

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 devised 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 of representing putative theories and then grinding away at them is to represent them as programs. When the process of trying to discover a theorem is represented as a program, the process of trying to discover a theorem is exploited so successfully in natural language. One obvious piece of information and deductive proof of a problem, to be taken into account by the system, is the goal. The system therefore gains general information from experiments to be believed in the deductive process. There is a close relationship between logical and AI circles to emphasize the tendency is also evident especially in the development of AI (Kowalski, 1973). The aim is to address the balance and to express principles of thought that are main. In examining this topic of lexical reasoning, proposition A few new experimental results are developing models of deduction.

LEXICAL REASONING

Perhaps the most obvious difficulty is hardly noticed in ordinary language. Such lexical items as nouns, verbs, and prepositions are, of course, often interrelated. The relations acts very much as a chinery revolve. If, for example, a tax, then from the statement "He must pay the tax" the inference is invalid: it lacks a life, however, human beings know that poodles are dogs, and they know the canons of formal logic.

Logicians have tended to ignore this difficulty. The device of meaning relations (1967) was developed to deal with the relations between words. Psychology

an suffice for all the ways in which . Yet, in spite of this reservation, any general ways of thinking that eductions?

al work provides a decidedly nega-
g problem seems to matter just as
g not only how a problem is repre-
ces that are made. Wason and
dence of such effects in a variety
otheses to reasoning with proposi-
ncreasingly popular conception of
AI).

o program computers to carry out
natic methods of theorem proving.
that would both translate natural
ate calculus and operate on these
ig procedures. Because it had long
to algorithm for proof within the
is of a heuristic nature. Very often,
spirit turned out to be more power-
a theorem could be proved, to find
las, no guarantee that the method
t was impossible to derive a given
: heart of Church's (1936) proof
r procedure for the predicate cal-
cedures have one glaring disadvan-
way at a problem, there is no way
timately come up with a solution.
ater discover it; but if there is no
re, the impetus behind such sophis-
ple and the hyperresolution princi-
ncrease the efficiency of programs
they existed, within a reasonable
here is another difficulty with gen-
go to work on a problem, it has
ilus; and it turns out that the busi-
expressions into their appropriate
dinary language does not wear its
re often surprising divergences be-
sion and its underlying logic. Once
ure for carrying out correct transla-

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 psycho-logical 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 re- dress the balance and to examine to what extent there may be general principles of thought that are independent of any particular problem do- main. 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 de- veloping 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 inter- relations acts very much as a smoothing oil to help the inferential ma- chinery 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, al- though 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*, then *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 experimentation of the transitivity of a relation.

There are other patterns of lexical relations, for instance, an intransitive relation (*R*), for instance

aRb and bRb

The relation "next in line to" is *c* in line to *b*, and *b* is next in line in line to *c*. A nontransitive relation nor the intransitive inference; for to *c*, then nothing follows about *a* be arranged in a circular fashion.

Another aspect of the logic of inferences is symmetrical if it permits an inference

aRb

The relation *next to* is symmetric; that *b* is next to *a*. A relation is symmetric if of the form

aRb

The relation *on the right of* is clearly nonsymmetrical if it permits neither of the above forms.

There are still other logical properties of transitivity and connectivity, but their complexity is negligible. However, because transitive relations have attributes, the lexicon already contains relations exemplified in Table 1 by a set of relations.

The semantic representation of relations about their transitivity and symmetry "beyond" must permit a transitive inference of "nearest to" must prevent it. Whether this representation is effected by how this representation is effected has stored with it in the mental lexicon value and another tag indicating, tagged as transitive, it permits an

aRb and

This conception evidently requires to be defined as adjuncts to the lexicon. A relation is symmetric if the transitivity of a relation self

er the guise of studying "semantic studies are perhaps worth delineat- ohnson-Laird, 1974). The over- ned nouns and, in particular, the They have shown that where there *poodle: dog: animal*, it may take tion. It may take time, in other s an animal. A variety of compet- in this phenomenon (e.g., Collins 972; Schaeffer & Wallace, 1970). ctory, if only because there are is easier to retrieve than its constit- harder to verify than "a poodle e included in the class of animals eless, it remains true that not all he lexicon with the same ease. It ndredths of a second, to retrieve k, of course, has the logical form Gibbs and I gave subjects an infer-

spray.
chemical spray.

