1

Reasoning without Logic

P. N. JOHNSON-LAIRD

MRC Applied Psychology Unit
Cambridge

INTRODUCTION

The psychology of reasoning should characterize both inferential competence—what the mind has to compute and why—and inferential performance—the mental processes underlying these computations. Unfortunately, there is no existing theory of what the mind computes in reasoning, and the resulting theoretical gap has been filled by logic. Psychologists suppose that since logic encompasses the set of valid deductions, it is reasonable to assume that ideal human competence is in accordance with logical principles. This assumption leads very naturally to the doctrine that the mind *contains* a logic, and to a theory of performance in which deductions are made on the basis of formal rules of inference, or inference schemata. Thus logic is implanted in the mind in much the same way that grammar is supposedly located there.

The doctrine of mental logic is highly plausible. It is embodied in the philosophical belief that laws of thought are nothing but laws of logic (Boole, 1854; Mill, 1874), and it is embraced by all current theories that assume that human beings are potentially rational (e.g., Beth & Piaget, 1966; Braine, 1978; Henle, 1962; Inhelder & Piaget, 1958; Osherson, 1975; Rips, 1983; Youniss, 1975). In addition, it was tacitly accepted by my colleagues and myself for many years (e.g., Johnson-Laird, 1975; Wason & Johnson-Laird,

REASONING AND DISCOURSE PROCESSES

ISBN 0-12-512320-5

1972). What I argue in this chapter, however, is that the doctrine is false. In my view, logic does not make a good theory of deductive competence, and theories of performance based on it are defective. It is necessary to give up, not the thesis that human beings are capable of rational thought, but the notion that what underlies this ability is logic. There *can* be reasoning without logic. Indeed, once the search for a logic of the mind is abandoned, it is possible to make better sense of the psychology of reasoning.

THE DOCTRINE OF MENTAL LOGIC

My first task is to examine the doctrine of mental logic in some detail, since it is important to make clear precisely what is at stake. Logicians formalize logic in many ways, but a crucial component of a formalization is the introduction of rules of inference, or inferential schemata. A typical rule of inference is *modus ponens*:

$$\begin{array}{l} p \\ p \rightarrow q \\ \hline \therefore q \end{array}$$

This rule stipulates that if there are premises of the form p, and p implies q, then the conclusion q is derivable.

A rule of inference can be applied to any set of premises, regardless of their contents, provided they have the appropriate logical form. Hence, according to the doctrine of mental logic, an individual makes the following deduction:

> The warning light is on.
> If the warning light is on, then the system is defective.
> Therefore, the system is defective.

by relying on an internalized representation of modus ponens. The first premise corresponds to p, and the second premise corresponds to $p \rightarrow q$, since 'if p then q' can be treated as an implication, and the conclusion corresponds to q. Proponents of mental logic assume that inferences are based on some such formal system, but they disagree among themselves about its precise nature (see, e.g., Braine, 1978; Johnson-Laird, 1975; Osherson, 1975; Rips, 1983).

The main virtue of mental logic is that it solves the otherwise profound riddle of how it is possible to reason validly. The question is puzzling because the invention of logic as an intellectual tool would seem to require the ability to reason soundly, and yet this ability would seem in turn to depend on a recourse to logic. The doctrine of mental logic resolves the riddle: people are able to reason validly because they have a logic in their minds, and the invention of formal logic depends on the 'externalization' of this inner system.

If one accepts the existence of mental logic, then a second riddle arises: How does it get into the mind in the first place? The puzzle here is that children might need to be able to reason validly in order to acquire a logic, but if they can reason validly, they might not need a mental logic. Some writers, inspired by Chomsky's theories, have drawn parallels between the acquisition of logic and the acquisition of language (e.g., Falmagne, 1980). Children may learn logic by encountering valid inferences in verbal guise, and by abstracting from them their underlying logical form in much the same way that grammatical rules are supposed to be acquired. The problem with this conjecture, at least as a complete explanation of the development of mental logic, is that it offers no account of how children recognise valid inferences when they encounter them. It is not as though adults engage in sustained demonstrations of valid deduction or take steps to instruct their offspring in the canons of logic. Consequently, there is a natural temptation to propose that logical competence is inborn (Fodor, 1980), just as the principles of 'universal grammar' are considered to be innate.

So far, this appeal to the powers of evolution has not been followed up by an account of how logic evolved. Indeed, the argument proceeds by default: there is no direct evidence for it, merely the lack of a convincing alternative. Piaget tried to escape the dilemma by arguing that logic is neither innate nor learned by the ordinary principles of reinforcement. He claimed that children 'construct' logic by internalizing their actions and by reflecting on each step in their operational development (see, e.g., Inhelder & Piaget, 1958). Unfortunately, it is unclear whether such a process could work, since it has never been formulated in an explicit algorithm. Nevertheless, the doctrine of mental logic—whether learned, innate, or constructed—remains appealing and has had no competitors. There are, however, three difficulties associated with it.

First, the existence of a mental logic would suggest that human beings are intrinsically rational. Yet experimental results, anecdotal observations, and the well-known 'budgets' of fallacies imply that few people have a secure grasp of the principles of valid inference. Proponents of mental logic explain such mistakes in the following way: they arise, not from faulty inferential principles, but from errors in the application of the rules, from misinterpreting premises, or from making unwarranted assumptions (see, e.g., Cohen, 1981; Henle, 1962; Revlis, 1975; Staudenmayer, 1975). Doubtless, people do misunderstand or forget premises; they do import extraneous or irrelevant considerations into their reasoning; and they do fail to stick to pure logic. But the trouble with this defence of reason is that it is hard to see how, in principle, any empirical findings could shake it. "I have never found errors which could be unambiguously attributed to faulty reasoning," remarks Henle (1978) characteristically. But it is always possible to provide alternative explanations for an error, and it is not usually clear how to decide which one is correct. The defence seems tendentious, as if it had sprung from a prior conviction that conscious thought is invariably rational.

Second, if there are formal rules of inference in the mind, they should apply regardless of the content of problems. However, there is evidence that human reasoners are affected by the content of premises (see, e.g., Wason & Johnson-Laird, 1972). The circumstances and the reasons for such effects remain highly controversial, but they undoubtedly occur and thus present an unsolved problem for the doctrine of mental logic.

Third, logic cannot determine which particular valid conclusions people draw spontaneously. A valid conclusion is one that must be true if the premises are true, and a complete logic sanctions all of the valid conclusions that follow from a set of premises. In fact, though, any set of premises implies an infinite number of different valid conclusions. For example, given the premises:

> If the warning light is on, then the system is defective.
> The warning light is on.

most people draw the valid conclusion:

> The system is defective.

But logic sanctions the following conclusion, too:

> The warning light is on, or if the warning light is on, then the system is defective.

and so on ad infinitum with any number of repetitions of the premises. Similarly, the rule of inference:

$$\frac{p}{\therefore\ p \text{ or } q}$$

sanctions such valid conclusions as:

> The system is defective or the government has abandoned monetarism.

No sane person would draw any of these conclusions, even though they are logically valid. Psychologists have therefore to specify a theory of competence that accounts for the inferences that people *do* make. This theory must embody principles that lie outside logic altogether, because they will determine which of a potentially infinite set of valid deductions the human inferential mechanism actually produces. Proponents of mental logic have had little, if anything, to say about this problem, probably because they have concentrated on experimental procedures in which subjects evaluate given conclusions or choose between a small set of different conclusions.

If there is a logic in the mind, then a major task for psychology is to discover which particular type of logic it is, how it is specified, and how it is acquired. Unfortunately, very little is known about these matters, and there is controversy about all of them. Still, these disagreements do not directly threaten the doctrine of mental logical. Indeed, there hardly appears to be a viable alternative to it, and it has seemed at times to be irrefutable (Cohen, 1981). Why, then, is the doctrine to be abandoned? The answer summarizes the main aims of the present chapter. It establishes three principal points:

1. Logic alone is not sufficient for a theory of competence. There must be extra-logical principles to account for spontaneous deductions.

2. Logic is not necessary for a theory of competence. All major types of deduction can be made without recourse to it or to formal rules of inference (or inferential schemata). This claim extends to types of deduction for which there are presently no formal calculi and to those for which there can be no complete calculi.

3. There is an alternative theory of competence that specifies which inferences people draw. It in turn leads to an alternative theory of performance that specifies the mental processes underlying inference. This theory, unlike those based on mental logic, ranges over all types of deduction. It is more parsimonious, and where the two approaches diverge, it is better supported by the experimental evidence.

In order to establish these three points in a convincing way, it is necessary to consider different sorts of inference. The chapter accordingly deals with all of the major types of inference that logicians and psychologists have traditionally analyzed. There are six main sections: First, it outlines a theory of semantic information, which is used to specify which inferences people draw spontaneously. Second, it considers the ways in which people reason with propositions and the connectives that can join them, and it contrasts the approach of mental logic with an alternative that makes no use of rules of inference. Third, it considers the same contrast for reasoning that depends on quantifiers (such terms as 'all', 'some', and 'no'). Fourth, it demonstrates the advantages of abandoning mental logic in elucidating relational inferences. Fifth, it draws the same moral for the 'common sense' inferences of daily life. Finally, the chapter draws some brief conclusions in which an important limitation on the present approach—namely, its inability to account for deductions about infinite sets—is spelled out.

