COGNITIVE SCIENCE 4, 71–115 (1980)

Mental Models in Cognitive Science

P. N. JOHNSON-LAIRD
University of Sussex

INTRODUCTION

If cognitive science does not exist then it is necessary to invent it. That slogan accommodates any reasonable attitude about the subject. One attitude—an optimistic one—is that cognitive science already exists and is alive and flourishing in academe: we have all in our different ways been doing it for years. The gentleman in Molière's play rejoiced to discover that he had been speaking prose for forty years without realizing it: perhaps we are merely celebrating a similar discovery. And, if we just keep going on in the same way, then we are bound to unravel the workings of the mind. Another attitude—my own—is more pessimistic: experimental psychology is *not* going to succeed unaided in elucidating human mentality; artificial intelligence is *not* going to succeed unaided in modelling the mind; nor is any other discipline—linguistics, anthropology, neuroscience, philosophy—going to have any greater success. If we are ever to understand cognition, then we need a new science dedicated to that aim and based only in part on its contributing disciplines. Yet pessimism should not be confused with cynicism. We should reject the view that cognitive science is merely a clever ruse dreamed up to gain research funds—that it is nothing more than six disciplines in search of a grant-giving agency.

Cognitive science does not quite exist: its precursors do, but it lacks a clear identity. Perhaps the major function of this conference should be to concentrate our minds on what that identity might be. At present, there appear to be two distinct ideas wrapped up in it: one topic-oriented, and the other methodological.

The topic-oriented idea is that workers from several disciplines have converged upon a number of central problems and explanatory concepts. George Miller and I became aware of this convergence when we were caught in the toils

of *Language and Perception*. It soon became clear to us that psychology was ill-equipped to provide a semantic theory for natural language, but that other disciplines were tackling some of the problems in a useful way. We, in turn, became embroiled with these different disciplines in an effort to create a psychological plausible lexical semantics. Very much the same process must have occurred, I imagine, in the LNR project (Norman, Rumelhart, *et al*, 1975), in the development of FRAN and HAM (Anderson & Bower, 1973) and in a number of other recent research projects.

Perhaps the most striking example of a concept that has been worked over in radically different fields is that of the *prototype*. Wittgenstein (1953) was the first (at least in modern times) to use the notion. He was reacting to the Fregean doctrine that predicates can be analyzed in terms of sets of necessary and sufficient conditions. Subsequently, Hilary Putnam (1970, 1975) took up the idea, amplified it, and came to the startling conclusion that if meanings are what determine the reference of terms then meanings are not in the mind.[1] Meanwhile, psychologists and anthropologists had been busy establishing the mental reality of prototypical information (see e.g. Berlin & Kay, 1969; Rosch, 1973); workers in artificial intelligence had devized programs for representing prototypes and for exploiting them in visual perception (Falk, 1972; Marr & Nishihara, 1976); and even certain linguists had taken up the idea (see Fillmore, 1975; Lakoff, 1977).

There are other cases where a particular problem or concept has been a focus for work in a number of different disciplines. The study of parsers has been pursued by mathematical linguists, psychologists, and computer scientists; rhythm has been investigated by linguists interested in prosody, psychologists interested in the mental structuring of events, and artificial intelligencers interested in music; decision making has been analyzed by logicians, statisticians, economists and psychologists. Doubtless, we all have our favorite examples, and there must be many more that show an increasing overlap in the research carried out in different academic departments. Unfortunately, cognitive science is unlikely to achieve very much if it is simply involves people with diverse intellectual backgrounds who happen to work on the same problems. "Well," the optimists will say, "there needs to be a *collaboration* between these different individuals." At this point, the question of methodology arises, for the nature of the collaboration calls for more than the interchange of results.

Part of the underlying motivation for Cognitive Science is a dissatisfaction with the orthodox methods of studying cognition, and an impetus to change the fashion in which we think about the mind and investigate its operations. It is tempting to demonstrate the shortcomings of experimental psychology and artificial intelligence, but there are already plenty of such arguments in the literature. The purpose of this paper is certainly to contribute to the process of change, but it

[1]For an attempt to repudiate this thesis, see Johnson-Laird (1979).

is more appropriate on this occasion, and more important in general, to show that we can learn from both experiments and intelligent software. Philosophers distinguish between a correspondence theory of truth and a coherence theory. An assertion is true according to the first theory if it corresponds to some state of affairs in the world; and it is true according to the second theory if it coheres with a set of assertions constituting a general body of knowledge. Psychologists want their theories to correspond to the facts; artificial intelligencers want their theories to be coherent; both groups have adopted the methods best suited to their aims. Cognitive science, however, needs theories that both cohere and correspond to the facts. Hence a rapprochement is required. I will have something more to say on this point later, but in case these observations strike you as ancient truths, my first task is to explore some of the major problems confronting cognitive science.

I will consider (1) the form of mental representations and the questions of whether images differ from sets of propositions, (2) the mental processes that underlie ordinary reasoning and the question of what rules of inference they embody, and (3) the representation of the meanings of words and the question of whether they depend on a decompositional dictionary or a set of meaning postulates. These three questions have stimulated much research, but we still do not know the answers. Moreover, although the questions have been independently pursued, they are intimately related to one another. Their answers all implicate the notion of a *mental model*.

The idea that an organism may make use of an internal model of the world is not new. Even before the advent of digital computers, Kenneth Craik (1943) wrote:

> If the organism carries a "small-scale model" of external reality and of its possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and the future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it.

The power of such a model is illustrated in a simple robot, designed by my colleague, Christopher Longuet-Higgins, which moves freely around the surface of a table, and which, whenever it reaches an edge, rings an alarm bell to summon its human keeper. It possesses neither pressure sensors for detecting edges, nor any sort of electronics. How then does it respond to the edge of the table? The answer turns—literally—on a model. As the robot travels around the table, two small wheels, driven by its main wheels, move a piece of sandpaper around on its baseplate. The position of the small wheels on the paper corresponds exactly to the robot's position on the table. The edge of the paper has a double thickness so that whenever one of the smaller wheels is deflected by it, a simple circuit is closed to ring the alarm. Few cognitive scientists are likely to doubt the power of internal models. What is more problematical is the way in which they are mentally represented and the use to which they are put in cognition.

INFERENCE AND MENTAL MODELS

Aristotle at least by his own account was the first to write on the processes of inference, and he remains in at least one respect in advance of many modern psychologists. Of course, as every schoolgirl knows, there has been an enormous growth in formal logic, particularly since 1879—the year in which both modern logic and experimental psychology began. But logic is not psychology. Aristotle's contribution was to formulate a set of principles governing the syllogism. Syllogisms are extremely simple, consisting of two premises and a conclusion, as this example from Lewis Carroll illustrates:

All prudent men shun hyaenas
All bankers are prudent men
All bankers shun hyaenas

Despite their logical simplicity, however, they have some interesting psychological properties. One such property can be illustrated by the following example. Suppose you are told that in a room full of various people:

Some of the parents are drivers
All of the drivers are scientists

and then asked to state what follows from these two premises. You may care to commit a conclusion to paper before reading on.

We have found in a number of experiments, and many informal observations, that the overwhelming majority of subjects are able to make a valid inference from these premises, but they show a very striking bias. They almost always draw a conclusion of the form:

Some of the parents are scientists

rather than its equally valid converse:

Some of the scientists are parents

This phenomenon, which I have dubbed the "figural effect," does not depend on the fact that the subject of the first premise is "Some of the parents," because it is also observed if the order of the premises is reversed. The results of one study that corroborated the figural effect are summarized in Table 1. The reader will observe that where a syllogism has the form $\frac{AB}{BC}$, as in the example above, 51.2% of the subjects drew a conclusion of the " . . . A . . . C," and only 6.2% drew a conclusion of the converse form. The effect is much less pronounced for syllogisms with symmetric figures:

$$\frac{AB}{CB} \text{ and } \frac{BA}{BC}$$

TABLE 1
The "Figural Effect" Observed in Syllogistic Inference
(from Johnson-Laird & Steedman, 1978)
The Percentages of A-C and C-A Conclusions as a Function of the Figure of the Premises

	Figure of Premises			
Form of Conclusion	A-B B-C	B-A C-B	A-B C-B	B-A B-C
A-C	51.2	4.7	21.2	31.9
C-A	6.2	48.1	20.6	17.8

Note: The table includes both valid and invalid conclusions: the effect is equally strong for both of them. The balance of the percentages corresponds almost entirely to responses of the form, "No valid conclusion can be drawn."

Although the figural effect is virtually unknown among psychologists, it was evident to Aristotle. He argued that a syllogism of the form:

All A are B
All B are C
$\therefore \lambda$ All A are C

was a "perfect" one, because the transitivity of the connection between the terms was immediately obvious. The validity of the argument, he claimed, is self-evident and requires no further support. Indeed, part of his doctrine of the syllogism is to show how arguments in other figures may be "reduced" to the perfect figure (see Kneale & Kneale, 1962, p. 67 et seq.). Unfortunately for psychology, this doctrine was largely supplanted by the rules of the syllogism developed by the medieval Scholastic logicians. Unlike Aristotle, they proposed a set of figures that did not contain the perfect one:

B - A	A - B	B - A	A - B
C - B	C - B	B - C	B - C
C - A	C - A	C - A	C - A

and psychologists have invariably followed this formulation with the result that for fifty years of experimentation they neglected half of the possible syllogisms[2] and failed to detect the potent effect of figure.

[2]Each statement in a syllogism has four possible forms, and hence there are $4^3 = 64$ possible "moods". Psychologists typically go on to claim: "Since each of two terms in each of two premises may appear either first or second, there are 2^2, or 4 possible figures. The variables of mood and figure combine to yield a total of 64×4, or 256 different syllogisms." This number is wrong. There are twice that number of syllogisms. Logicians ignored the order of the premises and made an arbitrary decision to cast their figures so that the subject of the conclusion, 'C' in the examples in the text, occurs in the second premise. Logic is not affected if the subject occurs in the first premise, but plainly the self-evidence of an argument may be affected.

The development of formal logic has not helped psychologists to elucidate the mental processes that underlie inference. There is of course a temptation to treat logic as model of "competence"—as a set of principles that human beings have somehow internalized but depart from occasionally as a result of "performance" limitations. This view is implicit in Boole's (1854) essay on the Laws of Thought, and in our time Piaget and his collaborators have rendered it wholly explicit. The trouble is there are many different logics—there is an infinite number of different modal logics; and any given logic can be formulated in many different ways. If formal logic is to be treated as a model of competence, we need to know which logic or logics human beings have internalized, and the nature of their mental formulation.

The orthodox formulation of a logical calculus consists of specifying (1) the syntactic rules governing well-formed formulae, (2) a set of axioms, and (3) a set of rules of inference that govern deductions from the axioms or from statements derived from them. Since ordinary human beings are little concerned in proving logical theorems, and more concerned with passing logically from one contingent assertion to another, the mental representation of logic should primarily consist of internalized rules of inference: axioms play little part in the logical business of daily life. But what rules of inference do we possess? We have no introspective access to them. It is unclear how we could have come to acquire them or pass them on to the next generation, especially since many everyday inferences appear, at least superficially, to be invalid. It is difficult to imagine that logic is innate—that merely passes the puzzle over to the geneticists—though perhaps an extreme Rationalist might opt for this alternative. The problem about the origin and transmission of rules of inference is so perplexing that I shall argue that there is something mistaken about any conception of reasoning that leads one to pose it.

Theories of syllogistic inference. Although psychologists have studied reasoning experimentally for over seventy years (see e.g. Storring, 1908, for an early study), only in the last five years have they got as far as venturing any hypotheses about the mental processes that underlie syllogistic inference. By far the most typical activity has been the investigation of the hypothesis that the "atmosphere" created by the premise predisposes an individual to accept certain conclusions rather than others. Although the original formulation of the hypothesis was complicated, (see Sells, 1936; Woodworth & Sells, 1935), its essence can be captured in two principles formulated by Begg and Denny (1969):

1. Whenever at least one premise is negative, the most frequently accepted conclusion will be negative; otherwise, it will be affirmative.
2. Whenever at least one premise is particular (i.e. contains the quantifier *some*), the most frequently accepted conclusion will be particular; otherwise it will be universal (i.e. contains either *all* or *none*).