e conventional studies of semantic *snake: reptile*) a transitive infer- ng its constituents; in other cases nference took less time than the

there between the meanings of ynonymy (e.g., *automobile-car*), s inclusion (e.g., *dog-animal*); and ng relations between sentences in t that studies of semantic memory s a potent relation because it leads e hierarchies can be generated by times by the relation *part of*. How- elations is comparative adjectives, e expressions of the general form o infer that if *a* is larger than *b*, han *c*. However, so much contro- the process (see Huttenlocher & a broader issues have been ignored

in the quest to explain experimental findings. One such issue, the represen- tation 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

$$aRb \text{ and } bRc \quad \therefore \text{not } (aRc)$$

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

$$aRb \quad \therefore bRa$$

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

$$aRb \quad \therefore \text{not } (bRa)$$

The relation *on the right of* is clearly asymmetrical. A relation is nonsym- metrical 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 reflexi- tivity 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

$$aRb \text{ and } bRc \quad \therefore aRc$$

This conception evidently requires inference schemata to be separately speci- fied 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 aRc would be self-evident from the joint representation of aRb and bRc . 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 c to be far too complicated to be h representation of the scene

The black ball is directly be;
the right of the cue ball,

Therefore, if I move so that t
ball, then the cue ball is o

Even if it is possible to frame infer-
ence to be made without the cor
it is most unlikely that this appr
inference. The only rules of infer
setting up a joint representation
for interrogating the joint represe
by the semantic information in the
ing lexical information to specify
equally well to abstract terms with
to visualize directly. With concre
joint representation is isomorphic
below in the analysis of quantific

Perhaps the most potent source
a language. The same sorts of relat
lexical items—relations such as a
clusion (e.g., *assassinate-murder*-
to express relations between sever
analysis of a relation and its conve
The additional complexity of ver
problems. Consider the followin
on the meanings of verbs:

Pat forced Dick to refrain fr
∴ Dick refrained from swea
∴ Dick did not swear.

Sam managed to prevent De
∴ Sam prevented Dean fron
∴ Dean did not pretend to l
∴ Dean was not naive.

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

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.

1
Exemplars of the
ary Relations in
guage

Intransitive	Symmetric
+	+
+	-
+	o
-	+
-	-
-	o
o	+
o	-
o	o

e = intransitive; o transitive

metric = asymmetric; o sym-

evident from the joint represent-
ing quantified statements (e.g.,
this property is described below.
tation for simple relational terms
relations. Consider the following

chair.
the chair.

ference is made involves setting
e depicted by the premises. This
fleeting abstract delineation—its
point is that its formal properties
o that the conclusion can be read
an actual array of objects. It may
on of the premises is unnecessary,
n appeal to general principles,
t items related by *between* must
that relational terms are tagged
they permit—founders on more

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 exp. gone, *or else* it has been sunk *and* what these connectives mean is ta inferences on the basis of the fo example, a speaker can hardly b of "or" unless he appreciates th

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

Therefore, it has gone.

The logic of connectives has be of the propositional calculus. The and a variety of different ways o informal exposition of the standa letters are allowed to range ove formalized by specifying what c 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. $(p \rightarrow q) \rightarrow [(r \text{ or } p) \rightarrow$

where the arrow is a sign for axioms, two rules of inference allows new formulas to be gene formula for a propositional variab of inference, the so-called law of

From a formula *A* together
B may be deduced.

It is fairly simple to show that all the formulas that are true connectives.

What does such a system state cally naive persons? The answe is worth dwelling on the system l tial psychologist, Piaget, has use (Beth & Piaget, 1966) and bec deduction is instructive.

Among the more obvious diff as a model of ordinary deductio

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:

$$\frac{A \quad \text{If } A \text{ then } B}{\therefore B}$$

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 tion and disjunction may be taken for them, and the remaining connectives and disjunction. From a psychological point of view, it may be foolish to seek parsimony at the expense of a set of psychologically basic principles.

Any decision about whether a principle is basic is ultimately an empirical matter. The inference is in the immediate grasp of most persons. An inference schema can be used if people are incapable of carrying out the inference in a few minutes, subsequently giving a decision on the conditions they have carried out to determine whether it is not enough evidence to determine the validity of inference. What can be done is to find an approximation to it, taking care to use only those known to cause difficulty to logic. The use of such an inference is feasible for the majority of persons, hardly a decisive proof: the inference is not basic. Only those inferences that are basic are therefore included in the calculus. The final decision must depend on further evidence.