Table 1.1
A Truth Table Defining the Truth-Functional Sense of *p* and *q*, where *p* and *q* are Variables Ranging over Propositions

p	*q*	*p* and *q*
True	True	True
True	False	False
False	True	False
False	False	False

SEMANTIC INFORMATION AND INFERENTIAL HEURISTICS

Truth Tables and Truth-functional Reasoning

There are many words in English that can serve the function of interrelating clauses expressing propositions (e.g., 'and', 'or', 'if', 'because', 'before', 'when', and 'until'). A subset of them have meanings that can be treated as mapping the truth values of the clauses they connect onto a truth value for the sentence as a whole. 'And', for example, can be treated as yielding a true sentence if and only if the two clauses it connects both express true propositions. (There are obviously other senses of 'and' that cannot be analyzed in this way, as in "Charles and Diana are a splendid couple".) This way of talking about truth values is not appropriate when one is specifying formal rules of inference such as modus ponens, which should be thought of as purely syntactic formulae that allow one expression to be derived from others. It is appropriate, however, when one is defining the semantics of connectives by stating the conditions in which they are true and those in which they are false. It is crucial in what follows to bear in mind this contrast between the derivation of formulae in a quasi-syntactic way and the definition of their truth conditions.

A familiar device for defining 'truth-functional' connectives is the so-called 'truth table'. Table 1.1 defines the truth-functional sense of 'and'. The table shows that an assertion of the form *p and q* is true if *p* is true and *q* is true, but it is false for any other combination of truth values for *p* and *q*.

A formal logic for truth-functional connectives can be formulated using such rules as modus ponens; it is known as the propositional (or sentential) calculus. This calculus has been assumed by some psychologists to constitute the logic of the mind: "Reasoning is nothing more than the propositional calculus itself" (Inhelder & Piaget, 1958, p. 305). Other psychologists have merely assumed that certain inferences depend on truth-functional connectives and have tried to devise a psychologically realistic set of inferential schemata, following the idea of a 'natural deduction' system in which there are schemata for each connective (see Braine, 1978; Johnson-Laird, 1975; Osherson, 1975; Rips, 1983). No matter how the logic is formalized, a theory of competence still needs to place appropriate constraints on the deductive machinery in order to characterize the class of deductions that people draw spontaneously. Such principles can be developed from the concept of semantic information.

A Measure of Semantic Information

Psychologists are familiar with the statistical concept of information that derives from the work of Shannon and others; it is normally taken to have nothing directly to do with the meaning of messages (Shannon & Weaver, 1949). Nevertheless, a measure of semantic information can be developed from the plausible assumption that the more states of affairs that a proposition rules out, the greater the amount of semantic information it contains (see Bar-Hillel & Carnap, 1952/1964). This measure is particularly suitable for the analysis of propositional reasoning. Given a truth table, a proposition that eliminated half of its contingencies would be more informative than one that eliminated only a quarter of them. It follows that a categorical proposition such as "The warning light is on" is more informative than the inclusive disjunction, "The warning light is on or the system is defective (or both)", since the categorical eliminates half of the contingencies of any truth table into which it enters, whereas the disjunction eliminates only a quarter of the contingencies. Likewise, the conjunction, "The warning light is on and the system is defective," is more informative than a categorical proposition, since the conjunction eliminates three quarters of the truth table.

Most people have an intuitive grasp of the these differences. What may not be obvious at first glance, however, is that if the measure of semantic content is defined as the proportion of contingencies in the truth table that an assertion eliminates, then the measures above are independent of the number of contingencies in the table. A simple categorical premise always eliminates half of the contingencies, a conjunction eliminates three quarters of them, and a disjunction eliminates one quarter of them. Table 1.2 presents a number of such assertions and shows the contingencies they exclude. There is a simple way to compute the semantic information conveyed by any expression in the propositional calculus (see Johnson-Laird, 1983).

Table 1.2

The Relative Informativeness (I) of Different Sorts of Proposition[a]

Contingencies			Propositions and I Values						
a	*b*	*c*	*a* or *b* or *c* (0.125)	*a* or *b* (0.25)	(*a* & *b*) or *c* (0.375)	*a* (0.5)	(*a* or *b*) & *c* (0.625)	*a* & *b* (0.75)	*a* & *b* & *c* (0.875)
T	T	T							
T	T	F					X		X
T	F	T						X	X
T	F	F			X		X	X	X
F	T	T				X		X	X
F	T	F			X	X	X	X	X
F	F	T		X		X	X	X	X
F	F	F	X	X	X	X	X	X	X

[a] X indicates that a premise is inconsistent with the corresponding contingency.

Inferential Heuristics

No valid inference can yield a conclusion with a greater amount of semantic information than the premises on which it is based. Three constraints that characterize the valid inferences that people draw spontaneously can now be stated:

(1) Conclusions should not be based on throwing away semantic information from the premises. Thus, for example, given the premise:

The system is defective.

one should not expect anyone to draw the conclusion:

> The system is defective, or the government has abandoned monetarism.

since this conclusion, though valid, contains less information than the premise.

(2) Conclusions should express the semantic content of premises more parsimoniously; that is, conclusions should contain fewer occurrences of the connectives, relations, quantifiers, predicates, or atomic propositions contained within the premises. Thus, for example, given premises of the form:

If *p* then *q*
p

people do not normally draw the following valid conclusion, although it does maintain the content of the premises:

p, and if not-q then not-*p*, and *p*.

This conclusion is plainly less parsimonious than the premises.

(3) As a corollary of the previous heuristic, categorical premises should not be repeated as constituents of conclusions, since they can be taken for granted. Given such premises as:

John is lazy or he is rich. (*p or q*)
John is not lazy. (*not-p*)

then the conclusion:

John is not lazy and he is rich. (*not-p and q*)

is parsimonious and maintains the content of the premises. However, reasoners generally draw the simple conclusion:

John is rich. (*q*)

since they apparently find it unnecessary to repeat the categorical premise, not-*p* (see Johnson-Laird & Tridgell, 1972). Hence, this third principle should be taken to apply to the formulation of the conclusion rather than to its derivation. Likewise, given the relational premises:

Ann is taller than Bette.
Bette is taller than Carol.

the conclusion that subjects tend to draw (Johnson-Laird & Bara, 1983) is:

Ann is taller than Carol.

rather than the repetitious conclusion:

Ann is taller than Bette, who is taller than Carol.

Once a theory of competence has been developed, there are two ways in which to go about specifying a theory of performance. First, like Chomsky (1965), one can simply assume that the principles of competence are built directly into the mind. Thus, a knowledge of the heuristics outlined above might be part of the inferential system and act to filter out any deductions that fail to meet their constraints. Second, one can develop an independent theory of performance from which the principles of competence are emergent properties. That is, the performance algorithm abides by the competence principles even though it does not directly incorporate them. In what follows, I will show—beginning with propositional reasoning—that there is a plausible theory of performance of this second kind, which also illustrates how valid deductions can be made without recourse to formal rules of inference.

PROPOSITIONAL REASONING

A Simple Algorithm

There is a simple algorithm for propositional reasoning that does not rely on formal rules of inference but rather on the truth conditions of connectives (see Johnson-Laird, 1983). This knowledge is required in any case in order to understand or verify assertions containing the connectives. It is instructive to consider how the algorithm works before taking up the question of the extent to which it is independent of logic. Its first step, given such premises such as:

p and not-*q*, or *r*

q

is to determine which premise (if any) conveys most semantic information. Here, the second premise, *q*, conveys more information than the first. The algorithm then checks whether the more informative premise occurs elsewhere as a constituent of another premise. Since, in fact, the second premise occurs as a constituent of the first premise, the value 'true' can be substituted for its occurrence there:

p and not-true, or *r*

The truth conditions of 'not' in the propositional calculus are such that if a proposition is not true, then it is false; and if it is not false, then it is true. This semantics justifies the following substitution:

p and false, or *r*

The truth conditions of 'and' are such that a conjunction is false whenever one of its conjuncts is false. Hence, the conjunctive constituent of the premise can be simplified to false, with the result that the premise as a whole becomes:

false or *r*

The meaning of 'or' ensures that if a disjunction is true and one of its constituents is false, then the other constituent must be true. The application of this semantics to our example yields the conclusion, *r*.

This semantic procedure automatically ensures that no information is lost (the first heuristic in the competence theory) and that any forthcoming conclusion is more parsimonious than the premises on which it is based (the second heuristic). Since the algorithm operates on the premise into which the truth value is substituted, it does not repeat the categorical premise on which the inference is based (the third heuristic).

The general algorithm can be made completely explicit. To check its consistency, I implemented it in a computer program written in the list-processing language, POP-10. The program requires four steps:

(1) Find the maximally informative premise (of those that have not yet been tried as the basis for an inference) according to the measure of semantic content described above. Ordinary reasoners are unlikely to make anything other than an intuitive assessment of informativeness. If all the premises have been tried with no clear result, then no conclusion is drawn. If there is no single premise that is maximally informative, then a premise is chosen at random from the set of those that are maximally informative.