These principles characterize the nature of a putative bias, but they say nothing about the mental processes that underlie it. Moreover, they closely resemble two

of the traditional laws of the syllogism formulated by the Scholastic Logicians (see Cohen & Nagel, 1934). This resemblance makes the atmosphere predictions difficult to test because they often correspond to valid conclusions, and it is accordingly necessary to examine the invalid inferences that people make. Unfortunately, there is little consensus in the literature: some experimenters claim to have confirmed the atmosphere effect (e.g. Begg & Denny, 1969) others claim to have disconfirmed it (e.g. Ceraso & Provitera, 1971; Mazzocco, Legrenzi & Roncato, 1974). One datum that is difficult to reconcile with the effect is that certain premises from which a valid conclusion can be drawn tend to be judged not to imply any conclusion. Here is an example:

Some of the beekeepers are artists
None of the chemists are beekeepers

When such premises were presented in one experiment, 12 out of 20 subjects declared that there was no valid conclusion that could be drawn from them (see Johnson-Laird & Steedman, 1978). In fact, there is a valid conclusion:

Some of the artists are not chemists.

and, moreover, it is entirely congruent with the atmosphere effect: particular because the first premise is particular, and negative because the second premise is negative. Only 2 out of the 20 subjects drew this conclusion. Such findings require at the very least some modification of the atmosphere hypothesis.

It is obviously more important to give an account of the mental processes that underlie syllogistic inference than to attempt to explain the putative effects of "atmosphere." In fact, three major theories have been developed in the last few years.

1. Erickson (1974, 1978) argues that the premises of a syllogism are mentally represented in a form that corresponds to Euler circles. He postulates that only a single representation is used for each premise and so, for example, he assumes that a premise of the form *All A are B* is represented by two co-incident circles on 75% of occasions, and by one circle, A, within another, B, on 25% of occasions. An inference is made by combining the separate representations of the two premises, though Erickson does not specify any effective procedure for making such a combination. It is generally possible to combine such representations in more than one way. In one version of his theory, Erickson supposes that subjects consider all the different possible combinations; in another version, he supposes that they consider only one selected at random from the set of possible combinations. Unfortunately, this latter procedure will always yield a conclusion, and the theory is accordingly unable to predict responses of the form, "There is no valid conclusion." Moreover, if only a single combination is constructed, then there will be occasions where an overlap between sets ought to lead to a conclusion of the form, "Some A are C", and other occasions where it ought to lead to a conclusion of the form, "Some A are *not* C." Erickson accordingly invokes the atmosphere effect to account for the fact that subjects tend to make the appropriate response. A major difficulty with both versions of

the theory is that Euler circles are symmetrical: if they correspond to a conclusion, *Some A are C*, then they equally correspond to the conclusion *Some C are A*. The theory is accordingly totally unable to account for the figural effect.

2. An alternative theory is based on the idea that subjects illicitly convert both *All A are B* to *All B are A*, and *Some A are not B* to *Some B are not A* (Chapman & Chapman, 1959). This notion has been elevated into an information-processing model by Revlis (1975a,b). In its most recent formulation (Revlin & Leirer, 1978), the theory assumes that, during the process of encoding the premises, the reasoner converts each premise unless the result is an assertion that is obviously factually false. The reasoner then applies entirely logical processes to the resulting representations in order to derive a conclusion (though the theory does not specify the nature of these processes). It follows that the premises:

All A are B
Some B are C

should be converted during their encoding to yield:

All B are A
Some C are B

which logically imply the conclusion:

Some C are A

though this conclusion, of course, fails to follow from the original premises. Unfortunately, the theory leads naturally to a prediction exactly contrary to the figural effect: if subjects automatically convert premises, then there is no reason to suppose that they will be biased towards a conclusion of one form rather than another.

3. Robert Sternberg and his colleagues have recently proposed a model that attempts to remedy some of the difficulties of manipulating Euler circle representations (Guyote & Sternberg, 1978; Sternberg & Turner, 1978). This theory assumes that subjects represent premises in a logically correct way. Hence, a premise of the form *All A are B* requires two separate representations: one corresponding to the inclusion of set A within B, and one corresponding to an equivalence in the extension of the two sets. The first of these representations has a form corresponding to:

$a_1 \rightarrow B \quad b_1 \rightarrow A$
$a_2 \rightarrow B \quad b_2 \rightarrow -A$

where the lower case letters denote disjoint, exhaustive partitions of the corresponding sets denoted by capital letters, and the arrow denotes class inclusion. Thus the left-hand side of the representation states that each of the two partitions of set A, a_1 *and* a_2, is included in set B, and the right-hand side of the representation states that one of the partitions of set B, b_1, is included in set A and the other

of the partitions of set B, b_2, is included in not−A, the complement of A; in other words, set A is a proper subset of B. The choice of the number of partitions is arbitrary. Although the representation of premises is logically correct, according to the theory their combinations can give rise to errors. In particular, Sternberg and his colleagues assume that a subject never makes more than four combined representations; the particular four depend on an ordering postulated by the theory. The final state of an inference requires the subject to find a verbal description that is consistent with the set of combined representations. If there is no such label, then the premises are indeterminate. If there are two such labels, the theory assumes that subjects are biased both by the atmosphere effect and by a preference for descriptions that are consistent with the smallest number of alternatives. The theory also proposes that subjects are prone to become confused if the set of final representations appears not to be consistent with any verbal description. Although the representations postulated by this theory are very much easier to manipulate than Euler circles, they share with them precisely the same difficulty of being unable to account for the figural effect. Any representation that leads to the conclusion *Some A are C* will lead equally to the conclusion *Some C are A*.

Criteria for Evaluating Theories of Syllogistic Inference

An adequate theory of syllogistic inference should satisfy the following points.

First, the theory should account for the systematic mistakes, and the habitual biases, including the figural effect, that are observed in experiments, and also for the fact that many valid inferences are drawn.

Second, the theory should be readily extendable so that it applies to all sorts of quantified assertions. It should accommodate sentences that contain more than one quantifier, e.g. "Every man loves a woman who loves him." It should also accommodate sentences that contain such quantifiers as *most*, *many*, *several*, and *few*.

Third, the theory should provide an account of how children acquire the ability to make deductive inferences.

Fourth, the theory should be at least compatible with the development of formal logic, that is to say, it should allow that human beings are capable of rational thought, and that they have been able to formulate principles that govern valid inference.

All three of the theories described above fare poorly on these criteria, and it is therefore worth considering a different approach based on the notion of a mental model (Johnson-Laird, 1975).

Syllogistic Inference as the Manipulation of Mental Models

One way in which you could interpret a pair of premises such as:

All of the singers are professors
All of the poets are professors

would be by actually gathering together a number of individuals—actors, say—in a room, and then assigning them the roles of singer, professor, and poet, in a way that satisfies the premises. Logical principles can determine whether a given conclusion is valid, but they cannot even in principle specify what particular conclusion to draw from some premises on a given occasion, because there are always infinitely many valid conclusions that could be drawn. Most of them are trivial, of course, such as a disjunction of the premises.[3] Hence, in order to derive a specific conclusion from the premises, you need some extra-logical principle to guide you. Let us suppose that you work according to the heuristic procedure of always trying to establish as many identities as possible between the different roles that you assign. This heuristic is designed to cut down on the number of actors that you have to employ by maximizing the number of connections that are formed between the different roles. It keeps matters simple. Thus, you get together, say, six actors. The first premise asserts that all of the singers are professors, and so you arbitrarily assign three actors to play the part of singers, and, in accordance with the premise, you specify that each of them is also a professor. Of course there may be professors in the room who are not singers, and so you arbitrarily assign that role to the remaining three actors, but since the premise does not establish that they definitely exist, these individuals represent only a possibility. You have accordingly interpreted the first premise by establishing the following scenario:

```
singer = professor
singer = professor
singer = professor
         (professor)
         (professor)
         (professor)
```

where the parentheses indicate that the relevant individuals may, or may not, exist. You interpret the second premise, *all of the poets are professors*, in a similar way, using your heuristic principle in order to establish as many identities as possible:

```
singer = professor = poet
singer = professor = poet
singer = professor = poet
         (professor)
         (professor)
         (professor)
```

At this point, you might conclude (invalidly) as did a certain proportion of the subjects in our experiment (Johnson-Laird & Steedman, 1978) that *all of the singers are poets*, or conversely that *all of the poets are singers* since the form of the premises is not such as to give rise to the figural effect. However, if you are

[3]The inability of logic alone to provide the formulation for a theory of inference has been overlooked in nearly every psychological theory of reasoning—most notably in the Piagetian school (cf. Inhelder & Piaget, 1958, 1964), but also in other theories (e.g. Martin, in press).

prudent, you might refrain from drawing a conclusion until you have checked its logical validity. You must establish whether the identities between the various roles are irrefutable: you must attempt to destroy them without doing violence to the meaning of the premises. You should discover that you can break at least one of the identities without violating the premises:

```
singer = professor = poet
singer = professor = poet
singer = professor
         professor = poet
        (professor)
        (professor)
```

At this point, you may be tempted—again like some subjects—to conclude (invalidly) that *some of the singers are poets*, or conversely that *some of the poets are singers*. However, if you are really prudent, you may try to extend your destructive manoeuvre to all the identities. This step leads to the following re-assignment of roles, in which all the original identities are destroyed:

```
singer = professor
singer = professor
singer = professor
         professor = poet
         professor = poet
         professor = poet
```

Since you have been able to arrange matters so that none of the singers are poets, and hitherto you had arranged them so that all of the singers are poets, now at last you should appreciate—as some subjects do—that you cannot draw any valid inference about the relations between the singers and the poets.

The present theory of quantified inferences assumes that you can carry out the whole of the above procedure as a "thought experiment." You construct a mental model of the relevant individuals, you form identities between them according to the heuristic, and, if you are logically prudent, you attempt to test your mental model to destruction.

An Evaluation of the Mental Model Theory of Inference

How does the present theory measure up to the criteria on our shopping list? First, it provides an account of both the figural effect and the systematic errors that tend to occur in syllogistic reasoning. The representation of identities such as: $a = b$, depends on a list-structure in which there is an asymmetry in ease of search: given a it is relatively easy to establish its identity with b, but given b it is relatively hard to establish its identity with a. Premises that give rise to the figural effect yield a uniform direction of search, whereas the others do not. Likewise, the theory obviously predicts that those premises for which the heuristic yields a valid conclusion should be easier to cope with than those premises for which a valid conclusion emerges only after submitting the model to a logical

test. This prediction was readily confirmed: 80.4% of responses to the first sort of premises were correct whereas only 46.5% of responses to the second sort of premises were correct, and this pattern of results was obtained from each of the subjects who was tested (see Johnson-Laird & Steedman, 1978, for a detailed account).

Second, mental models can obviously be generated so as to represent all sorts of quantified assertions. They accommodate multiply-quantified assertions such as "Every man loves a woman who loves him," which cannot be represented by Euler circles. They can even represent sentences that are claimed to demand "branching" quantifiers that go beyond the resources of the ordinary predicate calculus, such as "Some relative of each villager and some relative of each townsman hate each other," (see Hintikka, 1974.) They can accommodate such quantifiers as *most*, *many*, *several* and *few*. They enable distinctions to be drawn between *each* and *every*, and *any* and *all*, as Janet Fodor (1979) has independently shown in a theory with a striking resemblance to the present account. Models also allow a clear distinction to be drawn between class-inclusion and class-membership. The assertion:

John is a Scotsman

concerns class-membership and can be represented as:

John = Scotsman
Scotsman
Scotsman
.
.
.
.

The assertion:

Scotsmen are numerous

also concerns class-membership and can be represented as:

Scotsman
Scotsman = numerous
Scotsman numerous
. numerous
. .
.
.
.
.

In other words, the set of Scotsmen is identical to one of the members of the set of sets of numerous entities. The combination of the two premises

John is a Scotsman
Scotsmen are numerous

leads to a representation from which one can *not* conclude:[4]

John is numerous

Third, the theory of mental models does illuminate the way in which children learn to make inferences and the problematical question of the nature of the rules of inference that they internalize. The theory contains no rules of inference. Its logical component consists solely in a procedure for testing mental models: the aim is to establish the falsity of a putative conclusion by destroying the model from which it derives, but the manipulations that attempt to carry out this process of destruction are constrained in that they must never yield a model that is inconsistent with the premises. The reader will recall that a rule of inference specifies in an essentially "syntactic" way a set of premises and a conclusion that can be derived from them. No such rules are invoked by the theory. This claim may be confusing, so let me elaborate it.

