1

MODELS OF DEDUCTION

P. N. Johnson-Laird
University of Sussex

> Beyond the obvious facts that he has at some time done manual labour, that he takes snuff, that he is a Freemason, that he has been in China, and that he has done a considerable amount of writing lately, I can deduce nothing else.
>
> *Adventures of Sherlock Holmes*
> SIR ARTHUR CONAN DOYLE

It has become a truism that whatever else formal logic may be it is not a model of how people make inferences. It perhaps provides a standard, an ideal template, against which to assess the validity of inferences; and this view has a considerable appeal until one considers just which particular logic should play the role of the paragon. Logic is not a monolithic enterprise. There are many logics. Indeed, there are an infinite number of modal logics, a mere branch of the discipline. Although the different branches may be independent of one another, a choice of logic for, say, the temporal expressions of natural language is quite likely to have implications for a choice of logic for, say, such terms as "necessary" and "possible." Many of the different linguistic suburbs—tense markers, modal terms, connectives, quantifiers, and so on—are, for a logician, independent areas of interest; and, despite the surge of interest in them (e.g., Montague, 1970; Parsons, 1972), there is as yet no single comprehensive logic of natural language (just as there is as yet no complete grammar). It may even be supposed that no single coherent logic can suffice for all the ways in which language is used (van Fraassen, 1971). Yet, in spite of this reservation, a central question endures: are there any general ways of thinking that human beings follow when they make deductions?

The tenor of much recent psychological work provides a decidedly negative answer. The content of a reasoning problem seems to matter just as much as its logical structure, determining not only how a problem is represented but also the sorts of inferences that are made. Wason and Johnson-Laird (1972) have found evidence of such effects in a variety of tasks, ranging from the testing of hypotheses to reasoning with propositions. Such findings coincide with an increasingly popular conception of inference within artificial intelligence (AI).

One of the original aims of trying to program computers to carry out intelligent activities was to devise automatic methods of theorem proving. The intention was to devise programs that would both translate natural language into expressions of the predicate calculus and operate on these expressions with general theorem-proving procedures. Because it had long been established that there could be no algorithm for proof within the predicate calculus, much of this work was of a heuristic nature. Very often, however, methods devized in a heuristic spirit turned out to be more powerful. Some methods even guaranteed, if a theorem could be proved, to find a proof sooner or later. [There was, alas, no guarantee that the method would reveal, where appropriate, that it was impossible to derive a given conclusion; and this deficiency was the heart of Church's (1936) proof that there could be no general decision procedure for the predicate calculus.] It follows that general proof procedures have one glaring disadvantage: no matter how long they grind away at a problem, there is no way of knowing whether or not they will ultimately come up with a solution. If there is proof they will sooner or later discover it; but if there is no proof, they may never find out. Therefore, the impetus behind such sophisticated methods as the resolution principle and the hyperresolution principle (Robinson, 1965, 1966) was to increase the efficiency of programs so that they would find proofs, where they existed, within a reasonable amount of computing time. However, there is another difficulty with general proof procedures. Before they can go to work on a problem, it has to be represented in the predicate calculus; and it turns out that the business of translating natural language expressions into their appropriate symbolic form is extremely taxing. Ordinary language does not wear its logical heart on its sleeve, and there are often surprising divergences between the superficial form of an expression and its underlying logic. Once again, there is no known general procedure for carrying out correct translations (see Johnson-Laird, 1970).

One reaction to these difficulties has been to try a different tack. Instead of representing putative theorems in the notation of the predicate calculus and then grinding away at them with a general proof procedure, they are represented as programs. When such programs are executed they control the process of trying to discover the proof. This idea forms the basis of Hewitt's (1970) theorem-proving language PLANNER, which has been exploited so successfully in Winograd's (1972) program for understanding natural language. One obvious advantage of the method is that it allows information and deductive procedures, pertinent to the particular content of a problem, to be taken into account in the theorem-proving process. The system therefore gains greatly in efficiency; and, if the psychological experiments are to be believed, it is also a better model of the human deductive process. There is accordingly a general tendency in both psychological and AI circles to emphasize goal-oriented inferential procedures. This tendency is also evident in recent work on uniform proof procedures, especially in the development of predicate logic as a programming language (Kowalski, 1973). The aim of this chapter, however, is to attempt to redress the balance and to examine to what extent there may be general principles of thought that are independent of any particular problem domain. In examining this topic, three main sorts of inference are discussed: lexical reasoning, propositional reasoning, and reasoning with quantifiers. A few new experimental results are presented but the emphasis is on developing models of deduction.

LEXICAL REASONING

Perhaps the most obvious sort of inference—so obvious, in fact, that it is hardly noticed in ordinary discourse—involves simple relations between such lexical items as nouns, verbs, adjectives, etc. The meanings of words are, of course, often interrelated, and a speaker's knowledge of such interrelations acts very much as a smoothing oil to help the inferential machinery revolve. If, for example, a law states that all dog owners must pay a tax, then from the statement "He owns a poodle," it may readily be inferred, "He must pay the tax." From a formal point of view such an inference is invalid: it lacks the premise, "All poodles are dogs." In daily life, however, human beings do not behave like logicians; they know that poodles are dogs, and they exploit this knowledge without a thought to the canons of formal logic.

Logicians have tended to ignore this aspect of practical reasoning, although the device of meaning postulates (see Carnap, 1956; Bar-Hillel, 1967) was developed to deal with the logical consequences of the semantic relations between words. Psychologists, however, have recently been very active in investigating such relations under the guise of studying "semantic memory." A few salient points of these studies are perhaps worth delineating (for a more extensive review, see Johnson-Laird, 1974). The overwhelming majority of studies have concerned nouns and, in particular, the relation of class inclusion between them. They have shown that where there is a hierarchy of class inclusion, such as *poodle: dog: animal,* it may take time to grasp the transitivity of the relation. It may take time, in other words, to recover the fact that a poodle is an animal. A variety of competing theories have been proposed to explain this phenomenon (e.g., Collins & Quillian, 1969; Landauer & Meyer, 1972; Schaeffer & Wallace, 1970). None of these theories is entirely satisfactory, if only because there are occasions in which the transitive relation is easier to retrieve than its constituents, e.g., "a poodle is a mammal" is harder to verify than "a poodle is an animal" even though mammals are included in the class of animals (Rips, Shoben, & Smith, 1973). Nevertheless, it remains true that not all semantic relations are obtainable from the lexicon with the same ease. It is necessary to work, albeit for a few hundredths of a second, to retrieve more recondite relations. And such work, of course, has the logical form of an inference. Indeed, when Graham Gibbs and I gave subjects an inferential task, involving such material as

> Flowers are killed by this chemical spray.
> Therefore, roses are killed by this chemical spray.

We obtained results comparable to more conventional studies of semantic memory. In certain cases (e.g., *python: snake: reptile*) a transitive inference took longer than inferences involving its constituents; in other cases (e.g., *pine: conifer: tree*) a transitive inference took less time than the inferences involving its constituents.

What sort of semantic relations are there between the meanings of words? The simple relations include synonymy (e.g., *automobile–car*), antonymy (e.g., *man–woman*), and class inclusion (e.g., *dog–animal*); and these relations give rise to corresponding relations between sentences in which the words occur. It is no accident that studies of semantic memory have concentrated on class inclusion: it is a potent relation because it leads to transitive inferences. Similar transitive hierarchies can be generated by the relation of spatial inclusion and sometimes by the relation *part of*. However, the obvious source of transitive relations is comparative adjectives, e.g., *"larger than," "better than,"* and expressions of the general form *"more x than."* It is a simple matter to infer that if *a* is larger than *b*, and *b* is larger than *c*, than *a* is larger than *c*. However, so much controversy has arisen over various details of the process (see Huttenlocher & Higgins, 1971; Clark, 1971) that certain broader issues have been ignored in the quest to explain experimental findings. One such issue, the representation of the transitivity of a relational term, is considered below.

There are other patterns of lexical inference apart from transitivity. An intransitive relation (**R**), for instance, permits an inference of the form

$$a\mathbf{R}b \textit{ and } b\mathbf{R}c \qquad \therefore \textit{not } (a\mathbf{R}c)$$

The relation *"next in line to"* is obviously intransitive because if *a* is next in line to *b*, and *b* is next in line to *c*, then it follows that *a* is not next in line to *c*. A nontransitive relation, however, permits neither the transitive nor the intransitive inference; for example, if *a* is next to *b*, and *b* is next to *c*, then nothing follows about whether *a* is next to *c*—the items may be arranged in a circular fashion or in a line.

Another aspect of the logic of relations concerns symmetry. A relation is symmetrical if it permits an inference of the form

$$a\mathbf{R}b \qquad \therefore b\mathbf{R}a$$

The relation *next to* is symmetrical because if *a* is next to *b*, then it follows that *b* is next to *a*. A relation is asymmetrical if it permits an inference of the form

$$a\mathbf{R}b \qquad \therefore \textit{not } (b\mathbf{R}a)$$

The relation *on the right of* is clearly asymmetrical. A relation is nonsymmetrical if it permits neither of these inferences; the relation *nearest to* is clearly nonsymmetrical.

There are still other logical properties of relational terms, such as reflexitivity and connectivity, but their role in ordinary language appears to be negligible. However, because transitivity and symmetry are independent attributes, the lexicon already contains a variety of relations. They are exemplified in Table 1 by a set of spatial expressions.

The semantic representation of relational terms must include information about their transitivity and symmetry. For example, the representation of "beyond" must permit a transitive inference, whereas the representation of "nearest to" must prevent it. What has yet to be determined is precisely how this representation is effected. It is possible that each relational term has stored with it in the mental lexicon a simple tag indicating a transitivity value and another tag indicating a symmetry value. Where a term **R** is tagged as transitive, it permits an inference of the form

$$a\mathbf{R}b \textit{ and } b\mathbf{R}c \qquad \therefore a\mathbf{R}c$$

This conception evidently requires inference schemata to be separately specified as adjuncts to the lexicon. A more plausible system, however, renders the transitivity of a relation self-evident from its semantic specification,

TABLE 1
Spatial Expressions as Exemplars of the Logical Sorts of Binary Relations in Ordinary Language

	Transitive	Symmetric
In the same location as [as *x* as]	+	+
Beyond [more *x* than]	+	–
Not beyond [not more *x* than]	+	o
Next in line to	–	+
Directly on top of	–	–
Nearest to	–	o
Next to	o	+
On the right of	o	–
At	o	o

+transitive = transitive; –transitive = intransitive; o transitive = nontransitive;
+ symmetric = symmetric; – symmetric = asymmetric; o symmetric = nonsymmetric.

i.e., the conclusion *a***R***c* would be self-evident from the joint representation of *a***R***b* and *b***R***c*. A way of representing quantified statements (e.g., "All bankers are prudent men") with just this property is described below. The best evidence for this sort of representation for simple relational terms is provided by inference about spatial relations. Consider the following inference:

The box is on the right of the chair.
The ball is between the box and the chair.

Therefore, the ball is on the right of the chair.

The most likely way in which such an inference is made involves setting up an internal representation of the scene depicted by the premises. This representation may be a vivid image or a fleeting abstract delineation—its substance is of no concern. The crucial point is that its formal properties mirror the spatial relations of the scene so that the conclusion can be read off in almost as direct a fashion as from an actual array of objects. It may be objected, however, that such a depiction of the premises is unnecessary, that the inference can be made by an appeal to general principles, or rules of inference, which indicate that items related by *between* must be collinear, etc. However, this view—that relational terms are tagged according to the inference schemata they permit—founders on more complex inferences. An inference of the following sort, for instance, seems to be far too complicated to be handled without constructing an internal representation of the scene

> The black ball is directly beyond the cue ball. The green ball is on the right of the cue ball, and there is a red ball between them.
>
> ---
>
> Therefore, if I move so that the red ball is between me and the black ball, then the cue ball is on my left.

Even if it is possible to frame inference schemata that permit such an inference to be made without the construction of an internal representation, it is most unlikely that this approach is actually adopted in making the inference. The only rules of inference that are needed are a procedure for setting up a joint representation of separate assertions and a procedure for interrogating the joint representation. Much of the work can be done by the semantic information in the lexicon; and the same principle of allowing lexical information to specify directly the logic of a relation can apply equally well to abstract terms with meanings that are difficult or impossible to visualize directly. With concrete or abstract terms, the structure of a joint representation is isomorphic to its logic in a way that is exemplified below in the analysis of quantified inference.