(2) If there is another premise that contains the one selected in the first step, then substitute in it the value 'true' for the one selected. If the first premise is negative and its unnegated proposition is a constituent of the second, then substitute the value 'false' for that constituent. If no substitution is possible, return to step 1.

(3) Use the semantics of the connective governing the substituted truth value to simplify the premise compositionally. If the result is a proposition that contains a truth value and a connective, then continue to carry out this step until the process can go no further.

(4) If the final result is a proposition that does not contain a truth value, then this proposition is the conclusion to be drawn. If the final result is the truth value 'true', then the premises are consistent but there is no conclusion of interest to be drawn. If the final result is the truth value 'false', then the premises are inconsistent.

A psychologically plausible algorithm should refrain from making trivial, though valid, deductions that are never drawn by human reasoners. This algorithm does indeed refrain from drawing trivial conclusions. For example, given premises of the form:

p
p or *q*

it substitutes a truth value for *p* in the second premise:

true or *q*

The semantics of 'or' in its inclusive sense then leads to the following simplification: true. This end-product merely establishes that the premises are consistent with each other, and therefore no conclusion is drawn. The same result occurs with premises of the form:

p
q

since the procedure is unable to make any substitution.

An algorithm for drawing conclusions spontaneously is obviously distinct from one for evaluating given conclusions. However, there is a simple way of extending the algorithm to deal with the evaluation of conclusions. It relies on the fact that an inconsistent set of assertions simplify to the truth value 'false'. If a putative conclusion has to be evaluated with respect to some premises, all that is necessary is to add its negation to the premises and then to determine what follows in the usual way. For example, suppose that the deduction:

not-*q*
p or *q*
∴ *p*

has to be evaluated. The negation of the conclusion is added to the premises:

not-*q*
p or q
not-*p*

The procedure substitutes 'false' for *q* in the second premise and simplifies it to *p*. This is a constituent of the negated conclusion, *not-p*, and the procedure substitutes the value 'true' for it:

not-true

which simplifies to false. This result shows that the negation of the conclusion is inconsistent with the premises. That is, the conclusion follows validly from the premises.

Not all propositional inferences derive from treating one premise as a constituent of another. The valid deduction,

p and *q*
not-*p* or *r*
not-*q* or *s*
∴ *r* and *s*

contains no premise that is a constituent of another. There is a very simple strategy at work here: Break down the maximally informative premise into its two constituents (*p*, *q*), use each of them as separate categorical premises in the procedure to yield separate conclusions (*r*, *s*), and join these conclusions by whatever connective was in the original premise (*r* and *s*). This method can be used recursively in order to break down a complex proposition into constituents that in turn need to be broken down, and so on. It is important, however, that the connectives are symmetrical. Since 'if *p* then *q*' is not symmetrical (i.e., it is not equivalent to 'if *q* then *p*'), conditionals must be translated into a symmetrical, truth-functional form: *not-p or q*, which is equivalent to *q or not-p*.

Truth Conditions Versus Rules of Inference

The algorithm for propositional reasoning relies on the truth conditions of connectives rather than on formal rules of inference. What is the difference, if any, between these two approaches? The algorithm needs access, in effect, to any row of the truth tables that define the meanings of the connectives. For example, it requires a representation of conjunction that is equivalent in force, though not necessarily in form, to the truth table in Table 1.1.

A system based on formal rules of inference (e.g., Braine, 1978; Johnson-Laird, 1975; Osherson, 1975; Rips, 1983) has no access to truth conditions but relies instead on rules of inference or inferential schemata such as

$$\begin{array}{l} p \\ q \\ \hline \therefore\ p \text{ and } q \end{array} \qquad \begin{array}{l} p \text{ and } q \\ \hline \therefore\ p \end{array} \qquad \begin{array}{l} \text{not-}p \\ \hline \therefore\ \text{not } (p \text{ and } q) \end{array}$$

These rules can be used to derive a conclusion from premises without any knowledge of the truth conditions of the connective. They are part of the formal or syntactic formulation of a logical calculus, but truth tables and statements of truth conditions are part of a quite distinct enterprise—the formulation of a 'model-theoretic' semantics for the calculus. Validity is in fact a semantic notion, whereas derivability from formal rules is a syntactic notion. An important part of meta-logic is, indeed, to establish that the formal logic is complete (i.e., any inference that is valid according to the semantics is derivable within the formal syntactic system). A major logical discovery of the twentieth century is that certain calculi are incomplete, in that there is no way of framing formal rules that capture all and only the set of valid inferences (see, e.g., Boolos & Jeffrey, 1980).

An adherent of mental logic is likely to note that the rows of a truth table can be translated into formal rules of inference. For example, the first schema just shown corresponds to the first row of Table 1.1. Nevertheless, it does not follow that the two approaches are identical. This fact is readily grasped by logicians, who are familiar with the contrast between a formal calculus and a semantic model for it; however, 'mental logicians' often have difficulty understanding the contrast. The crux of the matter is that although a statement of truth conditions may be mapped unequivocally onto rules of inference, the converse does not follow. Once this point is grasped, the distinction between the two approaches becomes obvious.

Any formal calculus is open to more than one semantic interpretation. The propositional calculus, for example, has a different interpretation in which the connectives are given an 'intuitionistic' interpretation. Roughly speaking, instead of truth one talks of provability in a constructive way, and instead of falsity one talks of absurdity (see Kneale & Kneale, 1962). Likewise, to consider a more accessible case, if the variables *p*, *q*, and so on are taken to range over sets rather than propositions, then *p* can be taken to mean that the set *p* is not empty, and the connectives can be redefined. Thus '*p* and *q*' means that the intersection of the sets *p* and *q* is not empty; '*p* or *q*' means that the union of the sets *p* and *q* is not empty; and '$p \rightarrow q$', the formula for material implication, means that the intersection of *p* and the complement of *q* is empty. In other words, all *p* are *q*. The formal rule of inference, modus ponens, which together with a substitution rule and appropriate axioms, suffices for deriving all valid theorems in the calculus:

$$p$$
$$p \rightarrow q$$
$$\therefore q$$

means that given that p is not empty, and that all p are q, it follows that q is not empty.

The moral is clear: the meaning of a connective is not defined merely by specifying formal rules of inference for it. Admittedly, proponents of mental logic have usually stipulated that their systems are supposed to model reasoning with propositions, but any such stipulation merely smuggles some semantics in by the back door. My point is that semantics should be allowed in by the front door, because then formal rules can be thrown out of the house completely.

How can one decide which approach children adopt? Braine and Rumain (1981) argue that children acquire rules of inference for 'or' before they acquire the truth conditions of the connective. They base this claim on the fact that children are able to make inferences before they are able to evaluate the truth or falsity of disjunctive assertions. Although this finding is of considerable intrinsic interest, it fails to decide the issue. One must not beg the question either by assuming that inferences can be made only by relying on *rules* of inference (as opposed to a knowledge of truth conditions), or by assuming that failure in a verification task shows that the truth conditions are not known to the child.

It is difficult to establish that formal rules of inference are acquired prior to truth conditions, because it would be necessary to show that children make such inferences as:

John is lazy or John is rich.
John is not lazy.
Therefore, John is rich.

purely by virtue of their logical form:

$$p \text{ or } q$$
$$\text{not-}p$$
$$\therefore q$$

and without knowing the truth conditions of the connective. Braine and Rumain's experiment does not, of course, demonstrate this point. In this regard, it is worth remarking that tests of comprehension often reflect a poorer grasp of meaning than does spontaneous speech (Braine & Rumain, forthcoming), and even four-year-olds understand that the command, "Do A or do B," is satisfied by carrying out one of the tasks (Johansson & Sjolin, 1975). Clearly, children may master some aspects of truth conditions long before they are able to make inferences.

A disinterested reader might suppose that I am making a fuss about a very minor distinction—the difference between formal rules and semantic interpretations—but the divergence between the two approaches increases as the chapter proceeds. Even within the domain of sentential connectives, the semantic approach has several advantages. It solves at a stroke the problems of which particular logic is in the mind, how it is mentally represented, and how children acquire it. These problems simply fail to arise, because there is no logic in the mind. Of course, it is necessary to explain how the truth conditions of connectives are acquired, but this task would still be necessary even if there were mental rules of inference, because children have to learn these truth conditions in order to determine the truth or falsity of assertions and to use the connectives appropriately.

A single, uniform semantics can be offered for 'and' (Gazdar, 1980) which embraces both its truth-functional use in connecting sentences and its other uses in connecting other constituents. This semantics suffices for the algorithm for propositional reasoning, whereas special and additional rules of inference are required to handle all of its uses within a formal calculus. Moreover, there are some sentential connectives, such as 'because' and 'until', which are clearly not truth-functional in their meanings, and for which there are no existing formal calculi. Perhaps these facts make it easier to grasp the point that children may initially acquire a knowledge of the truth conditions for these connectives before they acquire (if they do at all) the as yet unknown formal rules of inference for them.