Some extremely simple inferences can be derived from a clause expressing the meaning of a proposition. From a clause expressing the same meaning as the previous clause, I shall write simply of propositions. I am dealing with inferences that are not basic. The conclusion *A* can be derived from the form *A* or *A*. These inferences are the following schemata:

Although inferences of this sort are not basic, their structure is not derived from anything more basic. They may not be too trivial to be included in the calculus to play, and a model of propositional

ordinary language simple con-
ed by a connective. A sentence

he calculus:

; an excellent rider.

for dealing with such sentences

le.

proposition. Another difficulty,
onal: the meaning of its connec-
h value they give to a complex
s of its constituents. The multi-
e.g., *because, before, although*)
truth-functional calculus. Nor,
tions be immediately accommo-
work.

calculi and the inferential ma-
espective functions. Calculi are
s. The aim of practical inference,
pass from one contingent state-
erence is likely to involve few,
r of rules of inference. A formu-
re appropriate abandons axioms
Gentzen's method of "natural"
ne influence in the development
l, 1967; Reiter, 1973). A system
ation of rules of inference in a
ns, for example, is stated in the

en B

and the conclusion appears below
ulates 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:

$$\frac{A \text{ and } B}{\therefore A} \quad (3a)$$

$$\frac{A \text{ and } B}{\therefore B} \quad (3b)$$

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

$$\frac{A}{\therefore A \text{ or } B} \quad (4a)$$

$$\frac{A}{\therefore B \text{ or } A} \quad (4b)$$

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

$$\frac{A \quad B}{\therefore A \text{ and } B} \quad (5a)$$

$$\frac{A \quad B}{\therefore B \text{ and } A} \quad (5b)$$

And the final pair permit negated conjunctions to be deduced:

$$\frac{A \text{ and not } -B}{\therefore \text{not both } A \text{ and } B} \quad (6a)$$

$$\frac{\text{not } -A \text{ and } B}{\therefore \text{not both } A \text{ and } B} \quad (6b)$$

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, and the first has already been taken for granted, then he

John ran out of the house (Lakoff 1969a).

All of John's children are intelligent (Lakoff 1973).

The existence of constraints on inference can hardly be doubted. Indeed, the fact that a conjunction can be used for that conjunction to be used best. However, there is no adequate theory of constraints. One solution is therefore to propose schemata that give rise to the familiar patterns of conjunctions and disjunctions. Unfortunately, the rules of inference. They are not

It is frosty.

If it is foggy or frosty, then

Therefore, the game will be cancelled.

For the time being, schemata (6a) and (6b) are "inferences," for reasons that will be described in the next section.

In contrast to the auxiliary inference patterns of inference that have been described, among them the familiar patterns of inference.

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

Therefore, he is intelligent.

There is good reason to suppose

A or B

A or B

is basic. A study by Hill (cited in Lakoff 1973) of a sample of 6-year-old children v

schemata. The first pair permit
nce in a conjunction:

(3a)

John ran out of the house and he got out of bed (Johnson-Laird, 1969a).

(3b)

All of John's children are bald and John has children (Karttunen, 1973).

be inferred from either one of

(4a)

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

(4b)

e inferred from the independent

It is frosty.

(5a)

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

Therefore, the game will be canceled.

(5b)

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.

itions to be deduced:

$\frac{-B}{\text{and } B}$

(6a)

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:

$\frac{B}{\text{and } B}$

(6b)

John is intelligent or he is rich.

He is not rich.

Therefore, he is intelligent.

erns of inference is to find a suit-
number of authors have recently
hat can reasonably be expressed
tion. It may be true, for example,
rew a stone at the frog, but the

There is good reason to suppose that its underlying schema

$$\frac{A \text{ or } B \quad \text{not } -A}{\therefore B} \quad (7a)$$

$$\frac{A \text{ or } B \quad \text{not } -B}{\therefore A} \quad (7b)$$

tone at the frog

ptable. It is customary to suit an
e applies to the relations between
en sentences. Hence, if a speaker

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:

$$\frac{\text{Not both } A \text{ and } B \quad A}{\therefore \text{not } -B} \quad (8a)$$

$$\frac{\text{Not both } A \text{ and } B \quad B}{\therefore \text{not } -A} \quad (8b)$$

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

$$\frac{A \quad \text{If } A \text{ then } B}{\therefore B} \quad (9)$$

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:

$$\frac{\text{Not } -B \quad \text{If } A \text{ then } B}{\therefore \text{not } -A}$$

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 than a single inference. They ma

If the safe is locked, then t
This light is not on.

Suppose the safe is locked.
It follows then that the light
But the light is not on (fro
Therefore, the assumption
of affairs.

Therefore, the assumption

This sort of argument is, of cot
an inferential schema of the for

A impl

It is certainly true that logical (Evans, 1972); and it can be a *tollens*, although a completely co be established at present.