Perhaps the most potent source of lexical inferences is the set of verbs of a language. The same sorts of relation obtain between them as between other lexical items—relations such as antonymy (e.g., *open–shut*), and class inclusion (e.g., *assassinate–murder–kill*). However, verbs can often be used to express relations between several arguments, rendering even the simple analysis of a relation and its converse (e.g., *buy–sell*) a complicated matter. The additional complexity of verbs does, indeed, lead to some interesting problems. Consider the following typical sorts of inference that depend on the meanings of verbs:

> Pat forced Dick to refrain from swearing.
> ∴ Dick refrained from swearing.
> ∴ Dick did not swear.

> Sam managed to prevent Dean from pretending to be naive.
> ∴ Sam prevented Dean from pretending to be naive.
> ∴ Dean did not pretend to be naive.
> ∴ Dean was not naive.

> John regretted that he had no chance to lie.
> ∴ John had no chance to lie.
> ∴ John did not lie.

These examples illustrate ways in which inferences may be drawn about the truth or falsity of a clause occurring as the complement of a verb. For example, if someone *forces* x to do z, then it may be inferred that x did z, whereas if someone *prevents* x from doing z, then it may be inferred that x did not do z. The validity of these inferences depends on the meaning of the verbs and, in particular, on the fact that their semantic representation contains a conjunction of separate elements of meaning. The essentially conjunctive nature of many verbs is perhaps more evident in the semantics of causal verbs:

> He moved the table.
> ∴ He did something and consequently the table moved.
>
> He showed us the picture.
> ∴ He did something and consequently we could see the picture.
>
> He gave her the book.
> He had the book and he did something and consequently she had the book.

The logic of these inferences can largely be captured by treating the concept of cause as a special sort of conjunction (see Miller & Johnson-Laird, 1975). Of course, it is very much more than a simple conjunction and seems to involve the following conditions in ordinary language (*pace* Dowty, 1972):

a caused *b* if and only if: (i) *a* happened;
(ii) *b* happened;
(iii) it is not possible for *a* to happen and *b* not to happen afterward.

The important point, however, is that it is seldom necessary to take the analysis so far in order to explain the inferential properties of causal verbs. A conjunctive analysis usually suffices.

In short, lexical reasoning is noteworthy not for the novelty of its patterns of inference but for the speed and smoothness with which its inferences occur. They are sometimes so immediate as to pass unnoticed. Their patterns include simple relational schemata and, especially in the case of verbs, simple propositional inferences.

PROPOSITIONAL REASONING

It has been realized since antiquity that one source of inferential relations is the manner in which sentences, or clauses, are combined. Language provides a variety of connectives, such as "and," "or," and "if," that can be used to combine clauses expressing propositions, e.g., "The boat has gone, *or else* it has been sunk *and* no trace of it can be found." To know what these connectives mean is tantamount to knowing how to draw certain inferences on the basis of the formal patterns in which they occur. For example, a speaker can hardly be said to have fully grasped the meaning of "or" unless he appreciates the validity of an inference such as

The boat has gone or else it has been sunk.
It has not been sunk.
Therefore, it has gone.

The logic of connectives has been most fully explored in the development of the propositional calculus. There are, in fact, a variety of different calculi and a variety of different ways of formulating them. However, a brief and informal exposition of the standard calculus will suffice here. If lower case letters are allowed to range over propositions, then the calculus can be formalized by specifying what counts as a well-formed formula, and by stating a set of axioms such as

1. $(p \text{ or } p) \rightarrow p$
2. $p \rightarrow (p \text{ or } q)$
3. $(p \text{ or } q) \rightarrow (q \text{ or } p)$
4. $(p \rightarrow q) \rightarrow [(r \text{ or } p) \rightarrow (r \text{ or } q)]$

where the arrow is a sign for material implication. In addition to the axioms, two rules of inference are necessary. The first rule of inference allows new formulas to be generated by substituting any well-formed formula for a propositional variable in an expression, and the second rule of inference, the so-called law of *modus ponens,* may be stated as follows:

From a formula A together with a formula *if A then B,* the formula B may be deduced.

It is fairly simple to show that these axioms and rules suffice to derive all the formulas that are true on the logical interpretations of the connectives.

What does such a system state about the reasoning of intelligent but logically naive persons? The answer must surely be: very little. However, it is worth dwelling on the system for a moment because at least one influential psychologist, Piaget, has used it as the basis of a model of reasoning (Beth & Piaget, 1966) and because the contrast between it and ordinary deduction is instructive.

Among the more obvious difficulties of using the propositional calculus as a model of ordinary deduction is the fact that its connectives can stand only between fully fledged propositions. In ordinary language simple constituents, such as noun phrases, may be linked by a connective. A sentence such as

> Mark and Anne are excellent riders

is easily translated into a form suitable for the calculus:

> Mark is an excellent rider and Anne is an excellent rider.

However, there is no comparable procedure for dealing with such sentences as:

> Mark and Anne make a splendid couple.

This sentence must be treated as a single proposition. Another difficulty, of course, is that the calculus is truth functional: the meaning of its connectives is defined purely in terms of the truth value they give to a complex proposition as a function of the truth values of its constituents. The multifarious connectives of ordinary language (e.g., *because, before, although*) cannot be completely captured in a purely truth-functional calculus. Nor, indeed, can the logic of commands or questions be immediately accommodated within its essentially assertive framework.

A further divergence between logical calculi and the inferential machinery of everyday life concerns their respective functions. Calculi are devised primarily for deriving logical truths. The aim of practical inference, however, is not to prove theorems but to pass from one contingent statement to another. Therefore, practical inference is likely to involve few, if any, axioms but a relatively large number of rules of inference. A formulation of the calculus that is therefore more appropriate abandons axioms in favor of a system of rules analogous to Gentzen's method of "natural" deduction, an approach that has had some influence in the development of theorem-proving programs (e.g., Amarel, 1967; Reiter, 1973). A system of natural deduction involves the specification of rules of inference in a schematic form. The rule of *modus ponens,* for example, is stated in the following schema:

$$\frac{A \qquad \textit{If } A \textit{ then } B}{\therefore B}$$ [1]

where the premises appear above the line and the conclusion appears below it. A parsimonious system, of course, stipulates the minimum number of such schemata from which all the others can be derived. For instance, negation and disjunction may be taken as primitive, inference rules stipulated for them, and the remaining connectives simply defined in terms of negation and disjunction. From a psychological point of view, however, it would be foolish to seek parsimony at the expense of plausibility. What is needed is a set of psychologically basic patterns of inference.

Any decision about whether a pattern of inference is psychologically basic is ultimately an empirical matter. It is necessary to find out whether the inference is in the immediate repertoire of mature but logically naive persons. An inference schema can hardly be considered as basic if most people are incapable of carrying it out or can only do so in a matter of minutes, subsequently giving a detailed resumé of a whole chain of deductions they have carried out to make the inference. Unfortunately, there is not enough evidence to determine the definitive set of basic patterns of inference. What can be done, however, is to build up a plausible first approximation to it, taking care not to include any inferential schema known to cause difficulty to logically naive subjects. The fact that an inference is feasible for the majority of people suggests that it is basic but is hardly a decisive proof: the inference may be the result of combining several other inferences. Only those inferences that seem *prima facie* to be basic are therefore included in the following set, but in many cases the final decision must depend on further investigations.

Some extremely simple inferences are considered first. It is obvious that from a clause expressing the meaning A one can immediately deduce a clause expressing the same meaning, A. (This way of writing in terms of clauses expressing meanings is excessively cumbersome; from now on I shall write simply of propositions, although it must not be forgotten that I am dealing with inferences expressed in natural language.) Similarly, the conclusion A can be immediately deduced from a proposition of the form A or A. These inferences are summarized in the following schemata:

$$\frac{A}{\therefore A} \tag{1}$$

$$\frac{A \text{ or } A}{\therefore A} \tag{2}$$

Although inferences of this sort may sometimes rely on complex lexical inferences, their structure is very simple and can hardly be derived from anything more basic. The question is whether these inferences may not be too trivial to be useful. In fact, they do have a role to play, and a model of propositional inference is defective without them.

The same may be said about some further schemata. The first pair permit a proposition to be inferred from its occurrence in a conjunction:

$$\frac{A \text{ and } B}{\therefore A} \tag{3a}$$

$$\frac{A \text{ and } B}{\therefore B} \tag{3b}$$

The second pair permit a disjunction to be inferred from either one of its constituents:

$$\frac{A}{\therefore A \text{ or } B} \tag{4a}$$

$$\frac{A}{\therefore B \text{ or } A} \tag{4b}$$

The third pair permit a conjunction to be inferred from the independent occurrence of its constituents:

$$\frac{A \qquad B}{\therefore A \text{ and } B} \tag{5a}$$

$$\frac{A \qquad B}{\therefore B \text{ and } A} \tag{5b}$$

And the final pair permit negated conjunctions to be deduced:

$$\frac{A \text{ and not } -B}{\therefore \text{ not both } A \text{ and } B} \tag{6a}$$

$$\frac{\text{not } -A \text{ and } B}{\therefore \text{ not both } A \text{ and } B} \tag{6b}$$

A real problem with these simple patterns of inference is to find a suitable way to curb their productivity. As a number of authors have recently pointed out, there are constraints on what can reasonably be expressed in the form of a conjunction or a disjunction. It may be true, for example, that boys eat apples, and that Mary threw a stone at the frog, but the conjunction

Boys eat apples and Mary threw a stone at the frog

is, as Lakoff (1971) argues, barely acceptable. It is customary to suit an utterance to its context, and this principle applies to the relations between clauses as well as to the relations between sentences. Hence, if a speaker follows one clause with another explicitly specifying what seems to have been taken for granted, then he creates an extremely odd conjunction, e.g.,

> John ran out of the house and he got out of bed (Johnson-Laird, 1969a).
> All of John's children are bald and John has children (Karttunen, 1973).

The existence of constraints on the topics of conjunctions and disjunctions can hardly be doubted. Indeed, the constraints on "but" proved sufficient for that conjunction to be used by Bendix (1966) as the basis of a semantic test. However, there is no adequate explication of a complete set of constraints. One solution is therefore to do away with the simple inference schemata that give rise to the free combination of propositions in conjunctions and disjunctions. Unfortunately, it is impossible to do without these rules of inference. They are needed in order to make such deductions as

It is frosty.
If it is foggy or frosty, then the game will be canceled.

Therefore, the game will be canceled.

For the time being, schemata (1) to (6) shall be called "auxiliary inferences," for reasons that will become clear when the method of curbing their power is described.

In contrast to the auxiliary inferences, there are a number of primary patterns of inference that have no restrictions placed on them. There is among them the familiar pattern exemplified in the following inference:

John is intelligent or he is rich.
He is not rich.

Therefore, he is intelligent.

There is good reason to suppose that its underlying schema

$$\frac{A \text{ or } B \qquad \text{not} - A}{\therefore B} \tag{7a}$$

$$\frac{A \text{ or } B \qquad \text{not} - B}{\therefore A} \tag{7b}$$

is basic. A study by Hill (cited in Suppes, 1965) found that 82% of a sample of 6-year-old children were able to make the inference correctly.

Johnson-Laird and Tridgell (1972) found that it led to errors only when the negative occurred in the disjunctive premise, e.g.,

John is intelligent or he is not rich.
He is rich.

With premises of this sort, some of their adult subjects inferred that John was not intelligent, whereas other subjects considered that no conclusion followed from the premises. Such a finding suggests, however, not that the schema is intrinsically difficult but that an unusual placement of negative information can disturb its smooth execution.

The patterns of inference in (7) are valid both for an inclusive disjunction, where both constituent propositions can be true, and for an exclusive disjunction, where this possibility is ruled out. There is a further rule of inference that applies only to exclusive disjunctions, e.g.,

Either Mary is a plagiarist or else she is a genius (but not both).
She is a genius.

Therefore, she is not a plagiarist.