Theories of mental logic and the propositional algorithm have a limited explanatory power. They are only as real as the phenomenon of truth-functional reasoning, and, as has been known for some time, people do not ordinarily think in this way. They are generally more interested in interrelating propositions within a model of causal or intentional relations. Wason and Johnson-Laird (1972, Chap. 7) show that when causal considerations conflict, as they often do, with those of truth-functional logic, subjects follow causality. This problem will be resolved later in the chapter by an account of a general semantic procedure that is applicable both to causal and propositional reasoning.

QUANTIFICATIONAL REASONING

Syllogisms

The validity of many deductions depends, not just on sentential connectives, but also on the internal structure of premises and on the occurrence within them of quantifiers (e.g., 'all', 'any', 'most', 'many', 'some', 'several', 'few', and 'none'). In this section I consider how people make these inferences, establish that the two approaches—the formal and the semantic—diverge from each other, and present some experimental evidence that supports the semantic approach.

Aristotle formulated a logic for a subset of quantified inferences that he called 'syllogisms', and they have been the focus of most psychological studies of reasoning with quantifiers. A syllogism consists of two premises and a conclusion. These assertions can occur in one of four 'moods':

All *X* are *Y*.
Some *X* are *Y*.
No *X* are *Y*.
Some *X* are not *Y*.

The arrangement of the terms in the premises can also occur in one of four 'figures':

A–B	*B–A*	*A–B*	*B–A*
B–C	*C–B*	*C–B*	*B–C*

Granted that there are four moods for each premise and four figures, there are 64 possible logical forms for the premises of a syllogism.

Although psychologists have studied syllogistic inference since the turn of the century (Störring, 1908), they have concentrated on explaining errors in evaluating one or more given conclusions (e.g., Chapman & Chapman, 1959; Wilkins, 1928; Woodworth & Sells, 1935) and have assumed that logic—and, strangely, the logic of the medieval Schoolman—characterizes human competence. It was not until recently that experiments were carried out in which subjects were asked to state in their own words what conclusion, if any, followed from syllogistic premises (Johnson-Laird, 1975; Johnson-Laird & Bara, 1984; Johnson-Laird & Steedman, 1978). A striking phenomenon emerged: some syllogisms almost always elicit a correct response, whereas others hardly ever do so. Here is an example of a pair of premises that almost always yields a correct response:

Some of the scientists are parents. (Some of the *A* are *B*)
All the parents are drivers. (All of the *B* are *C*)

Nearly everyone (90% of subjects) draws the valid conclusion:

Some of the scientists are drivers. (Some of the *A* are *C*)

and a few subjects (5%) draw the converse and equally valid conclusion:

Some of the drivers are scientists. (Some of the *C* are *A*)

The bias in the form of the response is an example of a general effect of figure. Given premises in the figure:

A–B
B–C

the vast majority of conclusions that are drawn take the form:

A–C

whereas given premises in the figure:

B–A
C–B

the vast majority of conclusions take the form:

C–A

The figural effect, which my colleagues and I have observed in every subject whom we have tested, was not previously noticed because experimenters relied on scholastic logic (which recognizes only conclusions of the form *C–A*) and on procedures requiring the evaluation of one or more given conclusions.

Here is an example of a pair of premises to which few subjects make the correct response:

All of the beekeepers are artists.
None of the chemists are beekeepers.

The responses of the subjects in a typical study (Experiment 2 of Johnson-Laird & Steedman, 1978) were as follows:

None of the chemists are artists.	(60% of subjects)
None of the artists are chemists.	(10% of subjects)
No valid conclusion	(20% of subjects)
Some of the chemists are not artists.	(10% of subjects)
Some of the artists are not chemists.	(0% of subjects)

The last of these responses, which was made by none of the subjects, is in fact the only valid conclusion. Clearly, a theory of deductive performance should account for the extreme and systematic range of difficulty of syllogisms.

Theories of the Syllogism

Theories of performance with syllogisms have only recently begun to be proposed. They have generally been based on well-known logical techniques, and they have been intended to account only for the evaluation of given conclusions. The most familiar technique for syllogistic inference is the method of Euler circles, which is often confused with the rather different method of Venn diagrams. Euler circles rely on diagrams in the Euclidean plane in which there is a different circle standing for each set referred to in the premises. A premise of the form, "All *A* are *B*," calls for two separate diagrams. In one, the circle standing for *A* lies entirely within the circle standing for *B* to represent the possibility that set *A* is wholly included within set *B*; in the other, the two circles lie on top of one another to represent the possibility that the two sets are co-extensive. Since 'some' is construed by logicians to mean 'at least some' and is accordingly consistent with 'all', a premise of the form, "Some *A* are *B*," needs four different diagrams: *A* overlapping *B*, *A* included in *B*, *B* included in *A*, and *A* co-extensive with *B*.

The Eulerian method depends on drawing the appropriate diagrams for each premise and then combining them. In order to guarantee that a conclusion is valid, it is necessary to check that it holds for all possible combinations of the diagrams for the two premises. This is a complex operation, because there is no simple algorithm for ensuring that all the combinations have been considered, and the total number of them is generally greater than the mere product of the numbers of separate diagrams for the two premises. For example, the reader may care to try to construct the set of combinations for the premise:

Some *A* are *B*.

which calls for four diagrams, and the premise:

All *B* are *C*.

which calls for two diagrams. In informal tests, I have not found anyone who has succeeded in constructing all of the logically distinct combinations (see below).

Erickson (1974) has proposed a theory of syllogistic inference based on the assumption that people form mental representations of the premises that are isomorphic to Euler circles. In order to deal with the embarrassing number of combinations, he assumed that subjects often fail to consider the full set of different ways in which an individual premise can be represented, and (in one version of his theory) that they construct only one of the many possible combinations of representations. Guyote and Sternberg (1981) devised a similar theory, based on strings of symbols corresponding to Euler circles, in which they limited the number of combinations that subjects consider to a maximum of four. Both theories accordingly view human beings as intrinsically illogical with respect to many syllogisms—including, of course, the easy one based on the two premises above. Both theories also rely on a fundamentally irrational mechanism, the production of conclusions that match the 'atmosphere' of the premises, in order to deliver valid conclusions for certain syllogisms. Hence, although subjects may reach the right conclusion, these theories imply that they hardly ever do so for the right reasons.

The major weakness of any theory based on Euler circles is that it is bound to make the wrong predictions about the relative difficulty of drawing the correct conclusion from different types of syllogistic premises. On the one hand, the easy problem that nearly everyone gets right, based on premises of the form:

Some of the *A* are *B*.
All the *B* are *C*.

requires 16 different Euler diagrams to be constructed. On the other hand, the difficult problem that hardly anyone gets right:

All the *B* are *A*.
None of the *C* are *B*.

requires only six different diagrams. If, following Erickson, we assume that premises of the form, "All *X* are *Y*," are represented by only one diagram (in which *X* is wholly included within *Y*), and that premises of the form, "Some *X* are *Y*," are represented by only two diagrams (one in which *X* and *Y* overlap, and the other in which *Y* is included in *X*), the position is hardly any better. The easy problem requires six different diagrams and the hard problem requires five. Evidently, the number of diagrams to be constructed bears no relation to the difficulty of the problems.

Euler diagrams have a further explanatory inadequacy: they cannot be used to represent inferences that depend on sentences containing more than one quantifier, such as:

Not all the critics liked all the pictures.
Therefore, some critic did not like some picture.

Thus the case against theories based on Euler circles seems to be overwhelming.

Proponents of mental logic will rightly reject any theory of syllogistic reasoning that is unable to account for the potential rationality of human beings. They have two options: They can either adopt a more tractable logical tool for syllogisms, or they can assume that mental logic consists of the standard (first order) quantificational calculus.

Newell (1981) chose the first option and put forward a psychological theory of syllogisms based on a symbolic equivalent of a superior topological technique—the method of Venn diagrams. Newell's theory, however, is more an account of competence in evaluating conclusions and an illustration of how such a theory can be expressed within his framework for problem-solving (Newell & Simon, 1972) than it is a full-fledged theory of performance. Thus, the theory makes no specific predictions about errors. Moreover, there is an alternative (and simpler) theory on the same general lines which depends on the close relation between Venn diagrams and truth tables (see the alternative interpretation of the propositional calculus, which I described earlier). For example, the premises:

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

yield the following evaluations in the table of contingencies based on *A*, *B*, and *C*.

	A	*B*	*C*	
(first premise establishes one or other)	+	+	+	
	+	+	−	(eliminated by second premise)
	+	−	+	
	+	−	−	
	−	+	+	
	−	+	−	(eliminated by second premise)
	−	−	+	
	−	−	−	

The first premise establishes that the intersection of *A* and *B* is not empty, and the second premise establishes that the intersection of *B* and the complement of *C* is empty. Since the first contingency in the table is thereby definitely established, it is clear that "Some *A* are C." There are 16 different combinations of the Euler diagrams representing the premises, since they correspond to the possible ways of combining those four contingencies that involve the members of at least one of the three sets, and that may or may not be empty because they are neither established nor eliminated by the premises (rows 3, 4, 5, and 7).