In addition to the formal or syntactic stipulation of rules of inference that enable certain formulae to be derived, a logician can give a semantic characterisation of a logical calculus. He can do so by providing a *model-structure* for it, which consists of a model—a set of entities that provide the referents for the terms in the calculus, an interpretation function that specifies the referents (in the model) for the terms and predicates of the language, and a set of rules governing the way in which the interpretations of complex expressions are built up from the interpretations of their simpler constituents. Any well-formed sentence in the calculus will have a determinate truth value with respect to the model-structure, whose function is precisely to provide such interpretations. A rule of inference should accordingly yield only *valid* conclusions, that is, if it is applied to premises that are true with respect to the model structure, then it should yield only conclusions that are also true with respect to the model structure. Logicians are seldom interested in a particular model structure: the principle of validity must hold over any and every model that can be formulated for the calculus. There is an interesting relation between the model structures of formal logic and the mental models postulated in the present theory. The psychological theory posits a process of inference that involves, not the mobilization of quasi-syntactic rules of inference, but the direct manipulation of a model of the assertions in the premises. The notion is not wholly foreign to formal logic: the theory of *natural deduction* is based essentially on the same principle (see Beth, 1971). It is perhaps for this reason that the formal aspects of natural deduction have had some popularity amongst psychologists (see Johnson-Laird, 1975; Osherson,

[4]Unfortunately, such inferences are sanctioned by the theory proposed by Guyote and Sternberg (1978). They remark: ". . . the choice of the number of partitions [in their representation] is arbitrary, and of course, the most accurate representation of a set would have as many partitions as there are members of the set". However, a partition is a subset of the set whereas a member is not a subset. Guyote and Sternberg represent "X is a B", where "X" denotes an individual, as X→B, that is, in exactly the same way as they represent the subset relation. The arrow stands for class-inclusion.

1975; Braine, 1978) and artificial intelligencers (see Bledsoe, Boger & Henneman, 1972; Reiter, 1973).

The reader may be tempted to suppose nevertheless that somewhere in the theory of mental models for syllogistic inference there lurks some machinery equivalent to a set of rules of inference. The temptation must be resisted. A computer program that I have devised works according to the theory and uses no rules of inference. Its power resides in the procedures for constructing and manipulating models—a power which in turn demands at the very least the recursive power of list-processing operations.

Fourth, and finally, although the theory contains no rules of inference it is entirely compatible with the development of formal logic. Another computer program devised by Mark Steedman showed that simplifying the operation of the psychological principles embodied in the theory by natural computational "short cuts" led to the recovery of all the traditional laws of the syllogism. For example, with affirmative premises, it transpires that whenever one identity can be broken, then, as in the example above, all of them can be broken. Steedman implemented an extremely simple version of this principle: the relevant procedure looked for a middle item that was not linked by an identity to any end items, and whenever such an item was found the program indicated that no valid conclusion could be drawn about the relations between the end items. This procedure sacrifices psychological plausibility for the sake of simplicity: it cuts out a whole series of processes that are likely to occur when logically naive individuals reason, and that are modelled in the first program. However, the abstraction that Steedman's program embodies corresponds directly to the traditional law that the middle term must be distributed at least once in a valid syllogism (see Cohen & Nagel, 1934, p. 79). A logician's conscious reflection on the invariant properties of his own deductions could well have played an analogous role in the development of logic. Aristotle's own procedure for demonstrating that a pair of premises does not yield a conclusion bears a striking resemblance to a consciously applied method of manipulating models (see Kneale & Kneale, 1962, p. 75). He compares two different instances of a syllogism of the same form. The syllogism

Every man is an animal
No stone is a man

∴No stone is an animal

might be thought to be valid, but he compares it with:

Every man is an animal
No horse is a man

∴No horse is an animal

Aristotle's technique is accordingly to show by such examples that premises of the form:

Every B is A
No C is B

are consistent with

Every	C is A (e.g. Every horse is an animal)
No	C is A (e.g. No stone is an animal)
Some	C is A
Some	C is not A.

The method is wholly semantic and, in effect, externalizes the method of destroying putative conclusions by manipulating models.

The theory of mental models is compatible with the origins of logic. It allows that human beings are capable of rational thought; that they may fall into error if they fail to carry out a comprehensive destructive test of the models that they create, and that their discovery of this tendency to err may have led, in part by reflection on the invariant properties of deduction, to the formulation of logical laws.

MEANING AND MENTAL MODELS

There is a controversy about the proper form of a psychologically adequate semantic theory that can be resolved by following through the implications of the theory of mental models. Psychologists have generally agreed that a major burden for the meaning of words is to account for the relation between such assertions as "Polly is a parrot" and "Polly is a bird"—if the first assertion is true, then plainly so is the second. What they disagree about is the nature of the semantic machinery needed to explain such relations.

One school of thought, whose recent ancestry can be traced back to the work of Katz and Fodor (1963)—though it has a much longer history reaching back into antiquity—holds that the meaning of a word such as "parrot" is represented in the mental lexicon as a set of semantic elements that includes, amongst others, those corresponding to "bird." The relation between the two sentences is accordingly captured by the decomposition of the entries in the mental lexicon. A wide variety of psychological theories of meaning are committed to some sort of decomposition into semantic primitives (Clark & Clark, 1977; Collins & Quillian, 1972; Miller & Johnson-Laird, 1976; Norman & Rumelhart, 1975; Schank, 1975; Smith, Shoben & Rips, 1974).

An alternative view is that there are neither semantic primitives nor decompositional lexical entries (Fodor, 1976; Fodor, 1977; Fodor, Fodor, & Gar-

rett, 1975; Kintsch, 1974; Lyons, 1977). Entailments that depend upon the meanings of words are, according to these theorists, captured by meaning postulates (see Carnap, 1956). Meaning postulates stipulate the semantic relations between words, e.g. *for any x, if x is a parrot then x is a bird*. Such rules are introduced into a model-theoretic semantics of a language in order to render some models inadmissible, namely, those for which the meaning postulates are not true. Latterly, the idea has been cut loose from formal semantics and imported into psychological theory. Kintsch (1974) and Fodor et al (1975) assume that sentences in a natural language are translated into 'propositional representations' in a corresponding mental language, and that meaning postulates couched in the mental vocabulary are used to make inferences from these propositional representations.

Two Problems for Meaning Postulates

Although there have been attempts to resolve the controversy about meaning experimentally, the results so far are equivocal. Some findings appear to count against decomposition (Kintsch, 1974; Fodor et al, 1975); other findings appear to count against meaning postulates (Clark & Clark, 1977; Johnson-Laird, Gibbs & de Mowbray, 1978). But, as yet, there are no results sufficiently decisive to resolve the issue. Indeed, there has been a tendency to accept the view of Katz and Nagel (1974) that there is no fundamental distinction between the two sorts of theory. There are, in fact, several arguments that could be made to establish a difference in their psychological plausibility. I shall present two: the first concerns simple inferences based on premises in ordinary language, and the second the relation between language and the world.

Consider the following simple inference:

The pencil is in the box
The box is in the envelope

∴ The pencil is in the envelope

Obviously, it is valid since no one in practice would doubt the truth of the conclusion given the truth of the premises. Meaning postulates provide an initially plausible basis for such an inference. The premises are translated into a propositional representation, which according to Kintsch (1974) might take the following sort of form:

(IN, PENCIL, BOX)
(IN, BOX, ENVELOPE)

and then the meaning postulate that captures the transitivity of "in":

For any x, y, z, (If (IN, x, y) & (IN, y,z)) then (IN, x, z)

is applied to yield the conclusion:

(IN, PENCIL, ENVELOPE)

And this propositional representation can, if necessary, be translated back into natural language.

Although the details of the various processes of translation have not been formulated explicitly by any theorist, they are not problematical as far as the present argument is concerned. It covers any processes that lead parsimoniously to propositional representations and to the application of meaning postulates to them. There is nothing privileged about meaning postulates here, they may be replaced by any rules of inference that apply to such propositional representations.

The heart of the argument depends on the following sort of inference:

Luke is on Mark's right
Mark is on Matthew's right

∴ Luke is on Matthew's right

It is not immediately clear whether this inference is valid. If the three individuals are sitting in a straight line on one side of a table, then the relation referred to by "on x's right" is transitive, and the inference is valid. But if they are sitting at equal intervals round a small circular table, then the relation referred to by "on x's right" is not transitive, and the inference is invalid.

A natural way to try to accommodate this phenomenon within the framework of a propositional theory is to propose two different meanings for "on the right" and its cognates, one to which a meaning postulate expressing transitivity applies, and one to which a meaning postulate expressing intransitivity applies. However, if a number of people are seated round a *large* circular table, then the previous inference could be valid, but one might have doubts about the following one:

John is on Luke's right
Luke is on Mark's right
Mark is on Matthew's right

∴ John is on Matthew's right

As more and more individuals are added round the table, there will inevitably come a point where transitivity breaks down. (As a matter of fact, there is likely to be a region of uncertainty, but this possibility merely exacerbates the problems of a meaning postulate theory.) In general, the particular relation referred to by "on the right" may be intransitive or the extent of its transitivity may vary over any number of items from three to an arbitrarily large number. Each of these extents would require its own separate meaning postulate with the number of premises in its antecedent directly correlated with the number of items over which transitivity holds—two premises for transitivity over three items, three premises for transitivity over four items, and so on *ad infinitum*. Because there is

no limit to the number of items at which transitivity ceases to hold, there is no limit to the number of separate meaning postulates that are required to cope with the semantics of this single term. This conclusion is psychologically unacceptable on the reasonable criterion, decisive in other contexts (Miller & Chomsky, 1963), that human beings do not have an unlimited capacity for storing information, or the ability to learn an infinite number of rules.

It should be emphasized that these difficulties are not peculiar to "right" and "left." English vocabulary is plagued by the same sorts of problem, and it is hard to find any simple spatial terms that have an unequivocal meaning. Inferences based on such terms as "at," "between," "near," "next to," "on," and "in" can all reflect the uncertainties of transitivity.

A proponent of meaning postulates might argue that once the transitivity of "on the right" ranges over some large number of items, say, 100, then it can be taken to have an unlimited extent. This *ad hoc* proposal has at least the virtue of limiting the required meaning postulates to a finite number. Yet, it does not solve the problem: no matter how large the radius of a circle and how densely the individuals are packed around it, it *is* a circle and transitivity must break down. Moreover, this proposal highlights another difficulty: how is the appropriate meaning postulate recovered by someone attempting to make an inference? It is clear that any feasible answer to this question will depend on some mechanism for determining the nature of the situation referred to explicitly or implicitly by the premises. In the case of our examples, it will depend on information about the table and the seating arrangements, which in turn will be used to select the appropriate meaning postulate.

Once the need to deal with reference situations is admitted, the second argument against the meaning postulate account can be made. The theory contains an obvious, though deliberate, gap which is again best illustrated by a simple example. Given the following arrangement of letters:

B A

any competent speaker of English knows that it is true to say of them, "A is on the right of B" and false to say of them, "A is on the left of B". This distinction reflects the difference in meaning between "right" and "left"; yet, there is no way to capture it using meaning postulates. One can, of course, establish that there is a difference in meaning between the two terms, e.g. *for any x and y, x is on the right of y if and only if y is on the left of x*, and *for any x and y, if x is on the right of y then x is not on the left of y*. These postulates establish that a difference exists, but they do not specify its nature. For that, it is necessary to make explicit what it is that underlies our knowledge that A is indeed on the right of B in the example above.

Procedures for Manipulating Mental Models

The idea lying behind the psychological exploitation of meaning postulates, and indeed most decompositional theories of meaning too, is that it is feasible to specify the semantic relations between words without considering how they relate to the world: intensions can be profitably pursued independently from extensions. The principle seems plausible for meaning postulates in their original context of formal semantics, where the real world is replaced by a model structure in which the extensions of terms are assigned directly. But the precedent is misleading for natural language where, as we shall see, the only way to account for the proper relations between words, and for inferences based upon them, is by giving a specification of their meanings that includes their relations to the world. What is missing in the meaning postulate account is a *definition* of how "right" and "left" relate to the world. The reason for this omission is obvious: the relations are so basic that there is no way to define them in ordinary English. It is for this reason that a complete theory of meaning must rely upon some sort of decomposition into more primitive notions.

Is it possible to save a propositional theory by sacrificing meaning postulates? The answer depends, of course, on what processes are used to make inferences in their stead. Any system that relies on rules that manipulate propositions will have to introduce some machinery to handle transitive relations, and hence it will be in imminent danger of falling into precisely the same difficulties. The only escape route will be a method for handling the facts of transitivity without relying on rules, postulates, or productions, for transitivity itself. Once again, we need to get rid of rules of inference. This prescription may seem to be impossible to fulfill; fortunately, there is at least one way in which it can be met.

The semantics of spatial terms and the uncertainties of their transitivity can be accommodated within a sort of decompositional theory that has come to be known as "procedural semantics" (see Davies & Isard, 1972; Johnson-Laird, 1977; Miller & Johnson-Laird, 1976; Woods, 1967, 1979). The theory can be illustrated by considering a computer program (written in POP-10) that I have devised in order to investigate spatial inference. The purpose of the program is to evaluate premises about the spatial relations between objects. It works by building up a two-dimensional spatial model that satisfies the premises given to it, and indicates whether a premise is implied by, or is inconsistent with, what it has already been told. It accordingly contains a number of *general procedures* for constructing, recursively manipulating, and interrogating sets of models. One procedure constructs a new model for any premise that refers only to entities that have not been mentioned previously. Another procedure, given the location in the model of one item mentioned in a premise, puts another item into the same model at a place that satisfies the premise. Another general procedure is used to verify whether the relation specified to hold between two items, say A and B,

obtains within a model. It works by locating B and then by looking along a line from B in order to determine whether or not A is somewhere on that line. If A is found to lie on the line then the premise is true, otherwise it is false. The verification procedure contains two parameters, DX and DY, whose values specify the direction of the line: they give the respective increments on the x and y axes of the model that define the locations to be examined. This use of parameters to specify directions is common to all the general procedures used by the program, including those for inserting new items into a model. This uniformity makes it possible to define the meanings of relational terms as procedures that work in a way that is utterly remote from meaning postulates and conventional decompositional theories.