The real force of this inference derives from the exclusivity of the two propositions in the disjunction. It is therefore plausible that the basic inferential schema should be formulated in the following way:

$$\frac{\textit{Not both A and B} \qquad A}{\therefore \ \textit{not} -B} \tag{8a}$$

$$\frac{\textit{Not both A and B} \qquad B}{\therefore \ \textit{not} -A} \tag{8b}$$

The main candidate for a basic pattern of inference involving the conditional is *modus ponens:*

$$\frac{A \qquad \textit{If A then B}}{\therefore \ B} \tag{9}$$

There is considerable evidence to suggest that this schema is basic, whereas a closely related pattern, known as *modus tollendo tollens*, is not (see Wason & Johnson-Laird, 1972). The latter inference has the following form:

$$\frac{\textit{Not} -B \qquad \textit{If A then B}}{\therefore \ \textit{not} -A}$$

Intelligent subjects can make inferences of this sort but they tend to do so with a greater difficulty than with *modus ponens* and it is natural to suppose that they are carrying out a sequence of inferential steps rather than a single inference. They may, in fact, be arguing in the following way:

> If the safe is locked, then this light is on.
> This light is not on.
>
> ---
>
> Suppose the safe is locked.
> It follows then that the light is on (by *modus ponens*).
> But the light is not on (from the premise).
> Therefore, the assumption leads to an impossible, contradictory state of affairs.
> Therefore, the assumption is false: the safe is not locked.

This sort of argument is, of course, a *reductio ad absurdum* and requires an inferential schema of the form

$$\frac{A \textit{ implies } (B \textit{ and not } -B)}{\therefore \textit{ not } -A} \tag{10}$$

It is certainly true that logically naive persons can argue by a *reductio* (Evans, 1972); and it can be accepted as basic instead of *modus tollendo tollens,* although a completely convincing justification for this choice cannot be established at present.

There are two subsidiary points about the *reductio* schema in (10). First, a conditional may equally well have been used in place of the implication, because if one proposition can be derived from another, then this fact can be expressed by a conditional

$$\frac{A \textit{ implies } B}{\therefore \textit{ If } A \textit{ then } B} \tag{11}$$

Second, where a *reductio* is used to establish the falsity of a negative proposition, it is necessary to be able to eliminate the resulting double negation, e.g., "It isn't the case that 5 is not odd" becomes "5 is odd." A simple schema makes this elimination possible:

$$\frac{\textit{not not } -A}{\therefore A} \tag{12}$$

Its seeming simplicity, however, may be deceptive. At least one school of logicians, the intuitionists, have excluded this rule from their canon. These logicians, represented by Heyting (1956), are primarily worried about certain sorts of mathematical reasoning. In particular, they are concerned with inferences involving infinite sets and argue that such inferences must involve constructive and intuitive principles. They claim that it is not sufficient, in order to demonstrate the existence of a mathematical property, to show that its universal denial leads to a contradiction. Hence, the intuitionists reject the law of the excluded middle, i.e., the principle that either a proposition or its negation is true. They consequently reject the related principle for eliminating double negations. The relation between the intuitionist and the classical calculus of propositions is not so straightforward as might be imagined; Gödel (1933) has shown that the classical calculus can nevertheless be treated as contained within the intuitionistic calculus! It shall be assumed here, however, that the elimination of double negations is a feature of ordinary reasoning.

The dozen inference schemata that have now been stated constitute a plausible set of psychologically basic patterns of deduction. There are other forms of inference that, although probably not basic, are well within the competence of most people, and a way must certainly be found for incorporating them into the model. One example of such an inference is the *simple dilemma*, e.g.,

The President is dishonest or he is incompetent.
If the President is dishonest, then he will be forced to resign.
If the President is incompetent, then he will be forced to resign.
―――――――――――――――――――――――――――――――
Therefore, the President will be forced to resign.

Such an argument places an adversary literally on the horns of a dilemma, because no matter which of the alternatives he chooses from the initial disjunction, he is forced to accept the same conclusion. The rhetorical force of such arguments was, indeed, recognized by Cicero (see Kneale & Kneale, 1962; p. 178). However, the argument can be considered, for psychological purposes, as merely a special case of a more general pattern of inference:

$$\frac{A \text{ or } B \qquad \text{If } A \text{ then } C \qquad \text{If } B \text{ then } D}{\therefore\ C \text{ or } D}$$

If C is substituted for D in this schema, then the derived conclusion becomes C or C, and this conclusion, in turn, is immediately reducible to C by an auxiliary inference. It is feasible that the simple dilemma is derived in this way from the more general argument. A comparable chain of inference, which indeed is not logically independent of the general dilemma, is the so-called *hypothetical syllogism*. This pattern of inference makes explicit the transitivity of conditional propositions

$$\frac{\text{If } A \text{ then } B \qquad \text{If } B \text{ then } C}{\therefore\ \text{If } A \text{ then } C}$$

Obviously, a way must be found to insure that the model permits such inferences to be drawn.

There are a number of simple equivalences that cannot be established by the present rules of inference, e.g., "Neither John can come nor Mary can leave" is equivalent to "John can't come and Mary can't leave." It would be a simple matter to introduce schemata for them, but it would be slightly odd to treat such relations by way of rules of propositional inference. A more sensible solution is to assume that inferences based on synonymy are just special cases of the schema $A / \therefore A$, and that synonymy is established on linguistic grounds. In other words, the complete mechanism of lexical inference is at the disposal of the propositional machinery. Indeed, it may be said that reasoning with propositions is simply a matter of grasping the meaning of those lexical items that happen to be connectives. This view is certainly suggested by considering the question of how patterns of inference are acquired in the first place. Where, indeed, do they come from? And how are they fitted together into a coherent system? One plausible conjecture is that the basis of the whole process is the acquisition of the truth conditions of the various connectives. Perhaps this notion should be broadened to include the extensional conditions for commands and questions, etc.; however, for the sake of simplicity only the truth conditions of assertions shall be considered here.

In the standard formalizations of the propositional calculus, including the method of natural deduction, nothing explicit is said about the truth conditions of the various connectives. When the calculus has been axiomatized, a *theorem* is defined as a formula that can be derived from the axioms by the rules of inference. This sort of definition, and the equivalent sort for the method of natural deduction, is essentially formal: it provides purely syntactic criteria, pertaining solely to the manipulation of symbols, for what counts as a theorem. It is also possible, however, to define a *valid* formula—a formula that is a logical truth. The usual way of carrying out such a definition, as demonstrated below, is to set up a semantical model in the spirit of Tarski (1956). This model may be treated as a mathematical entity involving certain marks on paper, such as "T" and "F," or alternatively it may be interpreted so as to involve certain concepts, such as truth and falsity. Logically speaking, a crucial issue is whether the calculus is complete. A proof of its completeness amounts to showing that the set of formulas derivable from the axioms is one and the same as the set of valid formulas defined by the semantical model. It is a fairly simple matter to show that the standard formalizations of the propositional calculus are, indeed, complete.

The issue of completeness has no obvious counterpart in the psychological modeling of inference. The reason it disappears is, in my view, simply that the whole system is semantically based. The conditions in which conjunctions, disjunctions, etc., are true and false are learned and, from these conditions, the basic patterns of inference are derived. A competent adult therefore has at his disposal both the inference schemata and their underlying semantic basis.

The development of a semantical model for the propositional calculus typically involves the following sorts of conditions:

1. A negative proposition, *not A*, is true if and only if *A* is false.
2. A conjunction, *A and B*, is true if and only if *A* is true and *B* is true.
3. A disjunction, *A or B*, is true if and only if *A* is true or *B* is true.

There are two difficulties, however, one linguistic and the other metalinguistic, in regarding such principles as part of a psychological basis for the semantics of connectives.

The metalinguistic difficulty is caused simply by the lack of any obvious psychological correlate of the logician's distinction between an object language and a metalanguage. In the truth conditions above, the reader will have noticed that the connectives themselves actually occur as part of their own definitions. Logically, there is nothing objectionable in this practice because the conditions for the object language connectives are being stated in a quite separate language, the metalanguage. However, it is rather unfortunate that this metalanguage turns out to be ordinary English. If it is claimed that learning the truth conditions of ordinary connectives amounts to learning rules of the sort illustrated above, then a vicious circle is created because these rules presuppose a knowledge of the meaning of ordinary connectives. This problem seems to have been overlooked by many of the linguists engaged in setting up semantical bases for natural language (e.g., Keenan, 1970). Its solution presumably involves some more abstract form of mental representation for metalinguistic information about natural language.

The linguistic difficulty with the semantical rules concerns the interpretation of conditional statements and it goes to the heart of the problem of using the propositional calculus as the basis of a psychological model. Conditionals in ordinary language are, of course, capable of a great many different sorts of interpretation. They may be used to state temporal, causal, or logical relations between propositions. It is only relatively rarely that they fit the requirements of the calculus, for example, in conveying a material implication. Such an implication is true provided its antecedent is false or provided its consequent is true, e.g., "If this picture isn't by Picasso, then it's by Braque." The majority of everyday conditionals, however, are not rendered true merely by establishing that their antecedents are false. A statement such as "If this picture is by Picasso, then it was painted in 1910" is simply irrelevant—neither true nor false—if the picture in question turns out not to be by Picasso. It is one of the fictions of the propositional calculus as a model of ordinary deduction that propositions always have a truth value. The calculus does not permit truth-value gaps.

The distinction between a material implication and a conditional with a truth-value gap may be considered trivial. In fact, however, it leads to a clear divergence between the logical calculus and ordinary inference. The following bizarre inference for instance, counts as valid if conditional statements are treated as material implications:

> You can't both hate Mailer and admire him.
> If you hate Mailer, then you will soon give up reading his work.
> If you admire Mailer, then you will read his entire works.
>
> ---
>
> Therefore, if you hate Mailer you will read his entire works, or if you admire Mailer you will soon give up reading his work.

The validity of the argument turns simply on the fact that a material implication is true whenever its antecedent is false, and one of the two conditional antecedents in the conclusion must be false according to the first premise.

A more plausible account of conditionals should permit them to lack a truth value. The semantical rule for the conditional connective might then be stated as

> A conditional *if A then B* has a truth value if and only if *A* is true; and it is true if and only if *B* is true.

The trouble with this analysis, however, is that it leaves out of account the strong intuition that there should usually be some sort of connection between *A* and *B* in order for a conditional of the form *If A then B* to be true. It also, of course, runs entirely counter to the evaluation of many conditionals that, *ex hypothesi,* have antecedents that are false or as yet unfulfilled, e.g., such counterfactual conditionals as "If Hitler had been a successful painter, then World War II would not have occurred," and such conjectural conditionals as "If the Russians invade West Germany, then World War III will occur." Evidently, these conditionals are not truth functional.

What happens when you evaluate a conditional appears to depend on whether or not you already assent to its antecedent, and whether or not you already assent to its consequent. As Ramsey (1950) pointed out long ago, if you have no view about the antecedent, then for the sake of argument you add it to your set of beliefs and then consider whether or not the consequent is true. This judgment is in turn reflected back to your evaluation of the conditional as a whole. In contrast, if you happen already to believe that the antecedent is true, then your task is simply to evaluate the consequent. A problem arises, however, if you happen already to believe that the antecedent is false. Its solution, as Stalnaker (1968) has shown, is simply to add the antecedent to your beliefs for the sake of argument and then to make minimal changes in your other beliefs in order to maintain consistency. The way is then clear for you to evaluate the consequent in the light of these hypothetical assumptions. Your decision about the consequent must obviously take into account any views you have about a causal connection, or any other sort of connection, between the antecedent and consequent.

Your prior attitude to the consequent of the conditional obviously plays a part in these proceedings. If you happen already to believe that it is true and can continue to do so in the light of your treatment of the antecedent, then you may very well assent to the conditional even if there appears to be little connection between its antecedent and consequent. In contrast, if you happen already to believe that the consequent is false and continue to do so in the light of your treatment of the antecedent, then you must evaluate the conditional as false. Finally, if you have no prior views about the consequent, your evaluation of it must depend on what connections, if any, you establish between it and the antecedent.

This unified approach to the evaluation of conditionals can be precisely formulated in terms of a semantical model resting on the notion of a "posible state of affairs" (see Stalnaker, 1968). Its details are not of concern here because their psychological realization is more plausibly thought of as a set of procedures for making assumptions, eliminating inconsistencies, and so on. The crux of the matter is simply that although such connectives as the conditional are not truth functional, their role in deductions can nevertheless be modeled by inference schemata. Rules of inference may, indeed, be learned by considering truth conditions but they can be applied without reference to them.