There are two subsidiary pr
First, a conditional may equally cation, because if one propositi fact can be expressed by a conc

Second, where a *reductio* is usec sition, it is necessary to be able e.g., "It isn't the case that 5 is schema makes this elimination :

Its seeming simplicity, howeve of logicians, the intuitionists, These logicians, represented b about certain sorts of mathema cerned with inferences involvin;

that it led to errors only when
 premise, e.g.,

adult subjects inferred that John
 s considered that no conclusion
 g suggests, however, not that the
 n unusual placement of negative
 ion.

lid both for an inclusive disjunc-
 can be true, and for an exclusive
 l out. There is a further rule of
 unctions, e.g.,

she is a genius (but not both).

from the exclusivity of the two
 fore plausible that the basic infer-
 e following way:

$$\frac{A}{\quad} \quad (8a)$$

$$\frac{B}{\quad} \quad (8b)$$

n of inference involving the condi-

$$\frac{\text{then } B}{\quad} \quad (9)$$

t that this schema is basic, whereas
modus tollendo tollens, is not (see
 latter inference has the following

$$\frac{A \text{ then } B}{A}$$

s of this sort but they tend to do
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

$$\frac{A \text{ implies } (B \text{ and not } -B)}{\therefore \text{ not } -A} \quad (10)$$

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 impli-
 cation, because if one proposition can be derived from another, then this
 fact can be expressed by a conditional

$$\frac{A \text{ implies } B}{\therefore \text{ If } A \text{ then } B} \quad (11)$$

Second, where a *reductio* is used to establish the falsity of a negative propo-
 sition, 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:

$$\frac{\text{not not } -A}{\therefore A} \quad (12)$$

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 con-
 cerned with inferences involving infinite sets and argue that such inferences

nciples. They claim that it is
 e existence of a mathematical
 eads to a contradiction. Hence,
 ded middle, i.e., the principle
 true. They consequently reject
 negations. The relation between
 propositions is not so straight-
 3) has shown that the classical
 ntained within the intuitionistic
 ; that the elimination of double
 e now been stated constitute a
 ns of deduction. There are other
 y not basic, are well within the
 t certainly be found for incorpo-
 f such an inference is the *simple*

ompetent.
 e will be forced to resign.
 i he will be forced to resign.
 d to resign.

rally on the horns of a dilemma,
 ves he chooses from the initial
 conclusion. The rhetorical force
 by Cicero (see Kneale & Kneale,
 be considered, for psychological
 e general pattern of inference:

If B then D

then the derived conclusion be-
 rn, is immediately reducible to
 at the simple dilemma is derived
 nt. A comparable chain of infer-
 ndent of the general dilemma,
 This pattern of inference makes
 ositions

B then C
 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 'The picture was painted in 1910' is simply irrelevant in question turns out not to be relevant to the propositional calculus as a model of inference. The conditions always have a truth value. The conditions

The distinction between a material and a truth-value gap may be considered as a clear divergence between the logic and the following bizarre inference for inferences are treated as material im-

You can't both hate Mailer
If you hate Mailer, then you don't admire Mailer
If you admire Mailer, then you don't hate Mailer

Therefore, if you hate Mailer you don't admire Mailer you

The validity of the argument turns out to be true whenever its antecedents are true. The conditional antecedents in the conclusion are true in the conclusion premise.

A more plausible account of the truth value. The semantical rule may be stated as

A conditional *if A then B*
and it is true if and only if *A* is false or *B* is true

The trouble with this analysis, however, is the strong intuition that there is a difference between *A* and *B* in order for the conditional to be true. It also, of course, runs counter to the intuition that, *ex hypothesi*, unfulfilled, e.g., such counterfactual conditionals, a successful painter, then World War III will occur.?' functional.

What happens when you evaluate a conditional whether or not you already assent to its consequent? you already assent to its consequent, then you add it to your set of

d. The conditions in which con-
 else are learned and, from these
 are derived. A competent adult
 ence schemata and their under-

el for the propositional calculus
 nditions:

ue if and only if *A* is false.
 and only if *A* is true and *B* is

only if *A* is true or *B* is true.

linguistic and the other metalin-
 art of a psychological basis for

mply by the lack of any obvious
 distinction between an object lan-
 onditions above, the reader will
 es actually occur as part of their
 objectionable in this practice be-
 ge connectives are being stated in
 ge. However, it is rather unfortu-
 ordinary English. If it is claimed
 ry connectives amounts to learn-
 n a vicious circle is created be-
 ge of the meaning of ordinary
 been overlooked by many of the
 bases for natural language (e.g.,
 involves some more abstract form
 istic information about natural

ical rules concerns the interpreta-
 s to the heart of the problem of
 is of a psychological model. Con-
 course, capable of a great many
 may be used to state temporal,
 sitions. It is only relatively rarely
 ulus, for example, in conveying
 on is true provided its antecedent
 re, e.g., "If this picture isn't by
 y of everyday conditionals, how-
 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 impli-
 cation is true whenever its antecedent is false, and one of the two condi-
 tional 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 argu-
 ment you add it to your set of beliefs and then consider whether or not