The meaning of "on the right of" consists of a single procedure: FUNCT(% 0, 1 %). This takes whatever general procedure is about to be executed, and which has been assigned as the value of the variable, FUNCT, and "freezes in" the value of 0 to its DY parameter and the value of 1 to its DX parameter. The decorated parentheses are a standard device in POP-10 for freezing in the values of parameters, with the effect of converting a general procedure into a new more specific procedure that takes fewer arguments—one less for each argument that has had its value frozen in. The effect of FUNCT(% 0, 1%) on the verification procedure is accordingly to produce a procedure that scans a specified sequence of locations lying in a particular orientation. Since DY = 0, they have the same y-coordinate as the object B; and since DX = 1, they are spelt out by successive increments of 1 on the X-coordinate. In other words, if you imagine the spatial array laid out on a table in front of you, the procedure examines a sequence of locations lying progressively further to the right of B: it looks to see whether A is on the right of B. The same process of freezing in the values of parameters is used to convert the program's other general procedures into specific ones that depend on the relation specified in a premise.

The program's lexical entries define how words relate to its model of the world; but they stipulate nothing about transitivity or intransitivity. However, in the program's simple rectilinear world, a relation such as "on the right of" has the emergent property of transitivity, that is to say, whenever A is on the right of B and B is on the right of C, then as a matter of fact A will be on the right of C, whether the program is building, manipulating, or interpreting a model. The program can accordingly make transitive inferences even though it contains no rules, postulates, or productions, for transitivity itself. This facility depends crucially on its use of spatial models and procedural definitions that relate directly to them. The definitions decompose meanings into the primitive components of specific coordinate values that are only interpretable with respect to the spatial models. The meaning of a word is accordingly not a procedure that can do anything by itself; it is a procedure that applies to other procedures. If the locus of the entities in a reference situation is circular rather than rectilinear, then exactly the same lexical procedures will give rise to transitivity locally, but sooner or later it will fail as the entities depart further and further from the required sequence of locations passing through the initial object in the series.

The program is intended neither as an exercise in artificial intelligence nor as a computer simulation of spatial inference. It is far too simple to be psychologically realistic—for example, human beings do not just consider single lines, and whether objects lie on or off them, in determining spatial relations. Its purpose is merely to establish the feasibility of a theory of semantics based on the assumption that the meanings of words are decompositional procedures that relate to mental models of the world, and, in particular, on the use of lexical procedures that interact with the general procedures for constructing manipulating and evaluating mental models. There is a twofold advantage of this approach over any theory based on meaning postulates. First, the procedural theory gives an account of the extensions of expressions, which meaning postulates are neither intended nor able to do. Second, the vagaries of transitivity, which the meaning postulate theory is presumably intended to handle, emerge in a wholly natural way from the operation of procedures on mental models.

IMAGES, PROPOSITIONS, AND MENTAL MODELS

The concept of a mental model, which has been used throughout this paper, has yet to be analyzed in any detail. Undoubtedly, it resembles some of the current conceptions of an image. However, there is little agreement about the properties of images other than that they give rise to an obvious subjective experience, whereas this characteristic is wholly irrelevant to mental models, which need not possess any immediately "pictorial" attributes. In order to specify their positive characteristics, however, I need to resolve the controversy about images and propositional representations.

Images versus Propositional Representations

Many human beings claim to be able to form and to manipulate mental images in the absence of corresponding visual stimuli. The phenomenon has been studied empirically for a century, dating from Galton's questionnaire on his correspondents' ability to imagine their breakfast tables (Galton, 1928, originally published in 1880). More recent studies have examined a variety of aspects of images, including their use as mnemonics (Bower, 1972; Paivio, 1971), their mental rotation and transformation (Cooper, 1975; Shepard, 1975), their suppression by other tasks (Brooks, 1967, 1968; Byrne, 1974), and their use in retrieving information about objects (Hayes, 1973; Holyoak, 1977; Kosslyn, 1975, 1976; Moyer, 1973; Paivio, 1975). No one seriously doubts the existence of the psychological phenomena of imagery. What is problematical, however, is the explanation of the phenomena and the ultimate nature of images as mental representations. It seems unlikely that they are simple pictures in the head,

because this conjecture leads to a number of undesirable consequences including the need for an homunculus to perceive the pictures, and thus to the danger of an infinite regress (Dennett, 1969). There remain two schools of thought.

On the one hand, there are those who argue that an image is distinct from a mere representation of propositions (Bugelski, 1970; Kosslyn & Pomerantz, 1977; Paivio, 1971, 1977; Shepard, 1975, 1978; Sloman, 1971). These authors attribute a variety of properties to images. The consensus, in so far as one can be detected, embodies the following points:

1. The mental processes underlying an image are similar to those underlying the perception of an object or a picture.
2. An image is a coherent and integrated representation in which each element of a represented object occurs only once with all its relations to other elements readily accessible.
3. An image is amenable to apparently continuous mental transformations, such as rotations or expansions, in which intermediate states correspond to intermediate states (or views) of an actual object undergoing the corresponding physical transformation. Hence, a small change in the image corresponds to a small change in the object (or its appearance).
4. Images represent objects. They are *analogical* in that the structural relations between their parts correspond to those between the parts of the objects represented. There may indeed be an isomorphism between an image and its object, though this claim makes sense only with respect to an object viewed as decomposed into parts with particular relations between them.

On the other hand, there are theorists who argue that the subjective experience of an image is epiphenomenal and that its underlying representation is propositional in form (Anderson & Bower, 1973; Baylor, 1971; Kieras, 1978; Morgan, 1973; Palmer, 1975; Pylyshyn, 1973). The main properties of such a representation, again in so far as there is a consensus, are as follows:

1. The mental processes underlying a propositional representation are similar to those underlying the perception of an object or picture.
2. The same element or part of an object may be referred to by many of the different propositions that constitute the description of the object. However, when propositions are represented in the form of a semantic network, then the representation is coherent and integrated, and each element of the represented object occurs only once with all its relations to other elements readily accessible.
3. A propositional representation is discrete and digital rather than continuous. However, it can represent continuous processes by small successive increments of the relevant variable(s), such as the angle of an object's major axis to a frame of reference. Hence, a small change in the representation can correspond to a small change in the object (or its appearance).
4. Propositions are true or false of objects. Their representations are *abstract* in that they do not resemble either words or pictures, though they may be needed to provide an interlingua[5] between them (Chase & Clark, 1972). Their structure is not analogous to the structure of the objects that they represent.

The critics of imagery often allow that an image can be constructed from its propositional description, but such an image does not introduce any new information, it merely makes the stored description more accessible and easier to manipulate. Gelernter's (1963) program for proving geometric theorems, and Funt's 1977) program for making inferences about the stability of arrangements of blocks, are both considerably enhanced by the use of procedures that operate on diagrammatic representations. However, Pylyshyn (1973) argues that picturelike representations are not necessary for such purposes: the same function can be served by propositional descriptions. This view has been pushed still further by Palmer (1975):

> The arguments in favor of analogical representations tend to emphasize the relative ease with which certain operations can be performed on them compared to the difficulty in performing the same operations on propositional representations. These arguments, however, generally overlook the fact that propositions can encode quantitative as well as qualitative information. In addition, it is not often recognized that propositions are capable of encoding an analog image.

Palmer then goes on to establish both a way in which a shape such as a triangle can be encoded propositionally and a method for rotating such representations once they have been decomposed into their propositional constituents.

Evidently, the two sorts of representation share a number of properties: they differ mainly on the fourth of the characteristics listed above—the function served by the representation. Otherwise, their apparent similarity and the view that they are readily transformed into one another has indeed led some commentators to conclude that the controversy is neither fundamental (Norman & Rumelhart, 1975) nor resolvable (Anderson, 1976, 1978). In particular, Anderson (1978) argues that "any claim for a particular representation is impossible to evaluate unless one specifies the processes that will operate on this representation." He shows that a theory based on images can be mimicked by one based on propositions provided that certain conditions are satisfied.

Anderson's Theorem on "Mimicry"

Anderson's argument is intended to establish that given a theory which embodies assumptions about mental representations and processes, it is possible, in principle, to construct other theories with different sorts of representations that

[5]There is danger of an infinite regress here. If an interlingua is needed to mediate between words and pictures, then perhaps a language is needed to mediate between words and the interlingua, or between the interlingua and pictures, and so on and on (see Anderson, 1978).

nevertheless behave in an equivalent manner. Suppose, for instance, that one wishes to show that with suitable mental operations, a propositional theory can mimic an imaginal theory. The trick is to embed the whole of the imaginal theory within the operations carried out on the propositional representations. The imaginal theory assumes, say, that a stimulus is encoded as an image, which can be mentally rotated in order to determine whether it coincides with another stimulus. The propositional theory assumes only that a stimulus is encoded as a set of propositions. The following operations can accordingly be postulated as part of the propositional theory:

1. Apply the inverse of the propositional encoding to the set of propositions in order to recover the original "stimulus" (i.e. its sensory image).
2. Apply the imaginal encoding to this stimulus in order to obtain the corresponding image.
3. Rotate the image,
4. Apply the inverse of the imaginal encoding to the rotated image in order to obtain the corresponding stimulus.
5. Apply the propositional encoding to the stimulus in order to obtain the set of propositions corresponding to the rotated image.

The decision about whether these propositions match the second stimulus can again, if necessary, relay on the imaginal theory:

6. Apply the inverse of the propositional encoding in order to obtain the stimulus corresponding to the rotated image.
   (This stimulus is, of course, identical to the one obtained in step 4.)
7. Apply the imaginal encoding to the stimulus to obtain the corresponding image.
   (This image is identical to the one obtained from step 3.)
8. Compare the image to the one obtained from the second stimulus, and make the appropriate response.

Although this chain of operations can be postulated, its feasibility depends on a crucial condition: it must be possible to apply the inverse of the propositional encoding to obtain the original stimulus, or, more plausibly, a sensory representation isomorphic to the original stimulus. However, since perception is likely to involve a many-one mapping, the inverse may fail to yield the original "stimulus." It is for this reason that Anderson imposes the condition that there must be a one-to-one mapping between the respective representations of the two theories. Granted this condition, the inverse of the propositional encoding can yield any of the "stimuli" that could have given rise to the relevant set of propositions, and it will not matter which stimulus is selected, because they will all be equivalent for the imaginal theory, too.

That a propositional theory can mimic an imaginal theory by importing the whole apparatus of images is plainly a trivial result. What is of interest is the possibility of a more direct method of mimicry that does not depend upon embedding one theory within another. Unfortunately, there is no guarantee that a direct method can always be found for two alternative representational theories.

Anderson makes only the modest claim: " . . . it seems we can usually construct [the required operation] more simply than its formally guaranteed specification." Moreover, if one theory encodes stimuli into classes that do not correspond one-to-one with the encodings of the other theory, then the whole system of mapping breaks down.

Considerable care needs to be exercised in drawing conclusions on the basis of Anderson's demonstration. He himself (Anderson, 1976, p. 74) makes the following claim:

> Any behavior that can be computed from inspecting semantic primitives can be computed with the aid of "meaning postulates" that interpret more complex semantic units. This follows from the theorem . . . that any representation can mimic the behavior of any other, provided they impose the same equivalence class on their inputs.

The first assertion has, of course, proved to be false: meaning postulates cannot handle the reference of expressions or the uncertainties of transitivity, but lexical entries based on procedural primitives can accommodate them. It follows that the two sorts of theory do not impose the same equivalence classes on their inputs. And this conclusion is clinched by considering sentences of the form: "A is in front of B, which is behind C." The sentence is unambiguous[6] and should accordingly receive a single propositional representation, but it is referentially indeterminate—the relation between A and C is unspecified—and can accordingly be represented by a number of different mental models. Once one has constructed a particular model, it is impossible to recover the original premises on which it is based. This distinction drives a wedge between sets of propositions and mental models that is not easily removed.