The present set of inference schemata are still deficient as a psychological model, because the power of the auxiliary rules of inference has yet to be curbed. The method that shall be adopted here depends on the idea of modeling the machinery of ordinary deductive inference by a set of computer-like procedures, an idea that goes back to the work of Miller, Galanter, and Pribram (1960), and Newell, Shaw, and Simon (see Newell & Simon, 1972). This conception of a program of interrelated deductive procedures has many advantages, not least that it provides a straightforward way of making auxiliary inferences dependent on primary inferences. The basic idea is simply that an auxiliary inference can be made only as

TABLE 2
The Set of Auxiliary Inferences

	Premises	Conclusions
1.	*A*	*A*
2.	*A* or *A*	*A*
3.	*A* and *B*	*A*
	A and *B*	*B*
4.	*A*	*A* or *B*
	A	*B* or *A*
5.	*A B*	*A* and *B*
	A B	*B* and *A*
6.	*A* and not *B*	Not both *A* and *B*
	Not *A* and *B*	Not both *A* and *B*
11.	*A* implies *B*	If *A* then *B*
12.	Not not *A*	*A*

a necessary precursor to a primary inference: it is an auxiliary aid that prepares the way for a primary inference. An example of the process may clarify the relation between the two sorts of inference.

Suppose that the deductive program is given the goal to make an inference from two premises of the form:

$$\frac{A \textit{ and } C}{\textit{If } A \textit{ then } B}$$

It is obvious that no initial conclusion can be derived using a primary inference because instead of a simple categorical premise *A*, required by *modus ponens,* there is only the conjunction *A and C*. However, because this conjunction is linked to a constituent of the conditional, i.e., they contain the proposition *A* in common, a subgoal can be set up to derive *A* from the first premise, *A and C*. The primary inferences are no help here but an auxiliary inference allows the inference to be made. Once the auxiliary inference is made, the way is clear to deduce the conclusion *B* using a primary inference.

It may be helpful at this point to summarize the two sorts of inference schemata. The auxiliary inferences are stated in Table 2. The primary inference schemata, however, are more conveniently summarized in the form of Table 3. This format enables procedures to be devised that allow access to the table by way of a premise or by way of a conclusion.

TABLE 3
The Set of Primary Inferences

Complex premise	Categorical premise: *A*	Not *A*	*B*	Not *B*
7. *A* or *B*		*B*		*A*
8. Not both *A* and *B*	Not *B*		Not *A*	
9. If *A* then *B*	*B*			? (Not *A*)

The cell entries give the form of the conclusion deduced from that particular combination of complex and categorical premise. The "?" refers to the *modus tollendo tollens* inference, which may be made directly but, as was argued in the text, is likely to involve a more complex procedure.

A more general statement of the relation between primary and auxiliary inferences is given in Fig. 1. This flow diagram is a simplified model of informal propositional inference. Its power is limited because the program terminates as soon as it succeeds in making a primary inference. However, its general principle is instructive. When a primary inference fails, a test is made to see whether it has failed because there has been a mismatch between the categorical premise and the relevant constituent of the complex

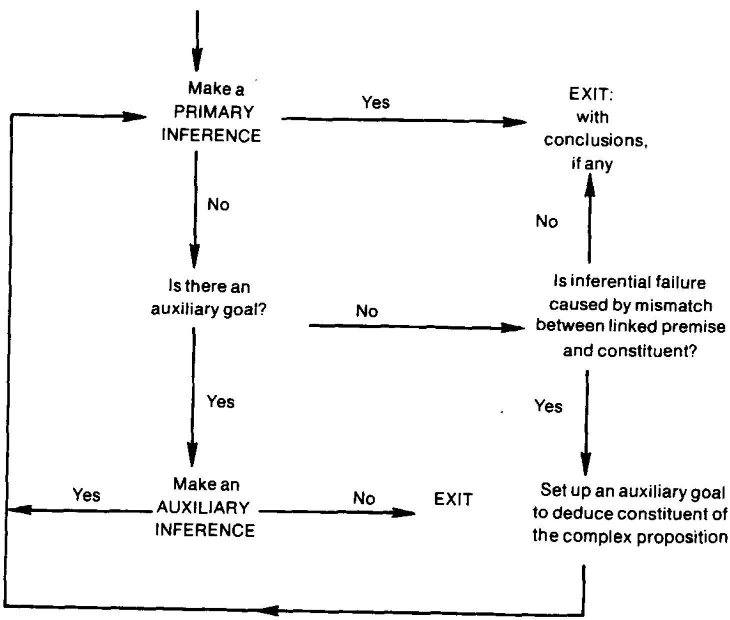


FIG. 1. A simplified model of propositional inference.

premise. If there is such a mismatch, an auxiliary goal is set up to infer the constituent from the categorical premise. If there is no such mismatch, the program can go no further. It is only when an auxiliary goal is set up that it becomes possible to try an auxiliary inference; and an auxiliary goal is only set up in order to try to make a primary inference. An everyday corollary of this relation is that auxiliary deductions do not normally occur in isolation, e.g., it is odd to argue "John's children are grown up," therefore, "John's children are grown up or it is raining in Manchester."

The model described in Fig. 1 is, of course, much less flexible than a human reasoner. Even if it were equipped to deal systematically with several premises, it would be unable to make certain complex deductions in more than one way. For example, faced with premises of the following form:

not D
C or D
A
If A and C then B

the program could deduce the conclusion *B* by proceeding in this fashion:

Make a primary inference from *not D*, and *C or D*, to *C*.
Attempt a primary inference from *C*, and *If A and C, then B*.
Set up an auxiliary goal to infer *A and C*.
Make an auxiliary inference from *A*, and *C*, to *A and C*.
Make a primary inference from *A and C*, and the conditional premise, to *B*.

However, if it attempts to proceed in this fashion,

Attempt a primary inference from *A*, and *If A and C, then B*.
Set up an auxiliary goal to infer *A and C*.

its failure is abrupt because in this case the auxiliary goal cannot be achieved by either a primary or an auxiliary inference. A human reasoner is unlikely to have too much difficulty in completing the chain of inference. What defeats the program is its lack of ability to set itself complex auxiliary goals requiring several inferences for their satisfaction. It cannot, in attempting to deduce *A and C*, set about trying to derive first *A*, and then *C*. Because this procedure is likely to place a considerable load on working memory, it will be interesting to know whether children are capable of it. Indeed, once a suitable modification to the program is made, a danger of the opposite sort is encountered. The program has no limit on the degree of recursion that can occur, whereas a human reasoner can presumably

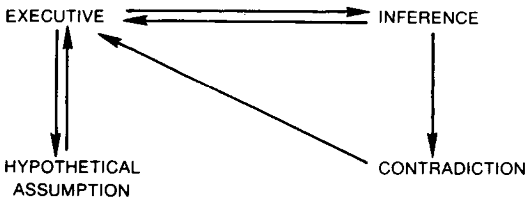


FIG. 2. The interrelations between the four components of the complete model of propositional inference.

tolerate only a certain degree. Such a limit can easily be introduced into the program, but what is humanly tolerable is an empirical matter that has yet to be determined.

A further severe limitation of the model is that it is totally incapable of making certain sorts of inference. It cannot make a deduction by a *reductio ad absurdum;* it cannot construct a hypothetical syllogism; and it cannot resolve a dilemma. These inferences require, first, a procedure that makes hypothetical assumptions and, second, a procedure that is sensitive to contradictions. In order to incorporate these procedures, however, it is necessary to consider the organization of a more complete model of deduction.

The complete model consists of four main components: an executive that controls the various attempts to make inferences, an inferential component that carries out primary and auxiliary inferences, a component for making hypothetical assumptions, and a component for detecting contradictions. The interrelations between these components are summarized in Fig. 2. The executive component, which is shown in Fig. 3, organizes the process of inference. When there are premises but no particular inferential goal, then the executive sets up a goal to make a deduction from the premises and then passes control to the inferential component. If an inference is made, then generally a new goal is created to try to deduce something from its conclusion. However, if no inference can be made, the executive passes control to the procedure that makes hypothetical assumptions. Only when the executive component runs out of premises to try does it cease to make any further attempts at inference.

The inferential component closely resembles the simplified model of inference. It has, as Fig. 4 shows, two main modifications. First, whenever a primary inference is made, control passes to the procedure for detecting contradictions. Second, a routine has been introduced for setting up complex auxiliary goals, thus remedying a major defect of the simplified model.

The procedure for making hypothetical assumptions will be grossly inefficient if it selects them at random, because it may take a long time to discover a fruitful assumption. The obvious heuristic is to find a proposition,

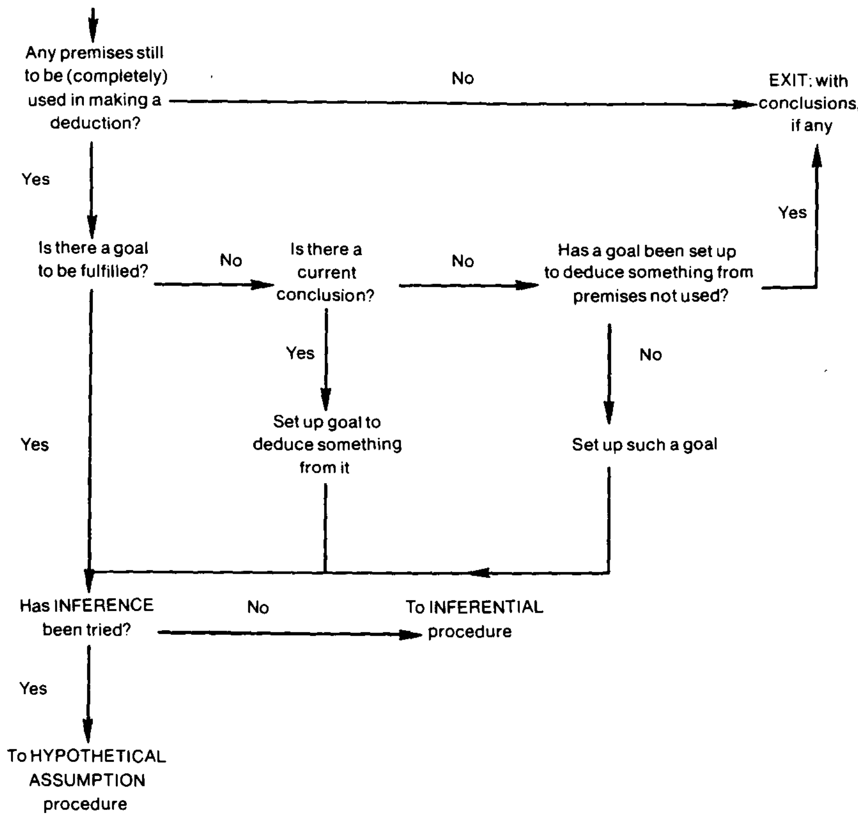


FIG. 3. The executive component of the complete model.

either a premise or a conclusion previously inferred, with an overall form that corresponds to one of the complex propositions of a primary inference, and then to assume an appropriate categorical premise (see Osherson, this volume). The form of this categorical premise can easily be ascertained by examining the primary inference decision table. For example, suppose that there is a premise of the form *If A or C, then B,* then it is clear from Table 3 that a conclusion can be drawn from this premise provided there is a categorical assertion of its antecedent. Therefore, it is this antecedent, *A or C,* that must be assumed. When one returns with this assumption to the inferential program, a primary inference yields the conclusion *B*.