The second option of building the predicate calculus into the mind has been implicitly adopted by linguists such as Chomsky (1977) for some time. They assume that assertions have a logical form that calls for the quantifiers and variables posited by the predicate calculus. The major difficulty with this assumption is the psychological implausibility of the rules of inference called for by the predicate calculus. There are, of course, different ways of formalizing the calculus, but they all rely on rules of inference that are not intuitively obvious.

The function of the rules is to get rid of quantifiers and then to reason propositionally, since the propositional connectives are part of the predicate calculus. The universal quantifier, 'for any *X*', is dropped by the rule of 'universal instantiation', which allows the quantifier to be replaced by any constant denoting an individual. This rule is readily grasped, since it merely formalizes the notion that if some predicate applies to everything, then it applies to, say, Fred. The rule of 'existential instantiation' for dropping the existential quantifier, 'there is at least some *x*', is not so obvious. It allows an existential quantifier to be dropped in favour of any constant denoting an arbitrary individual, provided that this individual has not been referred to in the argument. In other words, if a predicate applies to someone, then it is true of some particular individual, and we can assume that this individual is, say, Tom, provided that he has not been referred to anywhere else in the premises.

The notion of an arbitrary constant can cause trouble even to students of logic, and no psychologist is likely to assume that it, or the technique of dropping quantifiers, plays any part in ordinary reasoning. Unfortunately, the only way to avoid the rule of existential instantiation is to opt for a system based on a single rule of inference, the so-called 'resolution' rule, but this rule requires a very unnatural translation of assertions into a uniform disjunctive format in which existential quantifiers are represented by a special functional notation.

Although a mental logic based directly on the predicate calculus has not so far been explicitly advocated by any psychologist, Braine and Rumain (forthcoming) have constructed an ingenious set of inferential schemata that, in essence, build quantifiers into the required propositional rules. This theory, however, is primarily an account of rational competence; it makes few, if any, systematic predictions about errors in performance. Also, it cannot as yet account for the difference between the easy and the difficult syllogistic problems.

In summary, existing theories of reasoning with quantifiers divide into two main sorts. On the one hand, there are theories of performance which are based on logical techniques for the syllogism. These predict some of the errors that occur in performance but regard human beings as essentially irrational (Erickson, 1974; Guyote & Sternberg, 1981). On the other hand, there are theories of rational competence which are based either on techniques for the syllogism or the predicate calculus. These are often forced to assume psychologically implausible procedures and do not account for errors in performance (Braine & Rumain, forthcoming; Newell, 1981). All the theories, however, have one striking feature in common. There are simple inferences that they cannot, in principle, accommodate. For example, given the premises:

> More than half the teachers are men.
> More than half the teachers sing in the choir

even children can draw the valid conclusion:

> There is a man who sings in the choir.

Yet this deduction cannot be expressed in the first-order predicate calculus, because there is no way to capture the logic of the quantifier, 'more than half the *x*', in terms of quantification over individuals.

What is required is quantification over sets—that is, a higher order quantificational calculus (see Barwise & Cooper, 1981). Unfortunately there is a crucial problem with the higher order calculus: it is incomplete in that there is no way to specify rules of inference for it that will capture the complete set of valid deductions. Doubtless, a subset of the calculus can be formalized so as to allow the simple deduction above to be derivable from rules of inference. The fact that this problem arises with such a trivial deduction, however, should surely give pause to any advocate of mental logic. One way round it, together with a solution to marrying rational competence with errors in performance, is presented in the next section.

Reasoning as a Semantic Procedure

There is a general principle that governs all valid deductions: An inference is valid if and only if there is no way of interpreting the premises in which the conclusion is not true. This principle characterizes the semantics of validity, and all systems of logic are designed to capture it within their formal rules of inference. But not all systems, as we have just seen, can succeed completely in this task. In proposing an algorithm for propositional reasoning, I advocated a technique based on the semantics of connectives rather than on rules of inference for them, and elsewhere I have sketched a theory of syllogistic inference based on the semantics of a syllogism's premises (Johnson-Laird, 1980). I now intend to generalize the idea and to propose a theory of competence in which all deductions are made on the basis of the semantic principle of validity. What this assumption means in practice is that any deduction can be made using the following general procedure:

1. Construct a mental representation based on the meaning of the premises—that is, a model of the state of affairs that they describe.
2. Formulate, if possible, an informative conclusion that is true in all models of the premises that have so far been constructed. If none, then there is no relevant conclusion.

3. Try to construct an alternative model of the premises that renders the conclusion false. If there is such a model, abandon the conclusion and return to step 2. If there is no such model, then the conclusion is valid.

Where a conclusion has to be evaluated rather than drawn spontaneously, all that is required is a simplified version of step 3: If an alternative model can be constructed that renders the given conclusion false, then it is invalid; otherwise, it is valid.

The concept of a mental model is a subtle one that is of relevance to comprehension in general (see Johnson-Laird, 1983). Here I shall illustrate it only with respect to inferences with quantifiers. The algorithm, which I will describe informally, has in fact been implemented in the programming language LISP-80 (a dialect of LISP).

The way in which an individual can represent the state of affairs described by the premise:

Some authors are book-keepers.

is by imagining an arbitrary number of authors and then mentally tagging them in some way to indicate that they are book-keepers. Such a model may take the form of a vivid image or, alternatively, it may be outside conscious access. Regardless of its phenomenal content, however, it has a structure that can be mapped one-to-one onto an actual state of affairs in which, indeed, some authors are book-keepers:

```
 author  =  book-keeper
 author  =  book-keeper
 author  =  book-keeper
(author)   (book-keeper)
```

This model represents five individuals: three authors who are book-keepers, a possible author (as indicated by the parentheses) who is not a book-keeper, and a possible book-keeper who is not an author. Since the model is no more than a representative sample from the set of possible models of the premises, the system that constructs and interprets it must in this instance appreciate that the numbers are arbitrary, that is, the procedures for constructing the model can recursively revise the numbers of entities that it contains.

The information conveyed by a second premise:

All the book-keepers are cardplayers.

can be added to the model of the first premise. It is merely necessary to locate each individual who is a book-keeper and to add a further tag indicating that the same person is also a cardplayer:

author =	book-keeper =	cardplayer
author =	book-keeper =	cardplayer
author =	book-keeper =	cardplayer
(author)	(book-keeper =	cardplayer)
		(cardplayer)

The final parenthesized token represents the possibility that there may be cardplayers who are not book-keepers. At this point, step 1 in the deduction procedure has been completed: a model of the state of affairs described in the premises has been constructed. The second step calls for the formulation of an informative conclusion. The inferential heuristics described earlier in the chapter require that if the premises relate *A* to *B*, and *B* to *C*, then an informative conclusion should relate *A* to *C*. The maximally informative assertion about this relation is:

Some authors are cardplayers.

or its converse, Finally, step 3 requires us to search for a model of the premises that falsifies this conclusion. There is no need to make pointless manipulations of the number of individuals in the model, since validity in a syllogism does not depend on particular numbers. Since any mental model contains only a finite number of individuals, there will always be only a finite number of ways of re-arranging the elements in the model in ways consistent with the premises. In the present case, there is no falsifying model, since there is no way of destroying the links between the end terms that does not also violate the truth conditions of the premises. Thus the conclusion is valid.

In the case of premises of the form:

All *A* are *B*.
All *C* are *B*.

the initial model may consist of:

$$\begin{array}{c} a = b = c \\ a = b = c \\ (b) \\ (b) \end{array}$$

which yields the informative conclusion: "All *A* are *C*," or its converse. A search for a model of the premises that falsifies this conclusion will, if properly conducted, yield the model:

$$\begin{array}{lll} a = & b & \\ a = & b & = c \\ & b & = c \\ & (b) & \end{array}$$

On returning to the second step, it is possible to formulate a conclusion that is consistent with both models: "Some *A* are *C*," or its converse. A search for a falsifying model of this conclusion should yield:

$$\begin{array}{lll} a = b & & \\ a = b & & \\ & b = c & \\ & b = c & \end{array}$$

Step 2 now fails to yield any conclusion interrelating *A* and *C* which holds over all three models, because in the first model *A* and *C* are co-extensive, whereas in the third model they are wholly disjoint. There is therefore no informative conclusion to be drawn from these premises.

The algorithm readily generalizes beyond syllogisms to deductions that depend on multiple quantification or non-standard quantifiers. Hence, the premise: "More than half the teachers are men," yields a model of the form:

teacher = man
teacher = man
teacher = man
teacher (man)

and there is no way in which to incorporate the premise, "More than half the teachers sing in the choir," that does not identify at least one teacher as both a man and a singer in the choir.

It should be clear that the use of the general semantic procedure as a method of inference is very different from the use of a mental logic. The semantic procedure contains no formal rules of inference; it relies instead on the ability to construct models of premises, to formulate conclusions that are true of them, and to search for models that falsify those conclusions. The ability to construct models and to describe them, however, is presumably what underlies the production and comprehension of discourse. The only addition called for by deduction is a knowledge of the semantic principle of validity and the ability to put it into practice in the search for counter examples. Hence, the theory of reasoning without logic is certainly more parsimonious than the theory of mental logic, because it does not require any rules of inference. Is it, however, the correct theory of human inferential performance?

