibilities. One way in which performance can be strikingly improved is to use diagrams rather than verbal premises. Not any sort of diagram will do, however. The evidence suggests that the diagram must use graphical means to make the alternative possibilities more explicit. With diagrams that resemble electrical circuits, such as figure 65.1, subjects drew 30% more valid conclusions than with the equivalent verbal problems (Bauer and Johnson-Laird, 1993).

Reasoning and the brain

The mental model theory is particularly pertinent to the neuropsychology of thinking. Unlike theories that stress that reasoning is a verbal process (Polk and Newell, 1992) or is governed by formal rules (e.g., Rips, 1983), it assumes that a major component of reasoning is nonverbal—that is, the construction of models with a structure corresponding to the structure of situations. Hence, the theory predicts that the right cerebral hemisphere should play a significant part in reasoning (Whitaker et al., 1991).

In general, neuropsychological evidence bears out this prediction. Several studies have shown that damage to the right hemisphere impairs patients' ability to make inferences. Thus, Caramazza et al. (1976) have shown that such patients have problems in deducing the converse of relations. They fail such problems as:

John is taller than Bill.
Who is shorter?

Similarly, Read (1981) found that they are impaired in comparison with normals with such three-terms series problems:

Arthur is taller than Bill.
Bill is taller than Charles.
Who is shortest?

These studies were motivated by the possibility that visual imagery underlies performance (but cf. McDonald and Wales, 1986) and the knowledge that visuospatial thinking appears to depend on the right hemisphere. However, reasoning can also be based on models that have no perceptible correlates. Individuals who are capable reasoners often report that they have not experienced visual imagery, and yet their performance is entirely consistent with the predictions of the model theory: They find multiple-model problems difficult, and their errors are consistent with premises. If the construction of models depends on the right hemisphere, then patients with right-hemisphere damage should find it just as hard to reason about abstract matters as to reason about topics that are easy to visualize. Some neurological studies have examined the ability to make inferences that do not depend on visuospatial thinking. Thus, given the sentences:

Sally approached the movie star with pen and paper in hand.
She was writing an article about famous people's views about nuclear power.

normal individuals are likely to infer that Sally wanted to ask the star about nuclear power. Patients with damage to the right hemisphere, as Brownell et al. (1986) observed, infer that Sally wanted the movie star's autograph. They are misled by the first sentence and cannot make the bridging inference from the second sentence to revise their interpretation. In general, it seems that right-hemisphere damage leads to an inability to get the point of a story, to make implicit inferences establishing coherence, to grasp the force of indirect illocutions such as requests (see e.g., Wapner, Hamby, and Gardner, 1981; Beaman, 1993), although at least one study failed to detect effects of right-hemisphere damage on implicit inferences (Tompkins and Mateer, 1985). What complicates matters is that damage to the right hemisphere can lead to semantic difficulties in the interpretation of words (see, e.g., Joanette and Brownell, 1990), and so in consequence the comprehension of discourse may also be impaired. Conversely, there is also evidence from split-brain patients that either hemisphere is capable of nonverbal reasoning (Zaidel, Zaidel, and Sperry, 1981), though the left hemisphere is superior to the right in problem solving (Gazzaniga, 1992, 103).

The strongest evidence for the model theory's predictions comes from experiments on conditional reasoning carried out by Whitaker and his colleagues. In a study of brain-damaged patients, Whitaker et al. (1991) examined conditional reasoning in two groups. The patients in both groups had undergone a unilateral anterior temporal lobectomy to relieve focal epilepsy, one group to the right hemisphere and the other group to the left hemisphere. Those with right-hemisphere damage were poorer at reasoning from false conditional premises than those with left-hemisphere damage. Thus, given the following conditional:

If it rained the streets will be dry

and the categorical assertion:

It rained,

the right-hemisphere-damaged group had a reliable tendency to conclude:

The streets will be wet.

In other words, these patients were unable to carry through the process of deduction in isolation from their knowledge of reality. In an ingenious study, Savary, Whitaker, and Markovits (1992), have extended this research to normal individuals. They argued that if reasoning depends on a major nonverbal component, then it should interfere more than a verbal memory task with a nonverbal secondary task. The primary task was either reasoning with a conditional problem or memorizing a sentence. While engaged in a primary task, the subjects had to judge whether two shapes were similar (the nonverbal secondary task) or decide whether a visually presented string of letters was a word or not (the verbal secondary task). The experimenters also obtained response times to the two secondary tasks when they were performed alone, and the key comparisons concerned the difference between these control measures and those obtained while the subjects were performing a primary task. The results confirmed the prediction: Reasoning, unlike memorizing a sentence, slowed down the judgments of the similarity of shapes, whereas there was no difference between the two in their effects on lexical decision.