It is important to keep track of an assumption because, unless it can be independently inferred, any conclusion based on it cannot be asserted categorically: all that can be asserted is that the assumption implies the

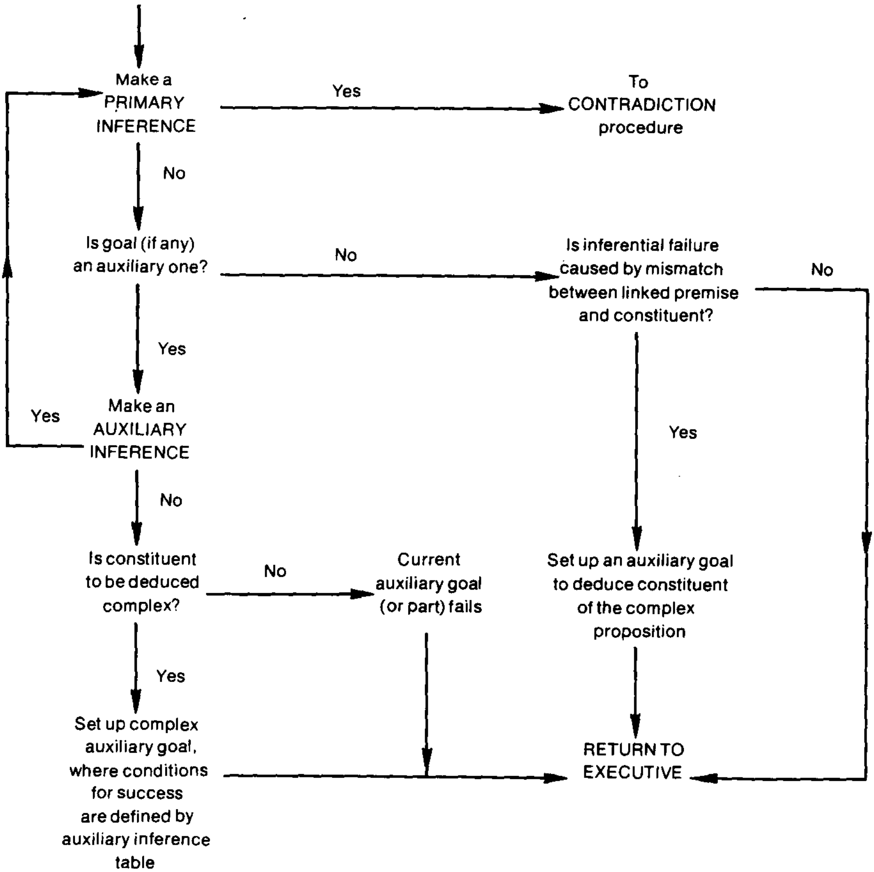


FIG. 4. The inferential component of the complete model.

conclusion. Where an assumption *A* implies a conclusion *B*, the program concludes, by an auxiliary inference, *If A then B*, One obvious moral is that it is no use making an assumption if it leads to no more than this initial deduction. For example, with a premise of the form. *If A then B*, the assumption of *A* can be used to deduce *B*, but if nothing else follows, the auxiliary inference leads merely to *If A then B*, a conclusion that is no more than what was known from the start. The hypothetical procedure has accordingly the capability of setting up a series of goals once an assumption has been made. The first goal is both to deduce a conclusion from the hypothetical assumption and to make a further inference from this conclusion. A failure to achieve this goal is made manifest if the hypothetical procedure is reentered with it, for such an event can only occur if the program has failed to make a primary inference. However, if at least an initial conclusion has been drawn, then a second goal can be set up to deduce a contradiction to it. In other words, the first goal may lead to a hypothetical syllogism, and the second goal to a *reductio ad absurdum*.

If the hypothetical procedure is reentered with the second goal, then it too must be abandoned in favor of a third goal. This situation may arise, for example, in the case of premises with the form of a dilemma:

A or B
If A then C
If B then D

The program may have proceeded as follows:

Assume *A*.
Make the primary inference *A*, *if A then C*, therefore *C*.
Attempt to deduce something from *C* (first hypothetical goal).
Attempt to deduce *not C* (second hypothetical goal).

Evidently, what is needed at this point is another assumption, but not just any assumption will do. A useful heuristic is to find another premise in which the original assumption also occurs and then to make a hypothetical assumption of its remaining major constituents:

Find A or B
Assume B

The third hypothetical goal is to deduce something from this assumption. This goal is fulfilled in the example:

Make the primary inference *B*, *If B then D*, therefore *D*.

Subsequently, the conclusion based on the assumption of *B* will be combined with the conclusion based on the assumption of *A*, and the combination will have the same logic as the premise in which both *A* and *B* occurred. Because this premise was *A or B*, the final conclusion will be *C or D*. The machinery for generating this conclusion, and the other sorts of hypothetical conclusion, is more conveniently located in the procedure for detecting contradictions. The procedure for making hypothetical assumptions is summarized in Fig. 5.

Whenever a primary inference is made, control passes to the procedure for dealing with contradictions that is summarized in Fig. 6. Contradictions that stem from an assumption are taken to imply the negation of that assumption, according to the *reductio* schema. If no assumption has been made, however, it follows that the premises are inconsistent. A conclusion cannot be categorically asserted unless it is established that it is not based

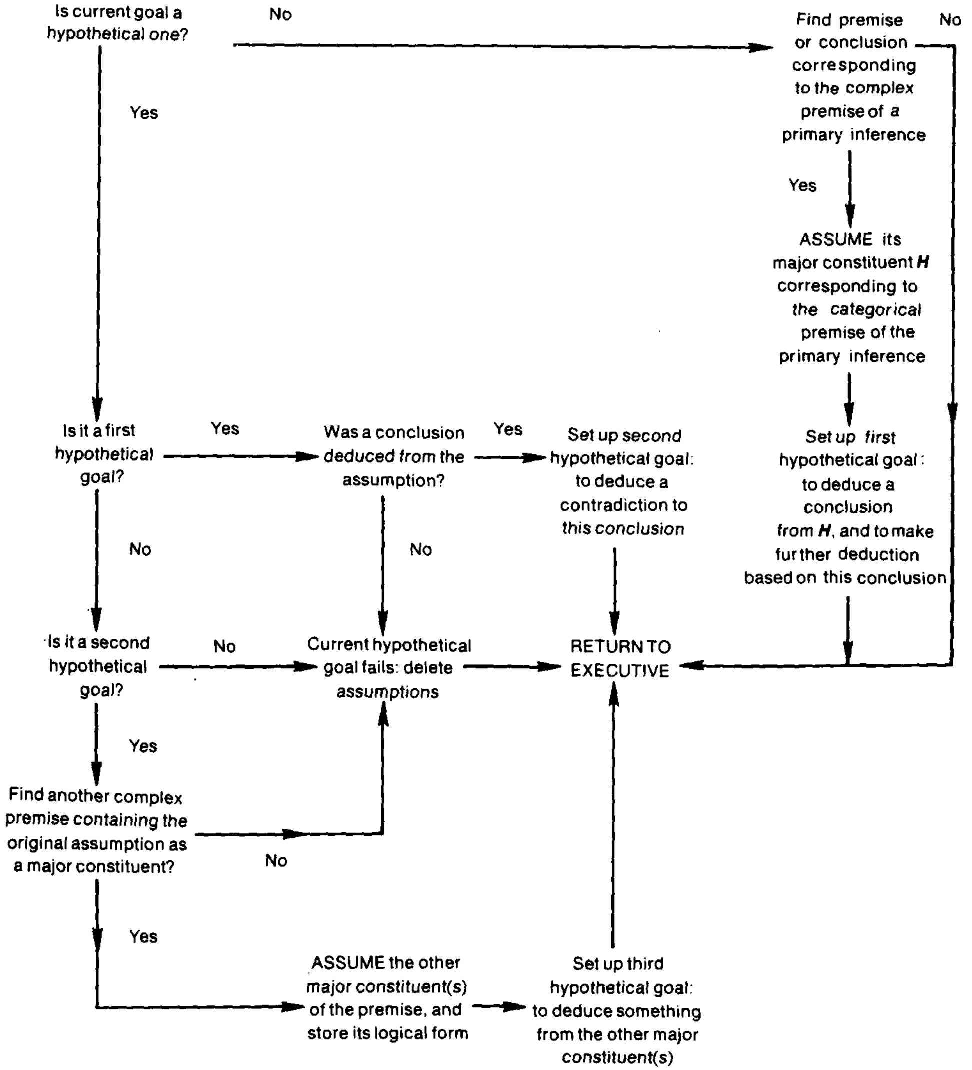


FIG. 5. The procedure for making hypothetical assumptions.

on an assumption. Where an assumption has been made, a conditional conclusion is drawn only if more than an initial consequence has been derived from it. Alternatively, if separate conclusions have been drawn from separate assumptions, then the conclusions may be combined according to the logical relations that originally obtained between the respective assumptions. This routine insures in the case of our previous example that the conclusion to the general dilemma is *C or D*.

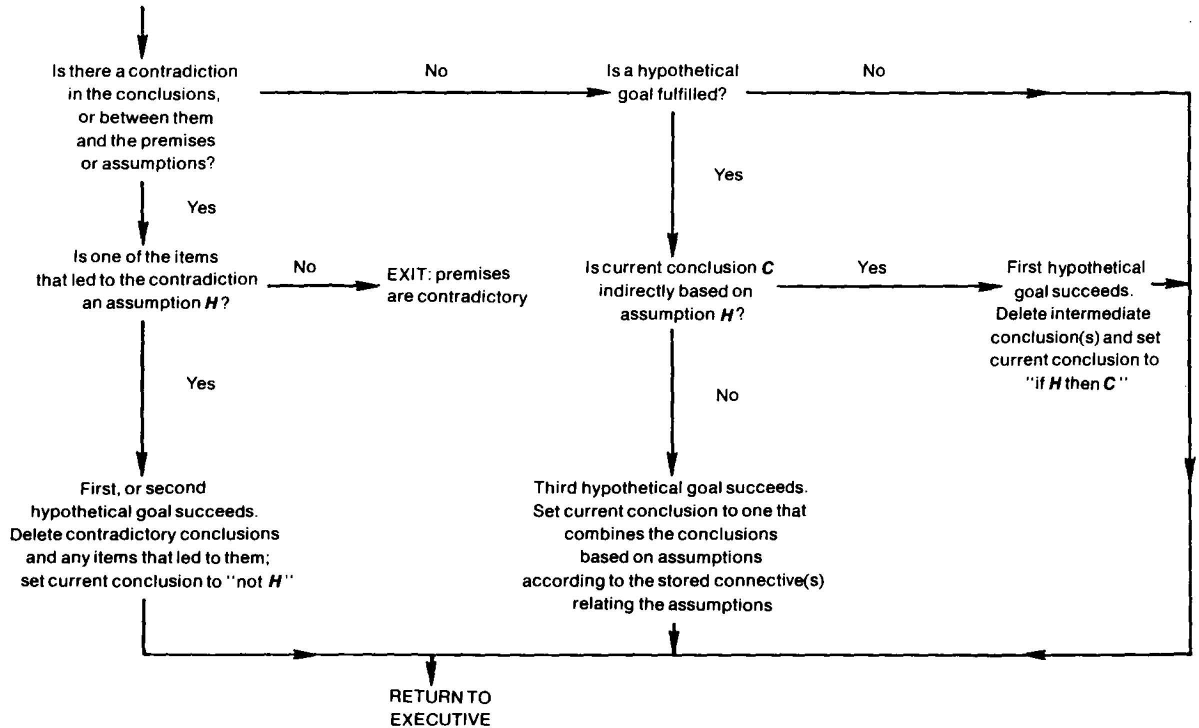


FIG. 6. The procedure for contradictions.

The description of the model is complete, but how is it to be tested empirically? There are three potential methods. First, a correspondence between the model and intelligent but logically naive subjects can be sought. It should be capable of inferences within their competence and it should be incapable of inferences outside their competence. This test may not be particularly stringent because what little evidence there is about human competence in propositional inference has been taken into account in the formulation of the model. Second, a detailed correspondence between the performance of the model and the performance of human reasoners can be sought. Do they go about the business of inference in the same way? Do they, for example, check each time they make a primary deduction to see whether it has yielded a contradiction? The methodology of making such tests of the model is not obvious, although human retrospections are likely to provide some useful evidence. Indeed, they already suggest one weakness of the model. Human reasoners seem to have the ability to gather a rapid global impression of premises and to use it to control their initial attempt at deduction. They do not appear to find it necessary, for example, to attempt a *reductio ad absurdum* before attempting a dilemma. The model's serial ordering of hypothetical goals may therefore be too strong an assumption.

If it proves possible to modify the model so as to increase its power of simulation, then a third method of testing may become feasible. This method involves using a model to make predictions about the relative difficulties of different sorts of inference. Such an approach is, unfortunately, some way off.

REASONING WITH QUANTIFIERS

The valid patterns of relational and propositional deduction do not exhaust the set of general structural rules of inference that human reasoners customarily follow. A new set of principles is required for inference with quantifiers. These terms include the familiar items "all," "some," "none," "many," "few," etc., and a wide range of implicit quantifiers, e.g., "usually," "often," "certain," "possible," and "permissible." Logically speaking, it is possible to develop the usual apparatus of axioms and rules of inference for quantifiers and to raise the customary question of completeness with respect to an appropriate semantical model. Psychologically speaking, however, matters are less clear-cut. Despite the many experimental studies of the syllogism, going back at least 70 years (see the work of Störring, cited in Woodworth, 1938), it is only very recently that actual models of syllogistic inference have been proposed (Erickson, 1973; Revlis, this volume). The majority of theories are about factors that create difficulty in dealing with syllogisms (e.g., Woodworth & Sells, 1935; Chapman & Chapman, 1959; Ceraso & Provitera, 1971). Indeed, the concentration of interest on the syllogism, that traditional but minor province of quantified inference, is symptomatic of the backward state of knowledge in this area.