Evidence in Favour of the Semantic Procedure

It is difficult to choose between the formal and the semantic approaches to reasoning on the basis of empirical observations of propositional reasoning. The results from experiments with syllogisms are more clear-cut, and they plainly support the semantic theory. There are three crucial observations:

(1) The errors that occur in syllogistic reasoning are proportional to the number of models that have to be constructed in order to make the correct response (Johnson-Laird, 1982). This result is to be expected on the reasonable assumption that the greater the number of models to be constructed, the greater the load on working memory and the more likely that a model will be overlooked. The difference between the easy and the difficult syllogistic problems presented earlier can indeed be explained in this way. The easy problem:

Some of the *A* are *C*.
All of the *B* are *C*.

has only a single model of the form:

$$\begin{array}{ccccc} a & = & b & = & c \\ a & = & b & = & c \\ (a) & & (b) & & c \\ & & & & (c) \end{array}$$

which yields the conclusions:

Some of the *A* are *C*.

or its equally valid converse. The difficult problem:

All of the *B* are *A*.
None of the *C* are *B*.

yields three models:

$$\begin{array}{ccc} (1) & (2) & (3) \\ \begin{array}{ccc} c & & \\ c & & \\ \hline b & = & a \\ b & = & a \\ & & (a) \end{array} & \begin{array}{ccc} c & & \\ c & = & a \\ \hline b & = & a \\ b & = & a \\ & & (a) \end{array} & \begin{array}{ccc} c & = & a \\ c & = & a \\ \hline b & = & a \\ b & = & a \\ & & (a) \end{array} \end{array}$$

where the horizontal barriers are a notational convention for representing the premise that none of the *C* are *B*.

(2) The erroneous conclusions that occur fit the assumption that subjects overlook the possibility of certain models. Thus, in the previous example, subjects who construct only the first model will respond:

No *C* are *A*. (60% of responses)

Subjects who succeed in falsifying this conclusion by constructing the second model will respond:

Some *C* are not *A*. (10% of responses)

And subjects who succeed in falsifying this conclusion by constructing the third model are likely to respond:

No valid conclusion. (20% of responses)

Only if subjects succeed in constructing all three models and in evaluating them in the opposite direction to the figural bias will they appreciate that the conclusion:

Some of the *A* are not *C*. (0% of responses)

holds over all the models of the premises. The predictions are stated here together with the percentages of spontaneous conclusions cited earlier (from Johnson-Laird & Steedman, 1978). These data are wholly typical in that the theory predicts the vast majority of responses that are made, and there are no responses made in significant numbers that are not predicted by the theory.

(3) The main effect of the figure of the premises is to create a bias towards certain conclusions, which is contrary to Revlis's (1975) hypothesis that subjects invariably make illicit conversions of the premises. The bias can be explained in terms of the operations required to construct a model of the two premises. It is necessary to form a model of one premise and then to incorporate within it the information from the second premise. The hinge on which the process depends is, of course, the term that occurs in both premises—the so-called middle term. William James (1890) pointed out that it was easy to integrate two relational assertions of the form *aRb* and *bRc* because the two middle terms occurred continguously. Hunter (1957) expands on this notion by introducing two mental operations that would be required to bring the middle terms into contiguity with premises in other figures. This idea can be extended to the integration of mental models. With premises in the figure, *A–B*, *B–C*, the process of integration can occur immediately. Working memory evidently operates on the principle that the first information into memory is the first information out of it (see Broadbent, 1958. p. 236). Hence, subjects will in this case be biased towards conclusions of the form *A–C*. With premises of the form *B–A*, *C–B*, an immediate integration is impossible, but if the initial model is based on the second premise, *C–B*, the information from the first premise, *B–A*, can then be integrated. The 'first in, first out' principle dictates that subjects will now be biased towards conclusions of the form *C–A*. With premises of the form *A–B*, *C–B*, the initial model can be based on either premise, but it is necessary to carry out a major operation—switching round the

interpretation of the other premise in order to bring the middle term into contiguity with its representation in the model. With premises in the figure *B–A*, *B–C*, the initial model cannot be fruitfully based on either premise. It is first necessary to switch round the interpretation of a premise and then to recall the other premise. Hence, the complexity of the mental operations required to form an integrated model increases over the four figures, and this increase is mirrored both in the latencies to one-model problems (for which there are sufficient correct responses to assess the reliability of the trend) and in the decline in correct valid conclusions to all the premises permitting them (see Johnson-Laird & Bara, 1984).

The semantic theory does not rule out the possibility that subjects may misinterpret or illicitly convert premises, forget them, or bring to mind irrelevant considerations. Nevertheless, the major causes of error in syllogistic reasoning appear to be, first, the need to construct alternative models of the premises, and, second, the effects of figure. The theory therefore appears to be superior to any current explanation based on the doctrine of mental logic.

RELATIONAL INFERENCES

Many deductions in daily life depend on simple relations. The following deduction, for example, depends on grasping a variety of temporal and geographical relations:

> John will stand in the local elections in Camden.
> Camden is a borough of London.
> London has its annual borough elections on Tuesday.
> Therefore, John stands in the elections on Tuesday.

A proponent of mental logic would argue that such deductions are drawn on the basis of formal rules of inference. Such rules, however, have yet to be formalized within any existing logic. Thus, for instance, although there are so-called 'tense logics', they are relatively remote from the logical properties of English tenses and modal auxiliaries. Nonetheless, it is clear how, in principle, a theory based on mental logic would account for the ability to make relational inferences. It would do so by analogy with the explanations of performance in three-term series problems, which are the type of relational inference that have received the most attention in the psychological literature.

Here is a typical three-term series problem:

> Anne is taller than Bertha.
> Bertha is taller than Charles.
> Therefore, Anne is taller than Charles.

The main puzzle that psychologists have sought to resolve is the nature of the mental representation of the premises and the mental processes by which the conclusion is derived. There has been considerable controversy over whether those representations and processes are primarily verbal or visual (see, e.g., Clark, 1969; Huttenlocher, 1968). Many experimental results suggest that different subjects may employ different strategies and that they can be induced to change strategies (see, e.g., Egan & Grimes-Farrow, 1982; Mayer, 1979; Ormrod, 1979; Sternberg & Weil, 1980). A crucial aspect of such inferences, regardless of the nature of their mental representation, is how reasoners grasp their logical validity. Advocates of mental logic are forced to argue either that there is a general schema of trasitivity:

For any x, y, and z, if xRy and yRz, then xRz.

to which particular relations such as 'taller than' are linked (Johnson-Laird, 1975), or else that specific schemata or 'meaning postulates' are acquired for each transitive relation (Kintsch, 1974):

For any x, y, and z, if x is taller than y, and y is taller than z, then x is taller than z.

The semantic approach suggests a different way in which such deductions are made. Reasoners build a model of the state of affairs described in the premises, generate an informative conclusion on the basis of that model, and then search for an alternative model of the premises in which the conclusion is false. The ability to construct models plainly depends on a grasp of the truth conditions of the premises—that is, a knowledge of the states of affairs in which the premises are true in principle. This knowledge, in turn, depends on grasping the contribution to these truth conditions of the meanings of relational expressions such as 'taller than'. Earlier in the chapter, a distinction was drawn between the truth conditions of propositional connectives and the rules of inference governing them. The same contrast can be drawn between the meanings of relational expressions and the schemata characterizing their logical properties. Since such relational expressions as 'taller than', 'greater than', 'kinder than', and their cognates are all transitive, asymmetric, and irreflexive, it follows that the differences in meaning between them cannot be distinguished merely by specifying their logical properties. Hence, a proper grasp of their meaning must depend on a knowledge of their truth conditions.

Although there is no need for inferential schemata once the truth conditions of an expression have been acquired, it might be argued that such schemata are useful and play a role in the psychology of inference. It is impossible to rule out this hypothesis in general, but there are domains where one can be certain that it is false. One such domain concerns spatial relations.

The spatial relation, 'on the right of', has two similar meanings (Miller & Johnson-Laird, 1976, Sec. 6.1.3). One meaning refers to the relation between objects from a particular point of view (usually the speaker's), as in the assertion:

The boulder is on the right of the road.

The other meaning depends on the fact that people (and many other entities) are conceived of as having an intrinsic right- and left-hand side:

St. John is on the right of Jesus in *The Last Supper*.

The first meaning plainly gives rise to transitivity. However, the second meaning has a very interesting property: In some cases, it is transitive; in others, it is intransitive; and in still others, it has a limited transitivity over a finite number of individuals. An example of the last sort arises with people seated round a circular table. It may then be valid to argue:

Arthur is on Guinevere's right.
Guinevere is on Lancelot's right.
Therefore, Arthur is on Lancelot's right.

But the seating arrangement may render transitivity over four or more individuals out of the question (for example, the fourth individual may be sitting directly opposite the first).