It might be supposed that the propositional representation could mimic the model representation, and yield two alternatives: one in which A is in front of C, and one in which C is in front of A. But, before such alternatives could be specified, it would be necessary to detect the indeterminacy in the first place. In general, a scheme for detection would have to be able to infer that the relation between certain items in a propositional representation was indeterminate. Unfortunately, this requirement leads straight back to the problems of transitivity: whether the relation between certain items is determinate or indeterminate may depend entirely on whether a transitive inference is valid or invalid. Since no finite system of rules based on a propositional representation can handle this problem, it follows that no such system can detect indeterminacies, or *a fortiori* set up alternative representations when they occur. Hence, a theory of propositional representation does not yield the same equivalence class of repre-

[6]Expressions such as "in front of," in fact, have two distinct spatial senses, a deictic sense that depends on the speaker's point of view, e.g. "Stand in front of the rock," and another sense that depends on the intrinsic parts of certain sorts of object, e.g. "The river was in front of the house" (see Fillmore, 1971; Miller & Johnson-Laird, 1976, Sec. 6.1.3). This complication is not relevant to the present argument and I have otherwise ignored it.

sentations as the class yielded by the theory of mental models. The wedge remains securely in place: there is a difference between the theory of mental models and the theory of propositional representations. The way is now clear to attempt to draw some lines of demarcation and to provide some evidence in support of them.

The Characteristics of Propositional Representations

The nature of a propositional representation obviously depends on what a proposition is. One view, which has much to commend it, is a generalization of the commonplace notion that to understand a proposition is to know what the world would have to be like for it to be true. If one considers all the different ways in which the world might be, as well as the way it actually is, that is, the set of all "possible worlds," then a proposition is, in principle, either true or else false of each member of the set. Hence, a proposition can be treated as a function from the set of possible worlds onto the set of truth values.[7] A logician might, in turn, treat this function as a set of ordered pairs, each comprising a possible world and a truth value (of the proposition in that world), but this conception is highly abstract since the set of possible worlds is plainly infinite. A mental *representation* of a proposition, however, can be thought of as a function which takes a state of affairs (perceived, remembered, or imaginal) as an argument, and whose body is capable of returning a truth value. The fact that a propositional representation is a function, however, does not imply that it is automatically evaluated every time the proposition is brought to mind. It does not even imply that the function could be evaluated. Many propositions may be only partial functions, yielding no truth values for certain states of affairs; many propositions may be functions for which there is no effective computational procedure. Yet, at least some propositional representations must sometimes be evaluated and return a truth value. Otherwise, propositional representations and truth itself would be idle wheels in our minds. A view common to many proponents of a "procedural semantics" is accordingly that grasping a proposition is analogous to compiling a function,

[7]We might also wish to include possible times and other aspects of the context in the domain of the function (see Lewis, 1972). An alternative way of handling pragmatics has been proposed by Kaplan (1977). He points out that it is necessary to distinguish the context of an utterance from the circumstances of its evaluation. For example, the sentence, "I am speaking now" is true in any context in which it is uttered, but it is not thereby logically true—the speaker does not necessarily have to be speaking. There are circumstances of evaluation—possible worlds and times—in which the sentence is false. The machinery for distinguishing context and circumstances of evaluation was originally provided by Hans Kamp (1971) in his analysis of that tricky word, "now." It depends on a system of double indexing in which a set of possible worlds (and a set of times) is used twice, once for context and once for circumstances of evaluation. Since a proposition is a function from possible worlds to a truth value, the *propositional concept* expressed by a specific utterance is a function from pairs of possible worlds to a truth value, that is to say, it is a function from possible worlds representing contexts to an intension, which in turn is a function from possible worlds representing circumstances of evaluation to a truth value (see also Stalnaker, 1978).

whereas verifying it is analogous to evaluating a function. This idea can be generalized to allow other mental operations based on propositions, and to allow functional representations for questions and commands (cf. Davies & Isard, 1972; Miller & Johnson-Laird, 1976; Woods, 1967, 1979).

If a proposition is a function, then its representation is the representation of a function. The way to represent a function is to express it in a language, and, as Fodor et al., (1975, & Fodor, 1976) have argued, it is useful to think of a propositional representation as an expression in a mental language. Although we may never delineate the details of the mental language, we do know that it must have both a syntax and a semantics. It must be capable, for example, of representing conjunction, and its mental syntax could take a variety of forms, e.g. "$(\alpha K \beta)$," "$K(\alpha,\beta)$," or "$(\alpha,\beta)K$," where the Greek letters range over representations of propositions, and "K" stands for some mental token representing conjunction. Whatever form the syntax takes, it must be associated with the appropriate semantics: the function representing a conjunction will return the truth value if and only if each of the functions representing the conjoined propositions returns the value true[8]. A crucial point about the mental representation of propositions, however, is that the choice of their syntactic structure, though perhaps innately determined, is not governed by any logical or analogical considerations. It is essentially free in the same way that the discursive structure of any language is free. That is to say, although nature may have decided that conjunction is represented by a structure of the form, "$K(\alpha,\beta$," she might just as well have settled for "$(\alpha K \beta)$." It will make no difference provided that the structure receives the appropriate semantic interpretation.

The same principle of *arbitrary syntactic structure* applies to simple propositions, and in particular to the way in which their predicates and arguments are syntactically arranged. This freedom of choice is actually exercised by the designers of programming languages: they determine the syntax of the language and how it relates to its semantics; they may even elect, perhaps unwisely, to lay down the syntactic rules independently of the semantic interpretation (Hamish Dewar, personal communication)—a strategy that Chomsky (1957) also adopted in his initial studies of natural language, but which has been emphatically repudiated by students of formal semantics (e.g. Montague, 1974).

The propositional description of a complicated state of affairs may consist of a large number of propositions. The question arises as to the nature of the structural relations between them. In fact, one paradigm case of a propositional representation is simply an unordered set of expressions in some symbolic language such as the predicate calculus. Uniform theorem provers will evaluate inferences made in such a formalism, relying on procedures that will search the

[8]I have assumed here a simple truth-functional account of conjunction. Natural language is more complicated: conjunction may require a more complex connective that is not truth-functional (cf. "and then"), or conversational principles that impose a further layer of interpretation on what is fundamentally a truth-functional connective.

set for any particular atomic proposition, looking within complex propositions to check whether it is a constituent of them (Robinson, 1965, 1979). However, advocates of propositional theories have often relied on some sort of semantic network (see Anderson, 1976, 1978; Anderson & Bower, 1973; Baylor, 1971; Kintsch, 1974, Moran, 1973; Norman & Rumelhart, 1975; Palmer, 1975). In a network, propositions about the same entity are gathered together and attached to the single node for that entity. Plainly this use of structure is not essential, it simply facilitates the processes that encode or retrieve information.

The Characteristics of Mental Models

Mental models and propositional representations can be distinguished on a number of criteria. They differ pre-eminently in their function: a propositional representation is a description. A description is true or false, ultimately with respect to the world. But human beings do *not* apprehend the world directly; they possess only internal representations of it. Hence, a propositional representation is true or false with respect to a mental model of the world. In principle, this functional difference between models and propositions could be the only distinction between them: there need be nothing to distinguish them in form or content. Model-theoretic semantics often uses the device of allowing a set of sentences to be a model of itself, because various neat proofs can thereby be established. Likewise, Hintikka (1963) has formulated a semantic theory of modal logic in which the model consists of a set of sentences. PLANNER, too, uses a set of assertions in its data-base (Hewitt, 1972). However, in the case of mental models, there is reason to suppose that their form is distinct from that of propositional representations. A model *represents* a state of affairs and accordingly its structure is not arbitrary like that of a propositional representation, but plays a direct representational or analogical role. Its structure mirrors the relevant aspects of the corresponding state of affairs in the world.

Mental models of quantified assertions introduce only a minimal analogical role for structure: the use of elements to stand for individuals in a one-to-one fashion, and links to stand for identities between them. But, they possess one other feature characteristic of models as opposed to propositional representations. They represent a set of entities by introducing an arbitrary number of elements that denote exemplary members of the set. Propositional representations of the sort proposed by Fodor et al., (1975) do not contain arbitrary features, whereas models based on verbal descriptions ordinarily do so. A model representing the assertion, "Two boys kissed one girl," might contain two elements standing for the boys, and one element standing for the girl; and the links between them might have a simple propositional label standing for the relation, "kiss." There might be nothing arbitrary about this representation, yet I should still be tempted to describe it as a (hybrid) model. It has a strong analogical feature: two elements to represent two boys, one element to represent one girl. The point to be emphasized is that the inferential heuristic of maximizing the

number of identities can only apply if there are entities to be identified: it demands the use of models, because it cannot operate on a propositional representation of the sort, following Kintsch (1974, p. 18) consisting of a formula: (KISS, BOY, GIRL) & (NUMBER, BOY, TWO) & (NUMBER, GIRL, ONE).

Images, like models, have the property of arbitrariness, which has often drawn comment from philosophers. You cannot form an image of *a triangle in general*, but only of a specific triangle. Hence, if you reason on the basis of a model or image, you must take pains to ensure that your conclusion goes beyond the specific instance you considered. Hume (1896, vol I) made the point, somewhat optimistically, in this way:

> For this is one of the most extraordinary circumstances in the present affair, that after the mind has produced an individual idea, upon which we reason, the attendant custom, revived by the general or abstract term, readily suggests any other individual, if by chance we form any reasoning that agrees not with it. Thus, should we mention the word triangle, and form the idea of a particular equilateral one to correspond to it, and should we afterwards assert, *that the three angles of a triangle are equal to each other*, the other individuals of a scalenum and isosceles, which we overlooked at first, immediately crowd in upon us, and make us perceive the falsehood of this proposition . . .

The heuristic advantage of a model is balanced by the need for procedures that test the conclusions that can be derived from it—a point that is borne out by the way in which the models for quantified assertions and spatial relations have to be manipulated in order to ensure validity.

Of course models can have a richer analogical structure than those required for quantifiers. They may be two- or three-dimensional; they may be dynamic; they may take on an even higher number of dimensions in the case of certain gifted individuals. One advantage of their dimensional structure is that they can be scanned in any direction, regular or irregular, since the dimensional variables controlling the search can be determined from moment to moment by any mentally computable function. In the case of a propositional representation, as Simon (1972) points out, direct scanning can be performed only in those directions that have been encoded in the representation. Simon also draws attention to the fact that people who know perfectly well how to play tic-tac-toe (noughts and crosses) are unable to transfer their tactical skill to number scrabble, a game which is isomorphic to tic-tac-toe. He comments:

> The number scrabble evidence is particularly convincing, not merely in pointing to semantic processing, but in showing how translation to an encoding that uses isomorphs of visual linear arrays to provide the (implicit) information as to the winning combinations causes a striking change in performance. Just as the collinearity of positions can be determined on an external tic-tac-toe array by visual scanning, so collinearity can be detected on an array in the "mind's eye" by an apparently isomorphic process of internal scanning.

This process of scanning is precisely what is modelled by the spatial inference program described above.

Models and propositions are interesting to compare on the criterion of

economy. If a series of assertions are highly indeterminate, and no profound inferences have to be drawn from them, it may be more economical to remember the propositions that were asserted rather than to interpret them in the form of a model: a single propositional representation will suffice, whereas many alternative models will be needed to represent the discourse accurately. Miller (1979) makes exactly this point, and suggests that discourse may accordingly be encoded in both sorts of representation. There is certainly a limit to the extent that human beings can manipulate models in order to ensure validity, and even certain syllogisms appear to be taxing for this reason.

The theory of mental models assumes that they can be constructed on the basis of either verbal or perceptual information, though only in the former case will their construction necessitate the introduction of arbitrary assumptions. It follows that images correspond to those components of models that are directly perceptible in the equivalent real-world objects. Conversely, models may underlie thought processes without necessarily emerging into consciousness in the form of images. Models are also likely to underlie the perception of objects by providing prototypical information about them (see Roberts, 1965; Marr & Nishihara, 1976) in a form that can be directly used in the interpretation of what Marr (1976) has referred to as 'the primal sketch,' the output of lower level visual processes.

LEVELS OF DESCRIPTION

Is it really true that images and models are not necessarily equivalent to sets of propositions? That was the conclusion of the previous section, but doubtless it will be resisted by propositional theorists. There is one way in which they can sustain their objection, but only at the cost of trivializing the whole controversy. It depends on a source of much confusion in theoretical discussions, the level at which a particular theory is described. The issues can be illustrated by considering the problem of how to characterize the computer program that embodies the theory of spatial inference.

One approach is that since the program must ultimately be translated into the machine language of a computer before it can be run, we should concern ourselves with what the machine language instructions cause to happen in the machine—the shifting of bits from one location in store to another, and so on. But this approach is misguided: the details of a specific implementation should not concern us. We should not worry about the particular computer and its machine code, since the program could be executed on some very different machines, and we do not want to make a separate characterization for all these different sorts of computer. An alternative approach is provided by Scott and Strachey (1971), the pioneers of formal semantics for computing languages:

> Compilers of high-level languages are generally constructed to give the complete translation of the programs into machine language. As machines merely juggle bit patterns, the concepts of the original language may be lost or at least obscured during this passage. The purpose of mathematical semantics is to give a correct and meaningful correspondence between programs and mathematical entities in a way that is entirely independent of an implementation.