The finding that certain sorts of diagrams can help reasoners (see the previous section) allows investigators to study reasoning without the need for verbal comprehension of premises. A major task for the future is to use brain-scanning techniques (see Kosslyn and Koenig, 1992) to investigate which areas of the brain are active during verbal and diagrammatic reasoning. The model theory predicts that both sorts of reasoning depend on the right hemisphere, and that diagrams should reduce the dependence on the left hemisphere.

Conclusions

The model theory is based on six main assumptions:

1. Entities are represented by tokens in models, their properties by properties of the tokens, and the relations between them by the relations between the tokens.
2. Alternative possibilities are represented by alternative models.
3. Negation is represented by a propositional annotation.
4. Implicit individuals and situations are represented by a propositional annotation that works in concert with an annotation indicating what has been represented exhaustively.
5. To account for counterfactual reasoning or reasoning about what is permissible, the epistemic status of a model can also be represented by a propositional annotation, such as: a model represents a real possibility, a counterfactual state of affairs, a permissible state of affairs, and so on.
6. Reasoning calls for the construction of models of premises, the formulation of conclusions based on them, and a search for alternative models to test validity.

Models based on the first two assumptions represent a class of situations (Barwise, 1993), and models that include propositional annotations can represent a finite set of alternative classes of situations.

The resulting theory makes three principal predictions. First, the greater the number of models called for to make an inference, the harder the task will be. Second, erroneous conclusions will tend to be consistent with the premises rather than inconsistent with them. Third, knowledge can influence the deductive process: Subjects will search more assiduously for alternative models when a putative conclusion is unbelievable than when it is believable. This chapter has illustrated the corroboration of these predictions in experiments from several domains of deduction. The model theory also makes a critical prediction about the role of the cerebral hemispheres in reasoning. As Whitaker et al. (1991) first pointed out, the construction of models is likely to depend on the right hemisphere. Although there is some evidence for this prediction, the crucial experiment has yet to be done. It calls for brain scanning during two sorts of reasoning, verbal and diagrammatic.

ACKNOWLEDGMENTS I am grateful for support from the James S. McDonnell Foundation. I thank Keith Holyoak and Nina Robin for useful criticisms of a draft of this chapter, and Malcolm Bauer and Ruth Byrne for help in carrying out much of the research.

REFERENCES

BARWISE, J., 1993. Everyday reasoning and logical inference. (Commentary on Johnson-Laird and Byrne, 1991) *Behav. Brain Sci.* 16:337-338.

BAUER, M. I., and P. N. JOHNSON-LAIRD, 1993. How diagrams can improve reasoning. *Psychol. Sci.* 4:372-378.

BEAMAN, M., 1993. Semantic processing in the right hemisphere may contribute to drawing inferences from discourse. *Brain Lang.* 44:80-120.

BRAINE, M. D. S., B. J. REISER, and B. RUMAIN, 1984. Some empirical justification for a theory of natural propositional logic. In *The Psychology of Learning and Motivation*, vol. 18. New York: Academic Press.

BROWNELL, H. H., H. H. POTTER, A. M. BIHRLE, and H. GARDNER, 1986. Inference deficits in right brain-damaged patients. *Brain Lang.* 27:310-321.

BYRNE, R. M. J., and P. N. JOHNSON-LAIRD, 1989. Spatial reasoning. *J. Mem. Lang.* 28:564-575.

CARAMAZZA, A., J. GORDON, E. B. ZURIF, and D. DE LUCA, 1976. Right hemispheric damage and verbal problem solving behavior. *Brain Lang.* 3:41-46.

CRAIK, K., 1943. *The Nature of Explanation.* Cambridge: Cambridge University Press.

FODOR, J. A., 1983. *The Modularity of Mind.* Cambridge, Mass.: MIT Press.

GARNHAM, A., 1987. *Mental Models as Representations of Discourse and Text.* Chichester: Ellis Horwood.

GAZZANIGA, M. S., 1992. *Nature's Mind: The Biological Roots of Thinking, Emotions, Sexuality, Language, and Intelligence.* New York: Basic Books.

HAGERT, G., 1984. Modeling mental models: experiments in cognitive modeling of spatial reasoning. In *Advances in Artificial Intelligence,* T. O'Shea, ed. Amsterdam: North-Holland.

HOLYOAK, K. J., and B. A. SPELLMAN, 1993. Thinking. *Ann. Rev. Psychol.* 44:265-315.

JOANETTE, Y., and H. H. BROWNELL, eds., 1990. *Discourse Ability and Brain Damage: Theoretical and Empirical Perspectives.* New York: Springer-Verlag.

JOHNSON-LAIRD, P. N., 1970. The perception and memory of sentences. In *New Horizons in Linguistics*, J. Lyons, ed. Harmondsworth, Middlesex, England: Penguin, pp. 261-270.

JOHNSON-LAIRD, P. N., 1983. *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness.* Cambridge: Cambridge University Press.

JOHNSON-LAIRD, P. N., 1993. *Human and Machine Thinking.* Hillsdale, N.J.: Erlbaum.