It is very difficult to define the logical properties of such relations using inference schemata because it is necessary to specify an infinite number of schemata ranging over all possible degrees of transitivity. A sensible way to proceed would be to postulate a meta-schema that allows the required degree of transitivity to be determined from information about the circumstances referred to in the premises. To concede this point, however, is to concede too much. If information about the situation were to be used in this way, then it would be better to abandon altogether the inferential schemata that mediate between one expression and another, and to rely instead on meanings that directly relate language to mental models of situations. Indeed, it is a simple matter to specify such a semantics for " on the right of" by defining the appropriate direction in a Cartesian framework: Increment the value of one horizontal co-ordinate whilst holding the other one constant. If the horizontal co-ordinates are those of a spatial layout as seen from a particular point of view, then these truth conditions yield transitivity as an emergent property of the relation. If the horizontal co-ordinates are centred on a particular individual in order to make an intrinsic interpretation, the same truth conditions yield transitivity over a collinear arrangement. Otherwise, as with a circular arrangement, transitivity may break down. In particular, *A* may be on *B*'s right and *B* on *C*'s right, but *A* may not be on *C*'s right because the line specifying those points on the right of *C* goes off at a tangent from the circle and clearly misses *A*, who is, say, nearly opposite to *C*.

COMMON-SENSE REASONING

Perhaps the most important purpose served by inference is to produce new knowledge. Suppose, for example, that I want to get my car, a Renault, serviced, but that I do not know a good garage. Someone may tell me that some of the people in the linguistics department own Renaults. I am then likely to try to contact these individuals. My behaviour is motivated by a simple deduction in which one premise is the assertion by my informant, and the other is a piece of general knowledge:

> Anyone who owns a Renault is likely to know where Renaults can be serviced.

The conclusion provides me with new information:

> Some of the people in the linguistics department are likely to know where Renaults can be serviced.

This use of deduction places constraints, which we have already encountered, on practical inference: conclusions should be informative—that is, they should establish relations between terms that are not explicitly related in the premises, and these relations should contain at least as much semantic information as is conveyed by the premises.

Another major function of inference is to serve the process of comprehension. Because people can make inferences, it is often unnecessary for speakers to spell everything out in detail; they can leave certain information unstated and rely on their listeners to infer it. Very often, however, this missing information can be filled in only by default. It can be assumed only because there is nothing to the contrary. The conclusion is not valid, since it may be overruled by subsequent information. Consider the following discourse:

> When I returned to my house last night, I discovered that I had lost my keys. There was no one there and the door was locked. I broke the glass and turned the lock from the inside. Someone heard the noise and came running.

In order to understand this passage, the reader will have drawn a series of inferences by default:

> The keys the speaker lost included the key to the door of his house.
> There was no one *in his house.*
> The door that was locked was the door to his house.
> He broke a pane of glass in the door to his house.
> He reached in through the resulting hole and unlocked the door.
> Someone (not in the house) heard the noise of the breaking glass and came running to investigate its cause.

Each of these inferences helps to tie the discourse together and enables the reader to construct a mental model that integrates the information from the separate sentences. None of these inferences, however, is valid; each of them could have been contraverted by subsequent information. For example, the passage might have made it clear that the speaker lost his car keys and had to break into his car by smashing a window.

Workers in artificial intelligence have suggested that certain default inferences depend on 'scripts'—that is, representations of the typical sequences of events that occur in certain stereotyped situations, such as dining in a restaurant (see, e.g., Schank & Abelson, 1977). Scripts, however, do not embrace all default inferences, as shown by some of the examples above, and they lack any explicit machinery for revising conclusions in the light of subsequent information. Attempts have accordingly been made to devise so-called 'non-monotonic' logics. Orthodox systems of logic are monotonic in that when a conclusion follows validly from a set of premises, it still follows validly if any additional premise is added to the set. However, if we reason by default, we draw a conclusion that a subsequent assertion may force us to abandon. Non-monotonic logics are designed to formalize this notion (see, e.g., McDermott & Doyle, 1980). Unfortunately, formal rules of inference run into apparently insurmountable problems with the underlying semantics of reasoning by default (Davis, 1980).

The whole point of the semantic approach is to marry reasoning with comprehension. Comprehension depends on the construction of mental models and is in fact akin to the process of maintaining a consistent database. It is necessary to make inferences from input information, and to be able to revise the database if such conclusions conflict with subsequent information.

Any particular assertion is interpreted by constructing an appropriate mental model or an appropriate addition to an existing model. Of course, an assertion may conflict with the current model of the discourse, perhaps because a default inference has been made earlier. Clearly, in the case of a conflict, an attempt should be made to construct an alternative model that is consistent with the discourse as a whole. Where there is such a model, the assertion conveys new semantic information; where there is no such model, the assertion is genuinely inconsistent with the previous discourse. This process is complementary to the semantic approach to deduction in which an attempt is made to construct an alternative model of the discourse that renders an assertion (the putative conclusion) false. Where there is such a model, the assertion does not follow validly from the previous discourse; where there is no such model, the assertion is a valid deduction from the previous discourse. Hence, both deduction and the revision of a default inference depend on recursive processes that search for alternative models.

There is one further hiatus between formal logic and the inferences of daily life. Suppose that the following evidence is presented at a murder trial:

> The victim was stabbed to death in a cinema.
> The suspect was travelling on a train to London when the murder took place.

Even children appreciate that this evidence provides a good alibi for the suspect. This conclusion is not valid, however, and so it cannot be drawn simply on the basis of a mental logic.

In some informal tests, I asked small groups of subjects what follows from these premises. They concluded that the suspect was indeed innocent, and when I inquired why, they pointed out a number of unstated assumptions:

> One person cannot be in two places at the same time.
> There are no cinemas on trains to London.
> Trains don't pass through cinemas.

When I queried whether the subjects were absolutely certain about their conclusion, an interesting phenomenon occurred: they began to try to construct scenarios in which the suspect *was* guilty. They argued that perhaps he left the cinema and boarded the train, that he might have thrown the knife, and so on. When I ruled out these hypotheses, the subjects suggested further possibilities, such as the use of a confederate, a spring-loaded knife in the seat, or a radio-controlled robot.

The only way to guarantee the validity of a conclusion is, of course, to eliminate all possible counterexamples. Logic is a formal machine designed to achieve this goal. In this case, however, logic cannot ensure that one has considered all the different ways in which the murder might have been accomplished. Like most everyday problems that call for reasoning, the explicit premises leave most of the relevant information unstated. Indeed, the real business of reasoning in these cases is to determine the relevant factors and possibilities, and it therefore depends on a knowledge of the specific domain. Hence, the construction of putative counterexamples calls for an active exercise of memory and imagination rather than a formal derivation of one expression from others. Yet the process is deductive: reasoners are trying to find an interpretation of the premises from which there is an informative conclusion that must be true.

Even if there were a complete logic for all utterances in natural language, it would still be necessary to spell out explicitly all the unstated assumptions that people rely on in making such inferences. Reasoning itself is largely concerned with the exploration of the tacit assumptions that flesh out the explicit premises. Reasoners draw a tentative conclusion and then seek to become aware of what they have assumed in order to ascertain whether it is necessarily true. Psychologists would have appreciated this aspect of practical reasoning long ago if they had not examined deductive performance through the distorting lens of formal logic.

CONCLUSIONS

Reasoning without logic consists of three simple steps: Interpret the premises by constructing a model based on their truth conditions, formulate an informative conclusion, and check the conclusion by searching for different models of the premises. Nevertheless, such reasoning is also very powerful. Any of the deductions that people are ordinarily able to make fall within its scope, and it is not necessary to devise different theories of reasoning for propositional connectives, quantifiers, and relational terms. Similarly, it is not necessary to formulate special theories for sentential connectives such as 'because' and 'if', which are used in senses that are not truth-functional. Once the truth conditions of a sentence have been grasped, they can be used as a basis for reasoning.

In the past, psychologists have assumed the doctrine of mental logic without argument. What they have argued about is different versions of the doctrine—no one ever collected data to investigate whether the doctrine itself was correct, and even to this day there are theorists who regard it as irrefutable (Cohen, 1981). In fact, however, we can now see that logic is neither necessary nor sufficient to account for rational competence, and that the experimental evidence favours a theory based on truth conditions. The semantic theory offers a superior account of syllogistic inference, relational deductions, and the common-sense and default inferences of daily life. Theories of performance based on mental logic should therefore be abandoned.

This step has three important consequences: First, it follows that current conceptions of children's intellectual development are wrong; children do not learn, construct, or inherit a mental logic. On the contrary, they acquire the truth conditions of words, and they use them to construct models of premises and to search for alternative models in accordance with the semantic criterion of validity. Second, the concept of 'logical form'—the form of a sentence as captured by a system of logic—has no role to play in accounting for deductive competence. In other words, there are no terms, such as the connectives 'and' and 'or' or the quantifiers 'all' and 'some', that have to be treated specially by the theory of reasoning. People need only to know the truth conditions of these terms in order to make deductions, just as they need only to know them for other so-called non-logical terms. Psychologists and linguists should give up the notion of logical form along with that of mental logic. Third, formal logic as an intellectual discipline does not depend on externalizing an inner system of logic. Rather, it is an attempt to formalize systematic rules that capture syntactically a class of deductively valid inferences that are initially grasped intuitively, incompletely, and unsystematically through the pre-theoretical manipulation of mental models.