There is a very important lesson for psychologist here: their subject can be pursued independently from neurophysiology (the study of the machine and the machine code) and other disciplines that reductionists often suppose underlie psychology. The argument also provides a useful antidote to the excessive scepticism that can be induced by theorems demonstrating how one sort of representational theory can be mimicked by another. In order to try to substantiate this claim, and to clear up the confusion over levels of description, let us continue the characterization of the spatial inference program.

The Reconstruction of a Theory at a Lower Level of Description

"It works by building up a two-dimensional array that satisfies the premises given to it." This description of the program is informal, but at a high level, the level of "psychological" discourse. You may wonder how exactly an array is represented by the programming language. It is, in fact, a data structure of one or more dimensions in which the elements can be accessed and updated by giving appropriate coordinates. (An array can also be represented by a function in POP-10, which permits it to be specified by a rule rather than an explicit table.) A programmer needs to know no more: one can write procedures for manipulating arrays simply by thinking of them as n-dimensional spaces where each location is specified by an n-tuple of integers. A student of the "psychology" of computers, however, may be curious about the invisible machinery that makes such an array possible. Its representation in the computer does not involve an actual physical array of locations in core store. That is quite unnecessary. Indeed, the physical embodiment of an array is irrelevant. What matters is that it should *function* as an array, that is, it has a set of addresses that are functionally equivalent to an array, its elements can be accessed as in an array, and its contents displayed or printed out in the form of an array. A psychological description should accordingly be a functional one.

Consider a program for spatial inference in which an assertion such as, "A is on the right of B" is represented by the following formulae: AT(A, 1, 6), AT(B, 1, 2) and the general procedure for verification works by looking for sequences of ordered pairs of integers as parts of such formulae. In order to verify the above assertion, it starts with B and its associated pair (1, 2), and then looks for formulae corresponding to the sequence: (1, 3), (1, 4), (1, 5) . . . up to some arbitrary number. If the program finds A associated with a pair of integers

in the series (which of course it will do in this example), then the assertion is true; otherwise, it is false. The series is defined by the procedure representing "on the right of," which freezes in the appropriate values for the incremental parameters of the verification process.

It should be clear that the whole of the original theory of spatial inference can be reconstructed in this way, even to the extent of coping with the problems of transitivity. Indeed, many adherents of propositional theories may wish to claim that a propositional theory of spatial inference has here been constructed that counters all the earlier criticisms. They would be wrong; but wrong in a way that is most instructive. The construction of the new propositional theory of spatial inference is in reality simply a reconstruction of the original theory *at a lower level of description*. The whole of the propositional apparatus, the ordered pairs of integers, the definition of "on the right of" in terms of incremental values of parameters, is parasitic upon the unacknowledged presence of a spatial array. Perhaps it is easiest to grasp this point by asking oneself how such a system could have been set up in the first place, how it could have been learned, and where the definition of "on the right of" could have come from. The program *functions* as though it uses an array, and one seen from a particular viewpoint, too.

Any Psychological Theory Can Be Based (Vacuously) on Propositional Representations

In general, a model is only a model at a certain level of description: that level at which it functions as one. A listing of the original spatial inference program in machine code is a level of description that obscures the program's use of models. The new "propositional" theory is similarly a redescription of the old theory at a level that obscures its reliance on models; it is a description that could well pass as a slightly more detailed account of how to set up and manipulate arrays in a certain programming language.

There is, of course, nothing inconsistent about calling such a representation a propositional theory. Indeed, the controversy can be resolved in a still more direct way to support the view that any plausible theory of any psychological phenomenon is propositional. If you accept Church's thesis that any "effective procedure" can be computed by a Turing machine, then it follows that the psychological theory, granted the reasonable criterion that it is intended to characterize an effective procedure, can also be computed by a Turing machine. This device, however, can be completely described by a set of propositions—linear strings of symbols from a defined alphabet—that characterize the rules governing its change of state and behaviour as a function of its current state and input (see e.g. Minsky, 1967, p. 106 *et seq*). The only form of representation required by a Turing machine is a tape divided into cells in which there is either a symbol, "1", or a blank: everything that can be computed at all can be computed on the basis of this preeminently propositional representation by a device that can

be specified propositionally in exactly the same code. To characterize a theory as propositional is accordingly to say nothing of any empirical consequence.

How to Give the Notion of a "Propositional Representation" an Empirical Content

If the term "propositional representation" is to have empirical content, then it must be constrained in some way. Hence, the view espoused earlier in this paper is that a propositional representation is based on symbols that correspond in a one-to-one fashion with the lexical items of natural language—a view proposed for other reasons by Kintsch (1964) and Fodor et al., (1975). It is unclear whether those who advocate propositional representations for images intend to make a trivial point of the sort that can be established directly by a reduction to machine code or by the parallel conceptual reduction to a Turing machine. What is noteworthy, however, is that they have freely introduced propositions expressing polar coordinates, vectors, and other spatial notions. Such concepts can obviously be expressed in scientific language, but there is no corresponding terminology for them in the ordinary language of simple shapes that they are being used to analyze. Hence, by the criterion introduced to ensure that "propositional representation" has an empirical content, what a theorist proposes in such cases is, not a propositional theory, but a reconstruction of a theory of mental models at a lower level of description.

The purpose of introducing lists, strips, arrays, and a whole variety of data structures and facilities into high-level programming languages is to enable the programmer to forget about the detailed implementation of something that can be functionally specified. Plainly these representations do not increase the computational power of the language or necessarily improve the actual running of the programs. What they do facilitate is the programmer's task of developing and testing programs. On the plausible supposition that the mind possesses the capability of devizing programs for itself (see Miller, Galanter & Pribram, 1960), precisely the same advantage is obtained from high level procedures for manipulating both models and propositional representations. My next task, having shown how they can be usefully distinguished in principle, is to examine some evidence that distinguishes them in practice.

EXPERIMENTS ON MENTAL MODELS AND PROPOSITIONAL REPRESENTATIONS

Ordinary discourse is often indeterminate. If you were to come across the following passage in a story, then you would probably form only a rather vague idea of the actual spatial layout:

> I opened the door and went in. The room was at the corner of the building and on my right there was a long window overlooking the bay. A plain but tasteful table ran the

> length of the room and there were chairs on either side. A large colour television set stood flickering on one side of the table beneath the window, and on the other side there was a small safe, its door ajar. At the head of the table facing the door, Willis sat deep in thought, or so it seemed. The room was very quiet. And Willis was very quiet, frozen in a posture of unnatural stillness.

A few details would stand out—the open safe, the TV, and the corpselike appearance of the man—but you would be unlikely to have gone beyond the description to have figured out whether the safe was on the right hand side or the left hand side of the room from where the narrator viewed it. Yet, if you read the passage again with the aim of determining the answer to this question, then you can form a very much more complete mental picture of the room. There accordingly appear to be different levels of representation, and the hypothesis that I wish to advance is that they differ in kind. The result of a superficial understanding is a propositional representation: a fairly immediate translation of the discourse into a mental language. A more profound understanding leads to the construction of a mental model which is based on the propositional representation, but which can rely on general knowledge and other relevant representations in order to go beyond what is explicitly asserted.

We have carried out a number of experiments in order to investigate this hypothesis. In one experiment (see Ehrlich, Mani & Johnson-Laird, 1979) the subjects listened to three sentences about the spatial relations between four common objects, e.g.:

The knife is in front of the spoon
The spoon is on the left of the glass
The glass is behind the dish

and then attempted to make a drawing of the corresponding layout using the names of the objects. We assumed that in order to carry out this task the subjects would construct a mental model of the layout as they heard each premise. Hence, we predicted that the task would be straightforward if the premises came in an order (like those in the example above) that permitted a model to be built up continuously, but that the task would be very much harder if the premises were arranged in a discontinuous order:

The glass is behind the dish
The knife is in front of the spoon
The spoon is on the left of the glass

in which the first two assertions refer to no item in common. In this case a subject must either construct two models and then combine them in the light of the third premise or else simply represent the premises in a propositional form until the time comes to make the drawing. The results reliably confirmed the prediction: 69% of the drawings based on continuous premises were correct, whereas only 42% of the drawings based on discontinuous premises were correct. It might be argued that the subjects only ever use a propositional representation of the premises and that it is easier to form such a representation from continuous premises

than discontinuous premises. One suggestive piece of evidence to the contrary is the relative ease of a third sort of ordering of the premises:

The spoon is on the left of the glass
The glass is behind the dish
The knife is in front of the spoon

in which the third assertion has nothing in common with the second. This ordering was not significantly harder than the continuous premises, yielding 60% of correct drawings. The point to be noted is that although the second and third premises are discontinuous, they always contain at least one item that would already have been represented in a mental model.

A further experiment corroborated the existence of two modes of representation. The subjects again listened to three assertions about the spatial relations between some common objects. They described either a determinate layout (as in the previous examples) or else an indeterminate one, e.g.

The knife is in front of the spoon
The spoon is on the left of the glass
The fork is on the right of the spoon

where the relation between the glass and the fork is undetermined. The subjects' task was rather different in this experiment. After each set of premises, they were shown a diagram of a layout and they had to decide whether or not it satisfied the description in the premises. I assumed that the subjects would be inhibited from forming a model of the indeterminate premises since they might easily form the "wrong" one, i.e. one that failed to correspond with the picture, though it was consistent with the premises. Hence, I predicted that they would use a propositional representation and would accordingly be better able to remember the premises. At the end of the experiment, the subjects received an unexpected recognition test of their memory for each set of premises. Each test involved the original premises, a paraphrase of them that had the same meaning, and two sets that differed in meaning from the originals. The major result was that my prediction was completely false: determinate premises were reliably better recalled than indeterminate premises. Not one of the twenty subjects that Kannan Mani tested went against this trend. However, there was an interesting incidental finding. If a subject remembers the meaning of the original premises, then he will pick out the originals and the paraphrases of them before he picks out the other two confusion items. In this case, it is possible to work out the likelihood that he can remember the original premises *verbatim*, picking them out prior to the paraphrases. This probability was 63% for the indeterminate problems, which was significantly better than chance; it was 57% for the determinate premises, which was not significantly better than chance.

A natural explanation for these results rests on the assumption that mental models are constructed from propositional representations. It follows, of course, that a greater amount of processing is required to construct a mental model than

to construct a propositional representation. We have found independently that other things being equal the greater the amount of processing the better an item will be remembered: the phenomenon applies both to individual words (Johnson-Laird, Gibbs & de Mowbray, 1978) and to sentences (Johnson-Laird & Bethell-Fox, 1978). It follows that in general mental models should be better remembered than propositional representations—as indeed the experiment established. However, a propositional representation is directly obtained from discourse: if it is recalled, then there should be a good chance that the original sentences on which it is based should be recalled *verbatim;* whereas a mental model, through relatively easy to recall, contains no direct information about the sentences on which it is based: even if it is recalled, there is no guarantee that they will be recalled *verbatim*.

Work in other laboratories provides similar support for two modes of representation for discourse. Scribner and Orasanu (1979), for example, examined their subjects memory for syllogistic premises, comparing trials on which a subject had answered a question about them that required an inference to be made with trials where the question did not require an inference to be made. They found that adults and older children tended to remember the premises more accurately when they had made an inference from them—a finding that corroborates the hypothesis that inferences depend on the construction and testing of mental models.

CONCLUSIONS

Language can be used to talk about real, imaginary, and hypothetical states of affairs: domains for which logicians and philosophers have often advocated a "possible worlds" semantics. However, a psychologically plausible account of such discourse cannot be based on an infinite set of possible worlds, but, as I have suggested elsewhere, should be founded on the mental ability to construct representations of alternative states of affairs to those that actually obtain (see Johnson-Laird, 1978). The same mode of representation can be used to represent beliefs about others' beliefs, and in general propositional attitudes about others' propositional attitudes (see Johnson-Laird, 1979). A crucial characteristic of discourse, whether conversation or text, is reference and referential continuity. The referents of expressions depend in part on context, and, as Alan Granham and I have recently argued, following in the steps of Karttunen (1976), Stenning (1978), and others, the real context of an utterance consists of the mental models of the current conversation that the speaker and the listener maintain. These models represent the relevant individuals, events, and relations. They also represent what is known about the other participants' state of mind. Hence, a speaker chooses his words partly on the basis of his model of the listener's discourse model; and a listener interprets these remarks partly on the basis of his model of the speaker's discourse model. A number of referential phenomena depend critically on the characteristics of mental models, as we were at pains to demonstrate (Johnson-Laird & Garnham, 1979). For example, what really controls the use of a definite description is, not uniqueness in the world, but uniqueness in a model. Hence, when a speaker remarks:

The man who lives next door drives to work

then the definite description should not be taken to imply that there is only one man living next door to the speaker. It designates the only neighbor who is relevant in the context.