The model of propositional inference incorporated the mechanism for lexical inference, and they must both in turn be contained within any model of quantified inference. Inferences based on synonymity are a very salient feature because the disposition of quantifiers allows the same basic fact to be expressed in a variety of ways, e.g.,

> Not all the critics admired all of his films.
> Some of the critics did not admire all of his films.
> Some of the critics did not admire some of his films.
> All the critics did not like all of his films.

A similar flexibility extends to terms that are implicit quantifiers, e.g.,

> You are not compelled to vote.
> You are allowed not to vote.
> It is not necessary for you to vote.
> It is possible for you not to vote.

There are at least two alternative ways in which these sorts of semantic relation may be handled. The first alternative is parasitic on a speaker's knowledge of how to give a surface form to an underlying semantic content. I have elsewhere specified a set of grammatical transformations that derive such synonymous sentences from a common underlying form (Johnson-Laird, 1970); and a number of other linguistic accounts of the synonymities involving quantifiers have been proposed (cf. Leech, 1969; Seuren, 1969; Lakoff, 1970; Jackendoff, 1972). Here is not the place to try to weigh up the respective merits of these accounts; what they appear to have in common is the realization that the behavior of quantifiers with negation conforms only in a covert and complicated way to the behavior of these items in a logical calculus. As in logic, one quantifier within the scope of a negation, e.g., "Not all of his films were admired," is equivalent to the alternative quantifier outside the scope of the negation, e.g., "Some of his films were not admired." The complexities arise because of the lack of clear devices in natural language for marking the scope of operators. Sometimes, for example, the scope of negation is indicated by the choice of quantifier, as in the contrast between the following sentences:

> I did not like any of his films.
> I did not like some of his films.

And sometimes scope is indicated by word order—although rarely definitively, because in a sentence such as "None of the critics like some of his films" the first quantifier is within the scope of the second.

The second approach to the problem of synonymity is to provide a direct semantic representation or model for each sentence. The equivalence between sentences is then established by noting that they give rise to the same representation rather than by first reducing the sentences to a common underlying linguistic structure, and then providing a semantic interpretation for it. The direct approach is relatively unexplored for natural language, although Julian Davies at Edinburgh (personal communication) has written a computer program with this sort of facility for quantifiers. The contrast between the two alternative approaches resembles, in many ways, the contrast between an intensional and an extensional semantics. For example, the linguistic approach is well suited to accounting for relations between sentences, whereas the direct approach is well suited to accounting for the relations between sentences and what they describe in the real world. It is too soon either to determine which approach makes the better psychological sense or to grasp the extent to which they are empirically distinguishable.

A more important question is whether it is possible to devise a general model of inference with quantifiers along the lines of the model for propositional inference. In fact, can that model be extended to take into account the internal structure of clauses and their quantifiers? The remainder of this paper is devoted to this problem, but because there is virtually no empirical data apart from the results of experiments on syllogisms, the main aim is to develop a model of how people cope with such syllogisms as the following typical example (from Lewis Carroll):

All prudent men shun hyenas
All bankers are prudent men
∴ All bankers shun hyenas

Psychological studies of the syllogism have been dogged by the baleful tradition of scholastic logic. Not that this logic is necessarily bad—it has simply been bad for psychologists, blinding them to some rather obvious points. Most introductory texts give a standard account of the syllogism, describing its four figures and its 64 moods, and concluding that there are 256 syllogisms. A psychologist, however, should recognize that there are exactly twice this number, because there is nothing God-given about the assumption, underlying the four traditional figures, that the predicate of the conclusion occurs in the first premise. A psychologist might be interested, for example, in the evaluation of a syllogism of the form

All bankers are prudent men
All prudent men shun hyenas
∴ All bankers shun hyenas

This syllogism, of course, is in a figure that does not correspond to any of the traditional four.

The early experimenters (e.g., Wilkins, 1928; Woodworth & Sells, 1935; Sells, 1936) relied on tasks that required given syllogisms to be evaluated rather than on tasks that insured a syllogistic inference was made. They selected the syllogisms from what they thought was a population of 256 in an arbitrary way. (More recent studies have sometimes added the further vice of presenting pooled data from syllogisms of the same mood but different figures, a habit that unfortunately has made it difficult to use the results in constructing a model of inference.) Nevertheless, the pioneering studies led to the important idea of an "atmosphere" effect in which negatives and the quantifier "some" exert a potent bias on the form of an acceptable conclusion to a syllogism—or so, at least, the protagonists of the theory believed.

In order to test the theory and, more importantly, to develop a theory of syllogistic reasoning a systematic study of syllogisms is required. There are 512 possible syllogisms but there are only 64 different combinations of premises (of which 27 yield valid inferences). If experimental subjects are asked to state what follows from each different premise combination, there is a strong presumption that they will be forced to make an inference; and, of course, the population of premise combinations is of a manageable size. Two recent studies, one of which is reported here have tested the ability of intelligent subjects to perform this task. The first study investigated only the 27 valid premise combinations. The second study, carried out in collaboration with Huttenlocher, investigated all 64 pairs of premises. In both these studies, the syllogisms were presented with a sensible everyday content but a content that lacked any perceptible bias toward particular forms of conclusion, e.g.,

> Some of the parents are scientists.
> None of the drivers are parents.

The patterns of inference that were made in both experiments were very similar, even though the experiments were carried out with different materials on opposite sides of the Atlantic. This discussion shall therefore concentrate on the second and more comprehensive study.

The most salient feature of the results is that there is a very wide divergence in the relative difficulty of syllogisms. To take two extreme examples, all 20 subjects presented with premises of the form

Some B *are* A
All B *are* C

correctly deduced a conclusion of the form *Some* A *are* C or its equivalent *Some* C *are* A. However, when these same subjects were presented with premises of the form

All B *are* A
No C *are* B

none of them gave the correct response, *Some* A *are not* C. Perhaps part of the fascination of syllogisms to psychologists is that the manipulation of a handful of variables can yield such very large differences in performance.

The results also demonstrated the inadequacy of the atmosphere hypothesis as a complete account of what goes on in syllogistic reasoning. As there is no point in belaboring this point, amply confirmed in another recent study (Mazzocco, Legrenzi, & Roncato, 1974), it can simply be stated that 40% of the conclusions drawn by the subjects are in accordance with the hypothesis, 8% of their conclusions are incompatible with the hypothesis, and the remaining 52% of their responses are neither compatible nor incompatible with the hypothesis because they consisted almost entirely of the response that no conclusion could be drawn from the premises. It may be objected that the atmosphere hypothesis, in fact, accounts for most of the results if one ignores those syllogisms for which no conclusion was drawn from the premises. However, this objection merely begs the question: how is it that subjects realize that no conclusion follows? The atmosphere hypothesis cannot explain this phenomenon.

There was one striking and unexpected aspect of the results. Certain figures of the syllogism exerted a strong influence on the form of the conclusion that subjects inferred, and this influence did not depend on the logic of particular syllogisms. Where the premises were of the form below where A–B designates the order in which terms A and B were mentioned, regardless of the quantifiers used),

A–B
B–C

85% of the conclusions that were drawn had the form A–C. Where the premises were of the form

B–A
C–B

86% of the conclusions that were drawn had the form C–A. In the case of the other two sorts of syllogisms, however, there were only slight biases, as the following percentages show:

$$\frac{\begin{array}{l}\text{B–A}\\ \text{B–C}\end{array}}{\text{A–C (54\%)}} \qquad \frac{\begin{array}{l}\text{A–B}\\ \text{C–B}\end{array}}{\text{C–A (67\%)}}$$

Although the "figural" effect provides an important clue to how people make syllogistic inferences, it is not the whole story. There is an interaction between the figure and the mood of the premises; and the main goal of a model of syllogistic inference must be to account for this interaction.

Why has this figural effect never been noticed before? The answer is simply because of the neglect of half the possible syllogisms, a neglect fostered by relying on a traditional account of the logic of syllogisms. Indeed, it is a pity that psychologists have not gone back to Aristotle, because the first of his figures has the form

All A *are* B
All B *are* C
∴ *All* A *are* C

This form of syllogism is the only one that Aristotle considered to be perfect, perhaps because the transitivity of the connection between its terms is obvious at a glance (see Kneale & Kneale, 1962, p. 73).

Aristotle's method of validating syllogisms involved their "reduction," by way of a variety of transformations, to the pattern of his perfect syllogism. Subsequent recipes for syllogistic inference have tended to be more mysterious. They are mechanical procedures that work, but their workings are in no way intuitive. It is natural to wonder how such procedures have been established as infallible. One possibility is that they have been tested by exhaustive searches for counterexamples. Another possibility, however, is simply that people, even logically naive people, are capable of syllogistic inference and, with sufficient care, can elucidate a syllogism of any form. This possibility obviously demands that a model of syllogistic inference be able to account for both valid and invalid deductions.

The essence of the model to be developed here is that an initial representation of the premises is set up, from which a conclusion may be read off. This initial representation, however, may be subjected to a series of tests. Where it is submitted to all of these tests, any ultimate conclusion corresponds in all cases to a valid inference. Where some of the tests are omitted, the conclusion may or may not be valid. The syllogisms that are easy to solve turn out either not to permit tests of the initial representation or else not to require their initial conclusions to be modified. The syllogisms that are difficult to solve do permit tests of the initial representation and invariably these tests call for a modified conclusion.

It is impossible from mere introspection to determine how the different sorts of syllogistic premises are mentally represented. They may be represented in a format resembling Euler's diagrams (see Erickson, 1973; Revlis, this volume; Neimark & Chapman, this volume). However, one difficulty with this representation is that it cannot account for the "figural" effect because, for example, the representation of *Some* A *are* B is identical to the representation of *Some* B *are* A. It is unfortunate that, in developing his interesting set-theoretic model of syllogistic inference, Erickson (1973) has overlooked the possibility of a "figural" effect by neglecting half the possible syllogisms. The fact that human reasoners often show a pronounced "figural" bias in stating their conclusions, even where such a bias is logically unwarranted, demonstrates the need to modify any simple representation of premises in the form of Euler's circles. However, instead of attempting such a modification, an entirely different format has been chosen for the present model. The model assumes that human reasoners represent a class by imagining an arbitrary number of its members. For example, a class of artists is represented by a set of elements that are tagged in some way as artists. The nature of the elements and their tags is immaterial—they may be vivid visual images or ghostlike verbal tags. The crucial point is simply that they are discrete elements. A statement such as "All the artists are beekeepers" relates two separate classes and it is represented in the following way:

$$\begin{array}{ccc} \text{artist} & \text{artist} & \\ \downarrow & \downarrow & \\ \text{beekeeper} & \text{beekeeper} & (\text{beekeeper}) \end{array}$$

where representatives of one class are mapped onto representatives of the other class, and the parenthetical item indicates that there may be beekeepers who are not artists. This representation is similar, but not isomorphic, to an Euler diagram. The discrepancy arises from the function of the arrows, which may be interpreted as pointers within a list-processing language. In other words, although the mapping represented by a single arrow is logically symmetrical, i.e., $a \rightarrow b$ is equivalent to $a \leftarrow b$, the two expressions are not psychologically equivalent. Intuitively, the item at the tail of the arrow can be thought of as having stored with it the address in memory of the item at the head of the arrow. Therefore, a fundamental assumption of the model is that it is easier to read off information from such representations proceeding in the direction of the arrows. It is possible to proceed in the opposite direction, but it is harder because it will be necessary to search memory for the item at the tail of the arrow. The model is accordingly very far from making the assumption that subjects tend readily to convert statements (*pace,* Chapman & Chapman, 1959). It does not even assume that they make valid conversions spontaneously during syllogistic reasoning, e.g., from "Some artists are bricklayers" to "Some bricklayers are artists." There is good reason to suppose that such pairs of statements are not always equivalent in ordinary discourse. The former statement, as Hintikka (1973, p. 69) has emphasized, presupposes that the field of search includes all artists, whereas the latter statement presupposes that the field of search includes all bricklayers. This divergence may even lead to a rather special interpretation of the predicate term, suggesting in the second example above, for instance, that some bricklayers are artists in their manner of laying bricks.