There is one class of deductions that the semantic theory, in common with all other psychological analyses, cannot explain: inferences concerning infinite sets. There is obviously no way in which an infinite number of entities can be directly represented in the mind. Hence, it is plausible to assume that these inferences are made possible by the invention of formal logic and mathematics. A primary motive for this invention is indeed our limited ability to reason. Too often, we make invalid inferences, not because we have forgotten the premises or distorted their meaning, but because our working memories are too overloaded for us to consider all the possible models of the premises. Logic is a consequence of our limited ability to reason, not the fundamental machinery on which this ability is based.

ACKNOWLEDGEMENTS

The work reported in this chapter was supported in part by a grant from the S.S.R.C. (G.B.) to Jane Oakhill and the author. I am particularly grateful both to Jane Oakhill and to Bruno Bara for help, ideas, and criticism. I am also grateful to Martin Braine, the doyen of the 'mental logicians', for stimulating discussions on the theory of inference.

REFERENCES

Bar-Hillel, Y., & Carnap, R. An outline of the theory of semantic information. In Y. Bar-Hillel, *Language and information.* Reading, MA: Addison-Wesley, 1964. Reprinted from Technical Report No. 247, MIT, Research Laboratory of Electronics, 1952.

Barwise, J., & Cooper R. Generalized quantifiers and natural languages. *Linguistics and Philosophy*, 1981, **4**, 159–219.

Beth, E. W., & Piaget, J. *Mathematical epistemology and psychology.* Dordrecht: Reidel, 1966.

Boole, G. *An investigation of the laws of thought.* London: Walton and Maberly, 1854.

Boolos, G. S., & Jeffrey, R. C. *Computability and logic* (2nd ed.). Cambridge: Cambridge University Press, 1980.

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

Braine, M. D. S., & Rumain, B. Development of comprehension of "or": Evidence for a sequence of competencies, *Journal of Experimental Child Psychology*, 1981, **31**, 46–70.

Braine, M. D. S., & Rumain, B. Logical reasoning. In J. Favell & E. Markman (Eds.), *Carmichael's Manual of Child Psychology* (4th ed.). New York: Wiley, forthcoming.

Broadbent, D. E. *Perception and communication.* London: Pergamon, 1958.

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

Chomsky, N. *Aspects of the theory of syntax.* Cambridge, MA: MIT Press, 1965.

Chomsky, N. *Essays on form and interpretation.* Amsterdam: North-Holland, 1977.

Clark, H. H. Linguistic processes in deductive reasoning. *Psychological Review, 1969,* **76**, 387–404.

Cohen, L. J. Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences,* 1981, **4**, 317–370.

Davis, M. The mathematics of non-monotonic reasoning. *Artificial Intelligence,* 1980, **13**, 73–80.

Egan, D. E., & Grimes-Farrow, D. D. Differences in mental representations spontaneously adopted for reasoning. *Memory and Cognition,* 1982, **10**, 297–307.

Erickson, J. R. A set analysis of behavior in formal syllogistic tasks. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium.* Potomac, MD: Erlbaum, 1974.

Falmagne, R. J. The development of logical competence: A psycholinguistic perspective. In R. H. Kluwe & H. Spada (Eds.), *Developmental models of thinking.* New York: Academic Press, 1980.

Fodor, J. A. Fixation of belief and concept acquisition. In M. Piatelli-Palmarini (Ed.), *Language and learning: The debate between Jean Piaget and Noam Chomsky.* Cambridge, MA: Harvard University Press, 1980.

Gazdar, G. A cross-categorial semantics for coordination. *Linguistics and Philosophy, 1980,* **3**, 407–409.

Guyote, M. J., & Sternberg, R. J. A transitive-chain theory of syllogistic reasoning. *Cognitive Psychology,* 1981, **13**, 461–525.

Henle, M. On the relation between logic and thinking. *Psychological Review,* 1962, **69**, 366–378.

Henle, M. Foreword. In R. Revlin & R. E. Mayer (Eds.), *Human reasoning.* Washington, DC: Winston, 1978.

Hunter, I. M. L. The solving of three-term series problems. *British Journal of Psychology,* 1957, **48**, 286–298.

Huttenlocher, J. Constructing spatial images: A strategy in reasoning. *Psychological Review,* 1968, **75**, 550–560.

Inhelder, B., & Piaget, J. *The growth of logical thinking from childhood to adolescence.* London: Routledge & Kegan Paul, 1958.

James, W. *The principles of psychology* (Vol. 2). New York: Holt, 1890.

Johansson, B. S., & Sjolin, B. Preschool children's understanding of the coordinators "and" and "or". *Journal of Experimental Child Psychology,* 1975, **19**, 233–240.

Johnson-Laird, P. N. Models of deduction. In R. J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults.* Hillsdale, NJ: Erlbaum, 1975.

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

Johnson-Laird, P. N. Ninth Bartlett memorial lecture: Thinking as a skill. *Quarterly Journal of Experimental Psychology,* 1982, **34A**, 1–29.

Johnson-Laird, P. N. *Mental models.* Cambridge, MA: Harvard University Press, 1983.

Johnson-Laird, P. N., & Bara, B. Syllogistic inference. *Cognition,* 1984, **16**, 1–61.

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

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

Kintsch, W. *The representation of meaning in memory.* Hillsdale, NJ: Erlbaum, 1974.

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

Mayer, R. E. Qualitatively different encoding strategies for linear reasoning problems: Evidence for single association and distance theories. *Journal of Experimental Psychology: Human Learning and Memory,* 1979, **5**, 1–10.

McDermott, D., & Doyle, J. Non-monotonic logic, I. *Artificial Intelligence,* 1980, **13**, 41–72.

Mill, J. S. *A system of logic* (8th ed.). New York: Harper, 1874.

Miller, G. A., & Johnson-Laird, P. N. *Language and perception.* Cambridge, MA: Harvard University Press, 1976.

Newell, A. Reasoning, problem solving, and decision processes: The problem space as a fundamental category. In R. Nickerson (Ed.), *Attention and Performance* (Vol. VIII). Hillsdale, NJ: Erlbaum, 1981.

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

Ormrod, J. E. Cognitive processes in the solution of three-term series problems. *American Jounral of Psychology,* 1979, **92**, 235–255.

Osherson, D. N. Logic and models of logical thinking. In R. J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults.* Hillsdale, NJ: Erlbaum, 1975.

Revlis, R. Two models of syllogistic reasoning: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behavior,* 1975, **14**, 180–195.

Rips, L. J. Cognitive processes in propositional reasoning. *Psychological Review,* 1983, **90**, 38–71.

Schank, R. C., & Abelson, R. P. *Scripts, plans, goals and understanding.* Hillsdale, NJ: Erlbaum, 1977.

Shannon, C., & Weaver, W. *The mathematical theory of communication.* Urbana: University of Illinois Press, 1949.

Staudenmayer, H. Understanding conditional reasoning with meaningful propositions. In R. J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults.* Hillsdale, NJ: Erlbaum, 1975.

Sternberg, R. J., & Weil, E. M. An aptitude-strategy interaction in linear syllogistic reasoning. *Journal of Educational Psychology,* 1980, **72**, 226–239.

Störring, G. Experimentelle Untersuchungen über einfache Schlussprozesse. *Archiv für die gasamte Psychologie,* 1908, **11**, 1–127.

Wason, P. C., & Johnson-Laird, P. N. *Psychology of reasoning: Structure and content.* Cambridge, MA: 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**, No. 102.

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

Youniss, J. Inference as a developmental construction. In R. Falmagne (Ed.), *Reasoning: Representation and process in children and adults.* Hillsdale, NJ: Erlbaum, 1975.

Reasoning and Discourse Processes

edited by

TERRY MYERS

Centre for Cognitive Science
University of Edinburgh
Edinburgh, Scotland

KEITH BROWN

Department of Language and Linguistics
University of Essex
Colchester, England

and

BRENDAN McGONIGLE

Department of Psychology
University of Edinburgh
Edinburgh, Scotland

1986

ACADEMIC PRESS
Harcourt Brace Jovanovich, Publishers
London Orlando New York San Diego Austin
Boston Tokyo Sydney Toronto

ACADEMIC PRESS INC. (LONDON) LTD
24/28 Oval Road
London NW1

United States Edition published by
ACADEMIC PRESS INC.
Orlando, Florida 32887

British Library Cataloguing in Publication Data
Reasoning and discourse processes.—
(Cognitive science series)
1. Discourse analysis
I. Myers, Terry II. Brown, K.
III. McGonigle, B. IV. Series
415 P302

Library of Congress Cataloging-in-Publication Data
Main entry under title:

Reasoning and discourse processes.

Includes index.
1. Reasoning (Psychology) 2. Rhetoric.
I. Myers, Terry. II. Brown, K. (Keith),
III. McGonigle, B., Date-
BF441.R38 1985 153.4'3 85-9223
ISBN 0-12-512320-5
ISBN 0-12-512321-3 (Pbk)

Typeset by Composition House Limited,
Salisbury, Wiltshire.

Printed in Great Britain by St. Edmundsbury Press,
Bury St. Edmunds, Suffolk.