Likewise, the most important characteristic underlying the coherence of texts is continuity of reference—a feature that was explicitly manipulated in the experiments on spatial inference. A simple illustration of this point is to consider the following text (after Rumelhart, 1976):

> Margie was holding tightly to the string of her beautiful new balloon. Suddenly, a gust of wind caught it and carried it into a tree. It hit a branch and burst. Margie cried and cried.

As Rumelhart points out, if the sentences are put into random order, their cohesion is destroyed:

> It hit a branch and burst. Suddenly a gust of wind caught it and carried it into a tree. Margie cried and cried.Margie was holding tightly to the string of her beautiful new balloon.

Obviously, the causal sequence of events is disrupted. Yet, if the original noun-phrases are replaced by ones that reestablish continuity of reference, the cohesion of the randomized text is greatly enhanced:

> Margie's beautiful new balloon hit a branch and burst. Suddenly, a gust of wind caught it and carried it into a tree. Margie cried and cried. She was holding tightly to the string of the balloon.

Moreover, if continuity of reference is destroyed by replacing the original noun-phrases with new ones, even in the original order the passage ceases to be cohesive:

> Margie was holding tightly to the string of her beautiful new balloon. Suddenly, a gust of wind caught a newspaper and carried it into a tree. A cup hit a wall and broke. John cried and cried.

There are of course other aspects of coherence, but none is likely to be so preeminent as referential continuity: if a text never refers to the same entity more than once, it rapidly acquires the characteristics of a telephone directory rather than a passage of prose.

Mental models evidently play a part in a variety of phenomena other than those that I have considered in detail in this paper. They appear to have a unifying role to play in Cognitive Science. To return to the three questions with which I began, first, there are indeed distinctions to be drawn between propositional representations and mental models:

1. A propositional representation is a description of a state of affairs, which may be true or false. It is evaluated with respect to a model representing that state of affairs.
2. The initial, and sometimes perhaps only, stage in comprehension consists in creating a propositional representation: a linear string of symbols in a mental language that has an arbitrary (and as yet unknown) syntactic structure and a lexicon that closely corresponds to that of natural language. This representation can be used to construct a mental model, which represents information analogically: its structure is a crucial part of the representation. Models can also be set up directly from perception.
3. A propositional representation encodes determinate and indeterminate information in a uniform way, and makes no use of arbitrary assumptions. A mental model of the state of affairs described in a proposition may embody a number of arbitrary assumptions since language is inherently vague. Indeterminate information is encoded either by utilizing a set of alternative models, or else by incorporating a propositional representation in a 'hybrid' way. The two sorts of representation do not necessarily yield the same equivalence classes, and hence there is no guarantee that a theory embodying one can be made to mimic the other.
4. A model represented in a dimensional space can be directly constructed, manipulated, or scanned, in any way that can be controlled by dimensional variables. A propositional representation lacks this flexibility and can be directly scanned only in those directions that have been laid down between the elements of the representation.

Second, there are likewise distinctions to be drawn between a decompositional semantics and a set of meaning postulates:

1. Insofar as language relates to the world, it does so through the action of the mind, and in particular through its innate ability to construct models of reality. The extension of such words as *right* and *left* is specified by decompositional procedures that operate on the general procedures for constructing and evaluating mental models. Meaning postulates are not intended to perform this function and contain no machinery for doing the job.
2. The logical properties of a term need not be specified within a procedural definition, rather they are emergent properties of that definition. Only in this way can such phenomena as the vagaries of transitivity be explained: they are not an intrinsic part of the meaning of the term, but properties that emerge in the construction of mental models. Meaning postulates, however, as rules that explicitly specify the logical properties of terms, and the logical relations between them.

Third, it is possible to account for the psychological principles underlying deductive reasoning:

1. The capacity to draw inferences rests fundamentally on the ability to construct and to manipulate mental models. The major inferential heuristic for quantified assertion can only be stated for a domain of individuals: it can be summarized in a principle of economy aimed to keep models simple by identifying individuals playing different roles. Inferential ability also depends on submitting putative conclusions to logical test by attempting to destroy the model on which they are based while maintaining its faithfulness to the premises.
2. Insofar as human beings have internal rules of inference that operate on propositional representations, they derive them from invariant outcomes in the manipulation of models e.g. whenever *a is greater than b* and *b is greater than c* then the resulting model is always such that *a is greater than c*.

3. The origins of formal logic as an intellectual discipline are likely to be found in the awareness of potential error as a result of failing to carry out the test procedures exhaustively, and in a self-conscious attempt to externalize such test procedures. Once a set of valid inferences has been determined in this way, an attempt can be made to formalize rules that characterize the set.

These conclusions have been based partly on the results of experiments and partly from ideas derived from developing computer programs. The reader will recall that at the outset I stressed the need for theories in cognitive science that are both coherent and correspond to the facts. The time has come to consider the arguments that favour the use of experiments, programs, and their methodological combination.

A Methodological Moral

There are many reasons for carrying out psychological experiments, and by no means all of them need concern the elucidation of mental phenomena. You may be primarily concerned with the practical application of your findings, as, for example, in the design of a more legible typeface, in the development of better procedures for teaching foreign languages, or in tests of the reliability of police identity parades. Such studies can be useful without directly revealing anything about mental processes. But even those investigations that have that as their primary aim can differ strikingly in how they achieve it. Experiments in cognitive psychology typically address specific hypotheses or sets of alternative hypotheses, and are designed to allow you to come to a decision about them. However, a view that is common amongst devotees of artificial intelligence is that psychological experiments are a waste of time because the theoretical alternatives are not sufficiently articulated to need to worry about experimental tests between them. The business of providing such theories can be pursued within AI on the basis of general knowledge and common observation. After a number of years of arguing with Max Clowes and other vigorous champions of AI, I confess to considerable sympathy with this view. One sort of experiment, however, still seems eminently worthwhile: it is that relatively rare variety that yields a significant pattern of results such as the figural effect, or the greater memorability of determinate descriptions, that is totally unexpected to you. Although experiments may be useful in corroborating your hypotheses, or in showing that they survive potentially falsifying tests, their major value is in causing a significant change in the way in which you think about a problem. An experiment should astonish you. Unfortunately, there are no methodological principles that can guarantee you success; but if you obtain a surprising result, then it may lead to an insight that could have been acquired in no other way.

Computer programming is too useful to cognitive science to be left solely in the hands of the artificial intelligenzia. There is a well established list of advantages that programs bring to a theorist: they concentrate the mind marvelously; they transform mysticism into information processing, forcing the theorist

to make intuitions explicit and to translate vague terminology into concrete proposals; they provide a secure test of the consistency of a theory and thereby allow complicated interactive components to be safely assembled; they are "working models" whose behavior can be directly compared with human performance. Yet, many research workers look on the idea of developing their theories in the form of computer programs with considerable suspicion. The reason for the suspicion is complex. In part it derives from the fact that any large-scale program intended to model cognition inevitably incorporates components that lack psychological plausibility. To take an example from a masterly program, Winograd's (1972) procedure for recovering the referents of pronouns is manifestly implausible.[9] Certain aspects of any such program must be at best principled and deliberate simplifications or at worst *ad hoc* patches intended merely to enable the program to work. The remedy, which I have struggled to express on a number of occasions (see e.g. Johnson-Laird, 1977), is *not* to abandon computer programs, but to make a clear distinction between a program and the theory that it is intended to model. For a cognitive scientist, the single most important virtue of programming should come not from a finished program itself, or what it does, but rather from the business of developing it. Indeed, the aim should be neither to simulate human behavior—often a species of dissimulation—nor to exercise artificial intelligence, but to force the theorist to think again. As Jackson Pollock remarked in a different context: the end product does not matter so much as the process of making it. The development of small-scale programs that explore part of a general theory can be a genuinely dialetical process leading to new ideas both about the theory and even about how to test it experimentally. Students of human reasoning would long ago have discovered that it is unnecessary to postulate a mental schema for transitivity, or other internalized rules of inference, if only they had attempted to devise some simple inferential programs.

Cognitive science does not exist: it is necessary to invent it. A crucial part of its invention may prove to be a methodological synthesis of experimental psychology and artificial intelligence. On the one hand, the experimenter's concept of truth exerts a dangerous pull in the direction of empirical pedantry, where the only things that count are facts, no matter how limited their purview. On the other hand, the programmer's concept of truth exerts a dangerous pull in the direction of systematic delusion, where all that counts is internal consistency, no matter how remote it is from reality. One way ahead is to develop general and comprehensive theories of the mind, couched in the theoretical vernacular of the discipline; to make explicit models of at least parts of them in the form of computer programs; and to combine this process with a regime of experimental investigation. This route may lead us to a discipline that is a general science of the mind.

[9]Til Wykes (1979), however, has found that very young children do interpret pronouns in a "syntactic" manner closely resembling the principles embodied in Winogard's programs.

ACKNOWLEDGMENT

This research was supported by a grant from the Social Science Research Council (G.B.) Many individuals have helped wittingly and unwittingly in the preparation of this paper. I am particularly indebted to Bruno Bara, Anne Cutler, Kate Ehrlich, Alan Garnham, Gerald Gazdar, Dave Haw, Steve Isard, Ewan Klein, Christopher Longuet-Higgins, George Miller, Don Norman, Stan Peters, Stuart Sutherland, Patrizia Tabossi, and Eric Wanner for many useful ideas and criticisms.

REFERENCES

Anderson, J. R. *Language, memory and thought*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Anderson, J. R. Arguments concerning representations for mental imagery. *Psychological Review*, 1978, *85*, 249–277.

Anderson, J. R. & Bower, G. H. *Human associative memory*. New York: V. H. Winston & Sons, 1973.

Baylor, G. W. Programs and protocol analysis on a mental imagery task. *First International Joint Conference on Artificial Intelligence*, 1971.

Begg, I. & Denny, J. P. Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning errors. *Journal of Experimental Psychology*, 1969, *81*, 351–354.

Berlin, B. & Kay, P. *Basic colour terms: Their universality and evolution*. Berkeley and Los Angeles: University of California Press, 1969.

Beth, E. W. *Aspects of modern logic*. Dordrecht, Holland: Reidel, 1971.

Bledsoe, W. W., Boger, R. S. & Henneman, W. H. Computer proofs of limit theorems. *Artificial Intelligence*, 1972, *3*, 27–60.

Boole, G. *An investigation of the laws of thought*. London: Macmillan, 1854.

Bower, G. H. Mental imagery and associative learning. In L. Gregg (Ed.) *Cognition in learning and memory*. New York: Wiley, 1972.

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

Brooks, L. The suppression of visualization by reading. *Quarterly Journal of Experimental Psychology*, 1967, *19*, 280–299.

Brooks, L. Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, 1968, *22*, 349–368.

Bugelski, B. R. Word and things and images. *American Psychologist*, 1970, *25*, 1002–1012.

Byrne, B. Item concreteness vs. spatial organization. *Memory and cognition*, 1974, *2*, 53–59.

Carnap, R. *Meaning and necessity: A study in semantics and modal logic*. Second edition. Chicago: University of Chicago Press, 1956.

Ceraso, J. & Provitera, A. Sources of error in syllogistic reasoning. *Cognitive Psychology*, 1971, *2*, 400–410.

Chapman, I. J. & Chapman, J. P. Atmosphere effect re-examined. *Journal of Experimental Psychology*, 1959, *58*, 220–226.

Chase, W. G. & Clark, H. H. Mental operations in the comparison of sentences and pictures. In L. W. Gregg (Ed.) *Cognition in learning and memory*. New York: Wiley, 1972.

Chomsky, N. *Syntactic structures*. The Hague: Mouton, 1957.

Clark, H. H. & Clark, E. V. *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich, 1977.

Cohen, M. R. & Nagel, E. *An introduction to logic and scientific method*. London: Routledge & Kegan Paul, 1934.

Collins, A. M. & Quillian, M. R. Experiments on semantic memory and language comprehension. In L. W. Gregg (Ed.) *Cognition in learning and memory*. New York: Wiley, 1972.

Cooper, L. A. Mental rotation of random two-dimensional shapes. *Cognitive Psychology*, 1975, *7*, 20–43.