In general, a universal affirmative statement, *All* A *are* B, is represented in the following way:

$$\begin{array}{ccc} a & a & \\ \downarrow & \downarrow & \\ b & b & (b) \end{array}$$

where the parenthetical item (b) indicates that there may be a b that is not a. The number of a's and b's in the representation is, of course, entirely arbitrary—we may just as well have linked 15 a's to 15 b's and included 30 parenthetical b's; for convenience, we have chosen two a's in representing each of the different sorts of premise. A particular affirmative statement, *Some* A *are* B, is represented in the following way:

$$\begin{array}{cc} a & (a) \\ \downarrow & \\ b & (b) \end{array}$$

where (a) indicates that there may be an a that is not b, and (b) indicates that there may be a b that is not a.

The representation of a negative statement involves a negative link: there is no mapping of the sort defined above and, moreover, none can be established by any subsequent manipulations of the representation. The representation of a universal negative statement *No* A *are* B requires an arbitrary number of negative mappings, which are here indicated by stopped arrows:

$$\begin{array}{cc} a & a \\ \underline{\downarrow} & \underline{\downarrow} \\ b & b \end{array}$$

If there is a negative link beetween a and b, neither of them may be involved in any positive links from a to b, or from b to a. A particular negative statement, *Some* A *are not* B, is represented by

$$\begin{array}{cc} a & (a) \\ \underline{\downarrow} & \downarrow \\ b & b \end{array}$$

where the positive mapping, $(a) \rightarrow b$ indicates that some a may be b.

The logic of these representations is as follows, where $a \in \mathsf{A}$, $b \in \mathsf{B}$, and R stands for the relation of identity

All A *are* B:	$(a)(\exists b) \quad (a\mathbf{R}b)$
Some A *are* B:	$(\exists a)(\exists b) \quad (a\mathbf{R}b)$
No A *are* B:	$(a)(b) \ \neg(a\mathbf{R}b)$
Some A *are not* B:	$(\exists a)(b) \quad \neg(a\mathbf{R}b)$

On the plausible assumption that A and B are never empty classes in ordinary language, these expressions seem to capture the obvious inferential properties of quantifiers. The logic of sentences in which the copula is replaced by some other relation (e.g., "All the artists married beekeepers") is easily accommodated by this notation.

It is a simple matter to write a program that sets up a representation for the first premise of a syllogism. The representation of the second premise is a more complicated matter because one term — the middle term of the syllogism — will have already been represented. The logical work, in fact, commences with the representation of the second premise because it is grafted onto the representation of the first premise. The process is perhaps best described by way of an example.

Suppose that the first premise of a syllogism is of the form *Some* A *are* B, and can accordingly be represented as

$$\begin{array}{cc} a & (a) \\ \downarrow & \\ b & (b) \end{array}$$

A crucial distinction is whether or not the middle term is the quantified item in the second premise. The representation of the premise *All* B *are* C simply involves mapping the existing members of B onto representative elements of C:

$$\begin{array}{ccc} a & (a) & \\ \downarrow & & \\ b & (b) & \\ \downarrow & \downarrow & \\ c & c & (c) \end{array}$$

The valid conclusion *Some* A *are* C may be read off from this representation, proceeding in the direction of the arrows. However, if the second premise is *All* C *are* B then it is necessary to set up some representative elements of C and to map them onto B. It is also necessary to allow that there may be other *b*'s that are not *c*'s; therefore, an initial representation of this syllogism is

$$\begin{array}{ccc} a & (a) & \\ \downarrow & & \\ b & (b) & (b) \\ \uparrow & \uparrow & \\ c & c & \end{array}$$

Because the mappings do not proceed in a uniform direction, there is no firm anchor on which to base the inference; and the model predicts that subjects will tend to be divided between concluding (invalidly) *Some* A *are* C and concluding (invalidly) *Some* C *are* A.

Both the representations that have been described reflect an initial bias of the model toward establishing transitive mappings. This feature has been introduced in order to account for the subjects' bias toward drawing conclusions where, in fact, none are warranted. Because many subjects are capable of a more sophisticated syllogistic performance, the model assumes that once an initial representation of the premises has been created, it may be submitted to tests before any attempt is made to read off a conclusion. These tests can be characterized as efforts to test to destruction any initial transitive mappings. The initial phase is analogous to a process of verification; the testing phase is analogous to a process of falsification and, as with falsification, it is often overlooked by subjects (see Johnson-Laird & Wason, 1970).

The procedure for falsifying a mapping involves trying to modify the representation of the second premise so that it is no longer connected to items that are themselves involved in a mapping relation. The procedure has no effect on the first illustrative syllogism but it is possible to modify the initial representation of the second illustrative syllogism from

$$\begin{array}{ccc} a & (a) & \\ \downarrow & & \\ b & (b) & (b) \\ \uparrow & \uparrow & \\ c & c & \end{array}$$

to

$$\begin{array}{ccc} a & (a) & \\ \downarrow & & \\ b & (b) & (b) \\ & \uparrow & \uparrow \\ & c & c \end{array}$$

The critical link has been broken; and because both its presence and its absence are consistent with the interpretation of the premises, it follows that no valid conclusion can be deduced from them.

The predictions of the model for the two illustrative syllogisms are summarized below, together with the numbers of experimental subjects (out of 20) deducing the predicted conclusions:

Some A *are* B		*Some* A *are* B	
All B *are* C		*All* C *are* B	
∴ *Some* A *are* C:	16	∴ *Some* A *are* C:	5
		∴ *Some* C *are* A:	5
		No conclusion follows:	9

A simple set-theoretic model does not account for the results with the first of these syllogisms because it predicts a response of *Some* A *are* C as often as a response of *Some* C *are* A.

Certain features of the present model can only be illustrated by considering the representation and testing of negative premises. Consider premises of the form

No B *are* A
All C *are* B

The first premise is represented as

$$\begin{array}{cc} a & a \\ \overline{\uparrow} & \overline{\uparrow} \\ b & b \end{array}$$

The second premise, of course, requires an additional *b* to be introduced and, when such an introduction occurs, the model bears in mind the universal nature of the first premise:

$$\begin{array}{ccc} a & a & a \\ \overline{\uparrow} & \overline{\uparrow} & \overline{\uparrow} \\ b & b & (b) \\ \uparrow & \uparrow & \\ c & c & \end{array}$$

It is a straightforward matter to read off the conclusion, *No* C *are* A, but a more difficult matter to read off the conclusion *No* A *are* C.

The initial representation of the premises

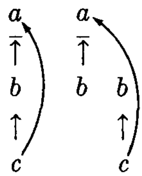
No B *are* A
All B *are* C

is set up in a similar way:

$$\begin{array}{ccc} a & a & \\ \overline{\uparrow} & \overline{\uparrow} & \\ b & b & \\ \downarrow & \downarrow & \\ c & c & (c) \end{array}$$

However, the mappings do not proceed in a uniform direction and the model predicts that subjects will be divided between the (invalid) conclusions *No* A *are* C and *No* C *are* A.

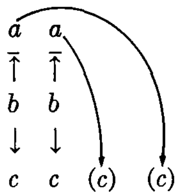
The falsification tests of a negative mapping consist in trying to establish that a transitive link can be set up between the elements in the representation. The only constraint on this maneuvre is that elements cannot be linked in inconsistent ways such as these in the following example:



because these links imply inconsistencies, such as that c both is and is not an a. The first syllogism survives the falsification tests unmodified. The second syllogism does not. Its initial representation

$$\begin{array}{ccc} a & a & \\ \overline{\uparrow} & \overline{\uparrow} & \\ b & b & \\ \downarrow & \downarrow & \\ c & c & (c) \end{array}$$

allows the following links to be established:



It is always possible to add new parenthetical items provided there are existing ones; it is never possible to add any items to a set that contains no parenthetical items. The distinction, in fact, corresponds to the traditional notion of a distributed term (no parenthetical items) and an undistributed term (parenthetical items). The new links in the representation are, of course, consistent. Yet subjects initially predisposed to conclude *No* A *are* C may find this second representation, which suggests *All* A *are* C, such a contrast that they may judge that no conclusion follows from the premises. More astute subjects, however, can appreciate that it is impossible to add further *a*'s to the representation and therefore that *Some* C *are not* A.

The predictions of the model for the two negative syllogisms are summarized below, together with the number of experimental subjects (out of 20) deducing the predicted conclusions:

No B *are* A		*No* B *are* A	
All C *are* B		*All* B *are* C	
∴ *No* C *are* A:	13	∴ *No* C *are* A:	4
∴ *No* A *are* C:	3	∴ *No* A *are* C:	3
		No conclusion follows:	4
		∴ *Some* C *are not* A:	7

The main features of the model of syllogistic inference have now been illustrated, and it should be obvious that it has the sort of flexibility needed to match the diversity of subjects' deductions. To generate quantitative predictions, however, would require the specification of various parameters and the estimation of their values from the experimental results. The exercise would not be very rewarding. A more appealing possibility is to try to derive assumptions about the relative value of these parameters from the processing properties of the model itself. The point can be illustrated by considering the figural effect in terms of the processing of lists. When the arrows in a representation lie in a uniform direction, they largely determine the direction in which conclusions are read off from it. It is possible to proceed in the opposite direction but at the cost of having to search memory for the items that are the tails of arrows. This asymmetry, which is reflected in human performance, is a simple consequence of the structure of lists. At a later stage of development of the model, it may be possible to derive some of fine-grain aspects of performance from similar sorts of information-processing considerations.

The reader familiar with the problems of manipulating Euler's circles can appreciate the computational advantage of the present style of representation. An Eulerian representation of, for example, *Some* A *are* B requires four separate diagrams to be created. When such representations are combined, the combinatorial consequences can become psychologically embarassing, particularly where several premises are involved, e.g.,

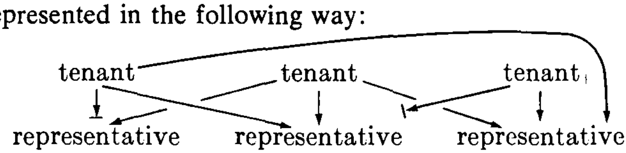
Some A *are* B
All B *are* C
All C *are* D

An Eulerian deduction from these premises involves considering at least $4 \times 2 \times 2 = 16$ different combinations of diagram, whereas the list representation is simply

$$\begin{array}{cccc} a & (a) & & \\ \downarrow & & & \\ b & (b) & & \\ \downarrow & \downarrow & & \\ c & c & (c) & \\ \downarrow & \downarrow & \downarrow & \\ d & d & d & (d) \end{array}$$

from which it is easy to infer *Some* A *are* D. It may be that in making such inferences human reasoners break the problem down into a series of syllogisms; but it is also feasible that they set up a complete representation of several premises in this way.

There are, of course, alternative models of syllogistic inference that can be couched in the list format. Mark Steedman (personal communication) has devised a model, involving a more elegant representation, in which invalid inferences arise not from a failure to test initial and improper representations of the combined premises but from actual errors in the representations of single premises. In particular, the model assumes that parenthetical items are often neglected so that universal premises come to be interpreted as *All and only* A *are* B. The list representations may also be extended in order to deal with quasi-numerical quantifiers (e.g., "many," "most," "few") and with statements involving several quantifiers. A statement such as "Some of the tenants will not vote for all the representatives" may be represented in the following way:



where the mapping represents the relation *votes for*. The combination of these diagrams is a complicated affair, but from the few studies of inference with such statements (Johnson-Laird, 1969b) it seems that untrained human reasoners have fairly restricted powers with them, too. It is likely that the list representation is sufficiently flexible to form part of a general model of quantified inference.

CONCLUSIONS

In an attempt to formulate models of deduction that operate on different aspects of the structure of statements, two essential inferential processes have been discerned: the transformation of information and the combination of information from separate sources. The transformation of information occurs primarily in lexical inferences and in the majority of auxiliary inferences. The combination of information occurs in three-term series problems, in inferences about complex spatial relations, in primary propositional inferences, and in syllogisms. When information is transformed, the process is usually governed by essentially linguistic rules. When information is combined, however, the process often seems to involve the creation of an internal "model" of the world. This procedure seems to be necessary for syllogistic inference; and it has the great advantage that the transitivity of relational terms can be made a direct consequence of their representation rather than an indirect consequence of an additional rule of inference. Indeed, it is seldom that it can be conclusively demonstrated that the transformation of information does not proceed by the construction of internal models.