JOHNSON-LAIRD, P. N., and R. M. J. BYRNE, 1991. *Deduction.* Hillsdale, N.J.: Erlbaum.

JOHNSON-LAIRD, P. N., and R. M. J. BYRNE, 1993. Precis of *Deduction,* and authors' response: Mental models or formal rules? *Behav. Brain Sci.* 16:323–333, 368-376.

JOHNSON-LAIRD, P. N., R. M. J. BYRNE, and W. SCHAEKEN, 1992. Propositional reasoning by model. *Psychol. Rev.* 99: 418-439.

JOHNSON-LAIRD, P. N., R. M. J. BYRNE, and P. TABOSSI, 1989. Reasoning by model: The case of multiple quantification. *Psychol. Rev.* 96:658-673.

KOSSLYN, S. M., and O. KOENIG, 1992. *Wet Mind: The New Cognitive Neuroscience.* New York: The Free Press.

LURIA, A. R., 1973. *The Working Brain: An Introduction to Neuropsychology.* Harmondsworth, Middlesex, England: Penguin.

McDONALD, S., and R. WALES, 1986. An investigation of the ability to process inferences in language following right hemisphere brain damage. *Brain Lang.* 29:68-80.

OAKHILL, J. V., P. N. JOHNSON-LAIRD, and A. GARNHAM, 1989. Believability and syllogistic reasoning. *Cognition* 31: 117-140.

OHLSSON, S., 1984. Induced strategy shifts in spatial reasoning. *Ada Psychol.* 57:46-67.

OSHERSON, D. N., 1974-1976. *Logical Abilities in Children,* vols. 1–4. Hillsdale, N.J.: Erlbaum.

POLK, T. A., and A. NEWELL, 1992. A verbal reasoning theory for categorical syllogisms. Mimeo, Department of Computer Science, Carnegie Mellon University.

READ, D. E., 1981. Solving deductive reasoning problems after unilateral temporal lobotomy. *Brain Lang.* 12:116-127.

RIPS, L. J., 1983. Cognitive processes in propositional reasoning. *Psychol. Rev.* 90:38–71.

SAVARY, F., H. WHITAKER, and H. MARKOVITS, 1992. Counterfactual reasoning deficits in right brain damaged patients. Unpublished paper, Department of Psychology, University of Quebec at Montreal.

SCHAEKEN, W., and P. N. JOHNSON-LAIRD, 1993. Mental models and temporal reasoning. Unpublished paper, Department of Psychology, Princeton University.

SHALLICE, T., 1988. *From Neuropsychology to Mental Structure.* Cambridge: Cambridge University Press.

SIMON, H. A., 1959. Theories of decision making in economics and behavioral science. *Am. Econ. Rev.* 49:253–283.

TOMPKINS, C., and C. A. MATEER, 1985. Right hemisphere appreciation of prosodic and linguistic indications of implicit attitude. *Brain Lang.* 24:185–203.

VAN DIJK, T. A., and W. KINTSCH, 1983. *Strategies of Discourse Comprehension.* New York: Academic Press.

WAPNER, W., S. HAMBY, and H. GARDNER, 1981. The role of the right hemisphere in the apprehension of complex linguistic materials. *Brain Lang.* 14:15–33.

WHITAKER, H. A., F. SAVARY, H. MARKOVITS, C. GROU, and C. M. J. BRAUN, 1991. Inference deficits after brain damage. Paper presented at the annual INS meeting, San Antonio, Texas.

ZAIDEL, E., D. W. ZAIDEL, and R. W. SPERRY, 1981. Left and right intelligence: Case studies of Raven's Progressive Matrices following brain bisection and hemi-decortication. *Cortex* 17:167-186.

THE COGNITIVE NEUROSCIENCES

Michael S. Gazzaniga, *Editor-in-Chief*

Section Editors: Emilio Bizzi
Ira B. Black
Colin Blakemore
Leda Cosmides
Stephen M. Kosslyn
Joseph E. LeDoux
J. Anthony Movshon
Steven Pinker
Michael I. Posner
Pasko Rakic
Daniel L. Schacter
John Tooby
Endel Tulving

A BRADFORD BOOK
THE MIT PRESS
CAMBRIDGE, MASSACHUSETTS
LONDON, ENGLAND

Third printing, 1996

This book was set in Baskerville by Asco Trade Typesetting Ltd., Hong Kong and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

The Cognitive neurosciences / Michael S. Gazzaniga, editor-in-chief;
section editors, Emilio Bizzi ... [et al.].
p. cm.
"A Bradford book."
Includes bibliographical references and index.
ISBN 0-262-07157-6
1. Cognitive neuroscience. I. Gazzaniga, Michael S. II. *Bizzi,* Emilio.
[DNLM: 1. Brain—physiology. 2. Cognition—physiology. WL 300 C6765 1994]
QP360.5.C643 1994
153—dc20
DNLM/DLC
for Library of Congress 93-40288
CIP