Craik, K. *The nature of explanation*. Cambridge: Cambridge University Press, 1943.
Davies, D. J. M. & Isard, S. D. Utterances as programs. In D. Michie (Ed.) *Machine intelligence 7*. Edinburgh: Edinburgh University Press, 1972.
Dennett, D. C. *Content and consciousness*. New York: Humanities Press, 1969.
Ehrlich, K., Mani, K. & Johnson-Laird, P. N. Mental models of spatial relations. Mimeo, Centre for Research on Perception and Cognition, Laboratory of Experimental Psychology, University of Sussex, 1979.
Erickson, J. R. A set analysis theory of behavior in formal syllogistic reasoning tasks. In R. Solso (Ed.) *Theories in cognitive psychology: The Loyola symposium*. Potomac, MD: Lawrence Erlbaum Associates, 1974.
Erickson, J. R. Research on syllogistic reasoning. In R. Revlin & R. E. Mayer (Eds.) *Human reasoning*. Washington, DC: V. H. Winston & Sons, 1978.
Falk, G. Interpretation of imperfect line data as a 3-dimensional scene. *Artificial Intelligence*, 1972, *3*, 101–144.
Fillmore, C. J. Toward a theory of deixis. Paper delivered to the Pacific conference on contrastive linguistics and language universals. University of Hawaii, 1971.
Fillmore, C. J. An alternative to checklist theories of meaning. *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, 1975, 123–131.
Fodor, J. A. *The language of thought*. Hassocks, Sussex: Harvester Press, 1976.
Fodor, J. D. *Semantics: Theories of meaning in generative grammar*. Hassocks, Sussex: Harvester Press, 1977.
Fodor, J. D. The mental representation of quantifiers. Paper presented to the Symposium on Formal Semantics and Natural Language, University of Texas at Austin, 1979.
Fodor, J. D., Fodor, J. A. & Garrett, M. F. The psychological unreality of semantic representations. *Linguistic Inquiry*, 1975, *4*, 515–531.
Funt, B. V. WHISPER: A problem-solving system utilizing diagrams. *Fifth International Joint Conference on Artificial Intelligence*, 1977, 459–464.
Galton, F. *Inquiries into human faculty and its development*. London: Dent, 1928 (originally published 1880).
Gelernter, H. Realization of a geometry-theorem proving machine. In E. A. Feigenbaum & J. Feldman (Eds.) *Computers and thought*. New York: McGraw-Hill, 1963.
Guyote, M. J. & Sternberg, R. J. A transitive-chain theory of syllogistic inference. Technical Report No. 5, Department of Psychology, Yale University, 1978.
Hayes, J. R. On the function of visual imagery in elementary mathematics. In W. G. Chase (Ed.) *Visual information processing*. New York: Academic Press, 1973.
Hewitt, C. Description and theoretical analysis of PLANNER. MIT AI Laboratory Report MIT-AI-258, 1972.
Hintikka, J. The modes of modality. *Acta Philosophica Fennica*, 1963, *16*, 65–82.
Hintikka, J. Quantifiers vs. Quantification theory. *Linguistic Inquiry*, 1974, *5*, 153–177.
Holyoak, K. J. The form of analog size information in memory. *Cognitive Psychology*, 1977, *9*, 31–51.
Hume, D. *A treatise of human nature*. Vol I. Edited by L. A. Selby-Bigge. Oxford: Clarendon, 1896.
Inhelder, B. & Piaget, J. *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structure*. London: Routledge & Kegan Paul, 1958.
Inhelder, B. & Piaget, J. *The early growth of logic in the child: Classification and seriation*. London: Routledge & Kegan Paul, 1964.
Johnson-Laird, P. N. Models of deduction. In R. J. Falmagne (Ed.) *Reasoning: representation and process in children and adults*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.
Johnson-Laird, P. N. Procedural semantics. *Cognition*, 1977, *5*, 189–214.
Johnson-Laird, P. N. The meaning of modality. *Cognitive Science*, 1978, *2*, 17–26.
Johnson-Laird, P. N. Formal semantics and the psychology of meaning. Paper presented at the Symposium on Formal Semantics and Natural Language, University of Texas at Austin, 1979.

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. & Garnham, A. Descriptions and discourse models. *Linguistics and Philosophy*, in press.
Johnson-Laird, P. N. & Steedman, M. J. The psychology of syllogisms. *Cognitive Psychology*, 1978, *10*, 64–99.
Johnson-Laird, P. N. Gibbs, G. & de Mowbray, J. Meaning, amount of processing, and memory for words. *Memory and Cognition*, 1978, *6*, 372–375.
Kamp, H. Formal properties of "Now." *Theoria*, 1971, *37*, 227–273.
Kaplan, D. Demonstratives: an essay on the semantics, logic, metaphysics and epistemology of demonstratives and other indexicals. Paper presented at the meeting of the Pacific division of the American Philosophical Association, March, 1977.
Karttunen, L. Discourse referents. In J. D. McCawley (Ed.) *Syntax and semantics Vol. 7: Notes from the linguistic underground*. New York: Academic Press, 1976.
Katz, J. J. & Fodor, J. A. The structure of a semantic theory. *Language*, 1963, *39*, 170–210.
Katz, J. J. & Nagel, R. Meaning postulates and semantic theory. *Foundations of Language*, 1974, *2*, 311–340.
Kieras, D. Beyond pictures and words: Alternative information-processing models for imagery effects in verbal memory. *Psychological Bulletin*, 1978, *85*, 532–554.
Kintsch, W. *The representation of meaning in memory*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.
Kneale, W. & Kneale, M. *The development of logic*. Oxford: Clarendon, 1962.
Kosslyn, S. M. Information representation in visual images. *Cognitive Psychology*, 1975, *7*, 341–370.
Kosslyn, S. M. Can imagery be distinguished from other forms of internal representation? Evidence from studies of information retrieval time. *Memory and Cognition*, 1976, *4*, 291–297.
Kosslyn, S. M. & Pomerantz, J. R. Imagery, propositions and the form of internal representations. *Cognitive Psychology*, 1977, *9*, 52–76.
Lakoff, G. Linguistic Gestalts. *13th Regional Meeting, Chicago Linguistic Society*, 1977.
Lewis, D. General semantics. In D. Davidson & G. Harman (Eds.) *Semantics of natural language*. Dordrecht, Holland: Reidel, 1972.
Lyons, J. *Semantics*. Vols. 1 and 2. Cambridge: Cambridge University Press, 1977.
Marr, D. Early porcessing of visual information. *Philosophical Transactions of the Royal Society of London*, Series B, 1976, *275*, 483–519.
Marr, D. & Nishihara, H. K. Representation and recognition of the spatial organization of three-dimensional shapes. MIT AI Laboratory Memorandum, 337, 1976.
Martin, E. The psychological unreality of quantificational semantics. In W. Savage (Ed.) *Minnesota studies in philosophy of science*, Vol. 9. Minnesota, in press.
Mazzocco, A., Legrenzi, P. & Roncato, S. Syllogistic inference: The failure of the atmosphere effect and the conversion hypothesis. *Italian Journal of Psychology*, 1974, *2*, 157–172.
Miller, G. Images and models, similes and metaphors. In A. Ortony (Ed.) *Metaphor and thought*. Cambridge: Cambridge University Press, 1979.
Miller, G. A. & Chomsky, N. Finitary models of language users. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.) *Handbook of mathematical psychology*, Vol. II. New York: Wiley, 1963.
Miller, G. A., Galanter, E. & Pribram, K. *Plans and the structure of behavior*. New York: Holt Rinehart & Winston, 1960.
Miller, G. A. & Johnson-Laird, P. N. *Language and perception*. Cambridge, Mass: Harvard University Press; Cambridge, Cambridge University Press, 1976.
Minsky, M. *Computation: Finite and infinite machines*. Englewood Cliffs, N.J.: Prentice-Hall, 1967.
Montague, R. *Formal philosophy*. Edited by R. H. Thomason. New Haven: Yale University Press, 1974.
Moran, T. P. The symbolic nature of visual imagery. *Third International Joint Conference on Artificial Intelligence*, 1973, 472–477.

Moyer, R. S. Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception and Psychophysics*, 1973, *13*, 180–184.
Norman, D. A., & Rumelhart, D. E., Memory and knowledge. In D. A. Norman, D. E. Rumelhart & the LNR Research Group, *Explorations in cognition*. San Francisco: Freeman, 1975.
Osherson, D. Logic and models of logical thinking. In R. J. Falmagne (Ed.) *Reasoning: Representation and process in children and adults*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.
Paivio, A. *Imagery and verbal processes*. New York: Holt, Rinehart & Winston, 1971.
Paivio, A. Perceptual comparisons through the mind's eye. *Memory and Cognition*, 1975, *3*, 635–647.
Paivio, A. Images, propositions and knowledge. In J. M. Nicholas (Ed.) *Images, perception and knowledge*. Dordrecht, Holland: Reidel, 1977.
Palmer, S. E. Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In D. A. Norman, D. E. Rumelhart & the LNR Research Group, *Explorations in cognition*. San Francisco: Freeman, 1975.
Putnam, H. Is semantics possible? In H. Putnam (Ed.) *Mind, language and reality: Philosophical papers*, Vol. 2. Cambridge: Cambridge University Press, 1975. (Originally published 1970.)
Putnam, H. The meaning of 'meaning'. In H. Putnam (Ed.) *Mind, language and reality: Philosophical papers*, Vol. 2, Cambridge; Cambridge University Press, 1975.
Pylyshyn, Z. W. What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 1973, *80*, 1–24.
Reiter, R. A semantically guided deductive system for automatic theorem-proving. In *Third International Joint Conference on Artificial Intelligence*, 1973.
Revlin, R. & Leirer, V. O. The effect of personal biases on syllogistic reasoning: Rational decisions from personalized representations. In R. Revlin & R. E. Mayer (Eds.) *Human reasoning*. Washington, DC: V. H. Winston & Sons, 1978.
Revlis, R. Two models of syllogistic reasoning: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behavior*, 1975(a), *14*, 180–195.
Revlis, R. Syllogistic reasoning: Logical decisions from a complex data base. In R. J. Falmagne (Ed.) *Reasoning: Representation and process in children and adults*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975(b).
Roberts, L. G. Machine perception of three-dimensional solids. In I. J. T. Tippett et al., (Eds.) *Optical and electro-optical information processing*. Cambridge, Mass: MIT Press, 1965.
Robinson, J. A. A machine-oriented logic based on the resolution principle. *Journal of Association for Computing Machinery*, 1965, *12*, 23–41.
Robinson, J. A. *Logic, form and function: Mechanization of deductive reasoning*. Edinburgh: Edinburgh University Press, 1979.
Rosch, E. Natural categories. *Cognitive Psychology*, 1973, *4*, 328–350.
Rumelhart, D. E. Notes on a schema for stories. In D. G. Bobrow & A. Collins (Eds.) *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.
Schank, R. C. *Conceptual information processing*. Amsterdam: North-Holland, 1975.
Scott, D. & Strachey, C. Toward a mathematical semantics for computer languages. *Proceedings of the Symposium on Computers and Automata*. Polytechnic Institute of Brooklyn, April, 1971.
Sells, S. B. The atmosphere effect: An experimental study of reasoning. *Archives of Psychology*, 1936, *29*, 3–72.
Scribner, S. & Orasanu, J. Syllogistic recall. Mimeo. Rockefeller University, 1979.
Shepard, R. N. Form, formation and transformation of internal representations. In R. Solso (Ed.) *Information processing and cognition: The Loyola symposium*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.
Shepard, R. N. The mental image. *American Psychologist*, 1978, *33*, 125–137.
Simon, H. A. What is visual imagery? An information processing interpretation. In L. W. Gregg (Ed.) *Cognition in learning and memory*. New York: Wiley, 1972.

Sloman, A. Interactions between philosophy and artificial intelligence: The role of intuition and non-logical reasoning in intelligence. *Artificial Intelligence*, 1971, *2*, 209–225.
Smith, E. E., Shoben, E. J. & Rips, L. J. Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 1974, *81*, 214–241.
Stalnaker, R. C. Assertion. In P. Cole (Ed.) *Syntax and semantics*. Vol. 9: *Pragmatics*. New York: Academic Press.
Stenning, K. Anaphora as an approach to pragmatics. In M. Halle, J. Bresnan, & G. A. Miller (Eds.) *Linguistic theory and psychological reality*. Cambridge, Mass.: M.I.T. Press.
Sternberg, R. J. & Turner, M. E. Components of syllogistic reasoning. Technical Report No. 6. Department of Psychology, Yale University, 1978.
Störring, G. Experimentelle Untersuchungen über einfache Schlussprozesse. *Archiv ges. Psychologie*, 1908, *11*, 1–127.
Winograd, T. *Understanding natural language*. New York; Academic Press.
Wittgenstein, L. *Philosophical investigations*. Oxford: Blackwell, 1953.
Woods, W. A. Semantics for a question-answering system. Mathematical linguistics and automatic translation report, NSF-19, Harvard computational Laboratory, 1967.
Woods, W. A. Procedural semantics. Paper presented at the Symposium on Formal Semantics and Natural Language, University of Texas at Austin, 1979.
Woodworth, R. S. & Sells, S. B. An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, 1935, *18*, 451–460.
Wykes, T. D.Phil. dissertation. Laboratory of Experimental Psychology, University of Sussex, 1979.