What of the role of content? Do different contents introduce perhaps difference principles of inference? It is certainly noticeable that a listener is able to draw on general knowledge to allow a speaker to leave many things unsaid. The sorts of inference that a listener can make are illustrated in the following examples culled from current work in a variety of disciplines:

He went to three drugstores.
Therefore, the first two drugstores didn't have what he wanted. (Abelson & Reich, 1969)

The mirror shattered because the child grabbed the broom.
Therefore, the child hit the mirror with the broom and broke it. (Bransford & McCarrell, 1972)

The policeman held up his hand and the cars stopped.
Therefore, the policeman was directing the traffic. (Collins & Quillian, 1972)

Harry is enjoying his new job at the bank, and he hasn't been to prison yet.
Therefore, Harry may be tempted to steal some of the money in the bank. (Wilson, 1972)

John gave Mary a beating with a stick.
Therefore, John wanted to hurt Mary. (Schank, Goldman, Rieger, & Riesbeck, 1973)

Janet needed some money. She got her piggybank and started to shake it.
Therefore, Janet got her piggybank and shook it in order to get some money from it. (Charniak, 1973)

Of course, the man may have visited three drugstores in order to enforce a protection racket; and the mirror may have shattered because it was balanced on top of the broom; and so on. The inferences are therefore plausible rather than valid. What appears to happen, however, is that people exploit a communal base of knowledge that includes such assumptions as:

Drugstores are shops that have certain sorts of goods.
People visit shops in order to buy goods.
If one shop does not have an item that it normally stocks, then another shop of the same sort may have it.

This knowledge will be automatically elicited by any utterance with a relevant topic, and it can be used by the inferential machinery in order to make good any gaps in the explicit discourse. The procedure relies on a convention that a speaker will draw attention to any special circumstances that render communal assumptions inappropriate.

If discourse is supplemented by common sense assumptions, it is unnecessary to postulate special rules of inference to deal with specific topics, although the content of statements may exert a selective bias on the availability of different rules of inference (see Wason & Johnson-Laird, 1972). The simple answer is to revert to Sherlock Holmes, who was himself a model of deduction. What he exploited, very much in the manner of PLANNER, were two special sources of information: an acute perceptual attention to detail, and an extensive specialized knowledge—the sort of knowledge to be expected in an individual who contributed to the literature on both cigar ash and tattoos. It was knowledge that was the foundation of his exceptional ability. The structure of his inferences was, indeed, elementary.

ACKNOWLEDGMENTS

This work was supported by a grant for scientific assistance from the Social Science Research Council. I am indebted to George Miller for some extremely useful ideas about the structure of the lexicon. I am also grateful to Keith Oatley and Mark Steedman for helpful discussions on spatial and syllogistic inference, to Gerard

Kempen for his work in programming part of the model of propositional inference, and to Janellen Huttenlocher for allowing me to describe some joint research. My interest in inference was originally stimulated by Peter Wason, and I would like to acknowledge here my debt to him.

REFERENCES

Abelson, R. P., & Reich, C. M. Implication modules: a method for extracting meaning from input sentences. *First international joint conference on artificial intelligence*. Washington, D.C., 1969.

Amarel, S. An approach to heuristic problem solving and theorem proving in the propositional calculus. In J. F. Hart & S. Takusu (Eds.) *Computer science and systems*, Toronto: University of Toronto Press, 1967.

Bar-Hillel, Y. Dictionaries and meaning rules. *Foundations of Language*, 1967, **3**, 409–414. Reprinted in Y. Bar-Hillel, *Aspects of language: Essays and lectures on philosophy of language, linguistic philosophy and methodology of linguistics*. Jerusalem: The Magnes Press, The Hebrew University. Amsterdam: North-Holland Publ., 1970.

Bendix, E. H. *Componential analysis of general vocabulary: The semantic structure of a set of verbs in English, Hindi and Japanese*. The Hague: Mouton, 1966.

Beth, E. W., & Piaget, J. *Mathematical epistemology and psychology*. Translated by W. Mays. Dordrecht-Holland: Reidel, 1966.

Bransford, J. D., & McCarrell, N. S. A sketch of a cognitive approach to comprehension: some thoughts about understanding what it means to comprehend. Paper presented at the Conference on Cognition and the Symbolic Processes, Pennsylvania State University, October 1972.

Carnap, R. *Meaning and necessity: A study in semantics and modal logic*. (2nd ed.) Chicago: Chicago University Press, 1956.

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

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

Charniak, E. Jack and Janet in search of a theory of knowledge. *Third international joint conference on artificial intelligence*. Stanford: Stanford Research Institute, 1973.

Church, A. A note on the Entscheidungs Problem. *Journal of Symbolic Logic*, 1936, **1**, 40–41.

Clark, H. H. More about "Adjectives, comparatives, and syllogisms": a reply to Huttenlocher and Higgins. *Psychological Review*, 1971, **78**, 505–514.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning & Verbal Behavior*, 1969, **8**, 240–247.

Collins, A. M., & Quillian, M. R. How to make a language user. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

Dowty, D. R. On the syntax and semantics of the atomic predicate CAUSE. In P. M. Peranteau, J. N. Levi, & G. C. Phares (Eds.), *Papers from the eighth regional meeting, Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 1972. Pp. 62–74.

Doyle, A. C. *The adventures of Sherlock Holmes*. London: J. Murray, 1917.

Erickson, J. R. A set analysis theory of behavior in formal syllogistic reasoning tasks. Paper presented to the Loyola Symposium on Cognition, Chicago, May 1973.

Evans, J. St. B. T. Deductive reasoning and linguistic usage (with special reference to negation). Unpublished doctoral dissertation, University of London, 1972.

Gödel, K. Zur intuitionistischen Arithmetik und Zahlentheorie. *Ergebnisse eines mathematischen Kaloquiums,* 1933, Heft iv, 34–38. Translated by M. Davis, On Intuitionistic Arithmetic and Number theory. In M. Davis (Ed.), *The undecidable: Basic papers on undecidable propositions, unsolvable problems and computable functions,* Hewlett, N.Y.: Raven Press, 1965.

Hewitt, C. PLANNER: a language for manipulating robots and proving theorems in a robot. Project MAC Artificial Intelligence Memo No. 168, Massachusetts Institute of Technology, 1970.

Heyting, A. *Intuitionism, an introduction.* Amsterdam: North-Holland Publ., 1956.

Hintikka, J. *Logic, language-games and information: Kantian themes in the philosophy of logic.* Oxford: Clarendon, 1973.

Huttenlocher, J., & Higgins, E. T. Adjectives, comparatives, and syllogisms. *Psychological Review,* 1971, **78,** 487–504.

Jackendoff, R. *Semantic interpretation in generative grammar.* Cambridge, Mass.: MIT Press 1972.

Johnson-Laird, P. N. "&" *Journal of Linguistics,* 1969, **6,** 111–114. (a)

Johnson-Laird, P. N. Reasoning with ambiguous sentences. *British Journal of Psychology,* 1969, **60,** 17–23. (b)

Johnson-Laird, P. N. The interpretation of quantified sentences. In G. B. Flores D'Arcais & W. J. M. Levelt (Eds.) *Advances in psycholinguistics.* Amsterdam: North-Holland, 1970.

Johnson-Laird, P. N. Experimental Psycholinguistics. In M. R. Rosenzweig & L. W. Porter (Eds.) *Annual Review of Psychology,* Vol. 25. Palo Alto, Calif.: Annual Reviews Inc., 1974.

Johnson-Laird, P. N., & Tridgell, J. When negation is easier than affirmation. *Quarterly Journal of Experimental Psychology,* 1972, **24,** 87–91.

Johnson-Laird, P. N., & Wason, P. C. A theoretical analysis of insight into a reasoning task. *Cognitive Psychology,* 1970, **1,** 134–148.

Karttunen, L. Presuppositions of compound sentences. *Linguistic Inquiry,* 1973, **4,** 169–193.

Keenan, E. L. *A Logical Base for a Transformational Grammar of English.* University of Pennsylvania, Transformations and Discourse Analysis Papers, No. 82, 1970.

Kneale, W., & Kneale, M. *The development of logic.* Oxford: Clarendon, 1962.

Kowalski, R. Predicate logic as programming language. Memo No. 70, Department of Computational Logic, School of Artificial Intelligence, University of Edinburgh, 1973.

Lakoff, G. Linguistics and natural logic. *Synthese,* 1970, **22,** 151–271.

Lakoff, R. If's, and's, and but's about conjunction. In C. J. Fillmore & D. T. Langendoen (Eds.), *Studies in linguistic semantics.* New York: Holt, Rinehart & Winston, 1971.

Landauer, T. K., & Meyer, D. E. Category size and semantic memory retrieval. *Journal of Verbal Learning & Verbal Behavior,* 1972, **11,** 539–549.

Leech, G. N. *Towards a semantic description of English.* London: Longmans, Green, 1969.

Mazzocco, A., Legrenzi, P., & Roncato, S. Syllogistic inference: The failure of the atmosphere effect and the conversion hypothesis. *Italian Journal of Psychology,* 1974, **1,** 157–172.

Miller, G. A., Galanter, E., & Pribram, K. H. *Plans and the structure of behavior.* New York: Holt, Rinehart & Winston, 1960.

Miller, G. A., & Johnson-Laird, P. N. *Perception and language,* 1975, in press.

Montague, R. English as a formal language. In B. Visentini *et al.* (Eds.), *Linguaggi nella società e nella tecnica.* Edizionidi Comunità, Milan, 1970.

Newell, A., & Simon, H. A. *Human problem solving.* Englewood Cliffs, N. J.: Prentice-Hall, 1972.

Parsons, T. A semantics for English. Mimeo., 1972.

Ramsey, F. P., General propositions and causality. In F. P. Ramsey (Ed.) *Foundations of mathematics and other logical essays.* London: Routledge & Kegan Paul, 1950.

Reiter, R. A semantically guided deductive system for automatic theorem-proving. In *Third international joint conference on artificial intelligence.* Stanford: Stanford Research Institute, 1973.

Rips, L. J., Shoben, E. J., & Smith, E. E. Semantic distance and the verification of semantic relations. *Journal of Verbal Learning & Verbal Behavior,* 1973, **12,** 1–20.

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. Automatic deduction with hyper-resolution. *International Journal of Computer Mathematics,* 1966, **1,** 227–234.

Schaeffer, B., & Wallace, R. The comparison of word meanings. *Journal of Experimental Psychology,* 1970, **86,** 144–152.

Schank, R., Goldman, N., Rieger, C. J., & Riesbeck, C. Margie: memory, analysis, response generation, and inference on English. *Third international joint conference on artificial intelligence.* Stanford: Stanford Research Institute, 1973.

Sells, S. B. The atmosphere effect: an experimental study of reasoning. *Archives of Psychology,* 1936, **29,** 3–72.

Seuren, P. A. M. *Operators and nucleus: A contribution to the theory of grammar.* Cambridge, England: Cambridge University Press, 1969.

Stalnaker, R. C. A theory of conditionals. *American Philosophical Quarterly, Monograph No. 2.* In Nicholas Rescher (Ed.) *Studies in Logical Theory,* 1968, 98–112.

Suppes, P. On the behavioral foundations of mathematical concepts. In L. N. Morrisett & J. Vinsonhaler (Eds.), *Mathematical learning. Child Development Monograph,* 1965, **30**(1), 60–96.

Tarski, A. The concept of truth in formalized languages. In *Logic, semantics, metamathetics: Papers from 1923 to 1938.* Translated by J. H. Woodger. London and New York: Oxford University Press, 1956.

van Fraassen, B. C. *Formal semantics and logic.* New York: Macmillan, 1971.

Wason, P. C., & Johnson-Laird, P. N. *Psychology of reasoning: structure and content.* Cambridge, Mass.: Harvard University Press, 1972.

Wilkins, M. C. The effect of changed material on the ability to do formal syllogistic reasoning. *Archives of Psychology,* 1928, **16**(102).

Wilson, D. Presuppositions II. University College London, Mimeo, 1972.

Winograd, T. *Understanding natural language.* New York: Academic Press, 1972. Also published as the whole of *Cognitive Psychology,* Part 1, 1972, **3,** 1–191.

Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

Woodworth, R. S., & Sells, S. B. An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology,* 1935, **18,** 451–460.