CHAPTER SIX

Inference and Mental Models

Philip N. Johnson-Laird

Department of Psychology, Princeton University, USA

INTRODUCTION

Suppose that the facts are:

> The defendant left town before the restaurant closed. The arsonist put a match to the cooking oil in the kitchen after the restaurant closed.

What conclusion would you draw? Probably, you would infer that the defendant could not have started the fire because, barring some remote scenario, the culprit must have been there to start the fire. The question for psychologists is: what mental process enabled you to reach your conclusion? In principle, it could be syntactic or semantic. Most psychologists who have studied the topic favour syntactic theories, but over the past decade a growing number have entertained the possibility of a semantic theory. My plan in this chapter is to describe one particular semantic theory—the theory of mental models—and to illustrate its power by applying it to three areas of reasoning: first, a hitherto uninvestigated area (temporal reasoning), secondly an established area (reasoning with sentential connectives), and thirdly the most studied area of all (Peter Wason's selection task).

Before we can compare syntax and semantics, we need to be clear about the difference between them. The clearest distinction has been drawn by 20th century logicians. Syntax concerns the form of expressions. Thus, a syntactic process operates on the form of expressions represented in some medium, such as marks on paper. Such a process occurs in constructing a

well-formed expression according to the rules of a grammar. It also occurs in deriving a conclusion from a set of premises according to formal rules of inference. The power of syntactic processes is that they apply to any content provided that it is expressed in the appropriate form. Semantics, in contrast, concerns the relations between expressions in a language and some domain outside the language—e.g. the relation between names in the language and the entities in the domain that they name. In a natural language, it is possible to use names, predicates, and other parts of speech to construct expressions that correspond to states of affairs in the world. The process is probably compositional with each syntactic rule in the grammar having an associated semantic rule that specifies how to combine the meanings of constituents according to the syntactic relations captured by the grammatical rule. In this way, the semantics of the language specifies the conditions in which a sentence would be true, and the actual state of affairs in the domain renders the sentence true or false depending on whether it satisfies these so-called "truth conditions". Of course, this picture is simplistic. Sentences in natural language do not have truth conditions, but rather their use in context has them. Similarly, certain aspects of idiomatic meaning are not compositional, and many speech acts do not seem to have truth conditions. We can be sure, however, that some assertions have truth conditions—otherwise, we should never be able to judge truth or falsity.

Syntax concerns form; semantics concerns truth. The burden of certain theorems in meta-logic is that an unbridgeable gulf exists between them (*pace* those theorists who seek to reduce meaning to syntax). As Gödel showed, there are truths in arithmetic that cannot be proved by any consistent syntactic method of inference. Now we can pose our central question—is the human inferential system syntactic or semantic? Does it manipulate expressions in a mental language according to purely syntactic principles akin to those of formal proofs, or does it proceed in a semantic way? And if the latter, what could the method be? Most theories of deductive reasoning have postulated a syntactic system of rules of inference (e.g. Braine & O'Brien, 1991; Macnamara, 1986; Osherson, 1974–76; Rips, 1983). Yet, for reasons that will become clear, the present paper defends the semantic theory developed originally by Johnson-Laird (1983) and later by Johnson-Laird and Byrne (1991).

THE THEORY OF MENTAL MODELS

A valid deduction, by definition, is one in which the conclusion must be true if the premises are true. What we need is a semantic method to test for this condition. Unfortunately, assertions can be true in indefinitely many different situations, and so it is out of the question to test that a conclusion

holds true in all of them. If, *per impossible*, it could be done, then there would be a general semantic method for evaluating deductions.

Here is how it could be done for certain everyday inferences. Consider, again, the opening example about the temporal order of events, which can be abbreviated as follows:

a before b
c after b

where "a" stands for "the defendant left town", "b" stands for "the restaurant closed", and "c" stands for "the arsonist put a match to the cooking oil in the kitchen". The assertions say nothing about the actual durations of the events or about the intervals between them. Instead of trying to build models of all the different possible situations that satisfy these premises, let us build a "partial" model that leaves open the details and that captures only the structure that all the different situations have in common:

a   b   c

where the left-to-right axis corresponds to time, but the distances between the tokens have no significance. This model represents only the sequence of events, and it is the only possible model of the premises—i.e. no other model corresponding to a different sequence of the three events satisfies the premises.

Now consider the further assertion:

a before c.

It is true in the model, and, because there are no other models of the premises, it must be true given that the premises are true. The deduction is valid, and because reasoners can determine that there are no other possible models of the premises, they can not only make this deduction but also know that it is valid (see Barwise, 1993). The same principles allow us to determine that an inference is invalid. Given, say, the following inference:

a happens before b
c happens before b
∴ a happens before c

the first premise yields the model:

a   b

but now when we try to add the information from the second premise, the relation between a and c is uncertain. One way to respond to such an indeterminacy is to build separate models for each possibility:

```
a       c     b             c     a     b             a     b
                                                      c
```

where the third model represents a and c as contemporaneous. The first of these models shows that the putative conclusion is possible, but the second and third models are counter-examples to it. It follows that event a may happen before event c, but it does not follow that event a must happen before event c.

One disadvantage of this procedure is that as the number of indeterminacies in premises increases so there is an exponential growth in the number of possible models. The procedure is intractable for all but small numbers of indeterminacies. Yet, even though the human inferential system is bounded in its powers (Simon, 1959), it may use an intractable procedure. Consider the following premises, which go beyond the simple transitive inference illustrated above:

  a happens before b
  b happens before c
  d happens while a
  e happens while c
  What's the relation between d and e?

The premises call for the construction of a single model:

```
a       b     c
d             e
```

which supports the conclusion:

  d happens before e.

The model theory predicts that this one-model problem should be easier than a similar inference that contains an indeterminacy. For example, the following premises call for several models:

  a happens before c
  b happens before c
  d happens while b
  e happens while c
  What's the relation between d and e?

The premises are satisfied by the following models:

```
a     b    c        b    a    c        a    c
      d    e        d         e        b
                                       d    e
```

In all three models, d happens before e, and so it is a valid conclusion. The model theory also predicts that the time subjects spend reading the second premise, which creates the indeterminacy leading to multiple models, should be longer than the reading time of the second premise of the one-model problem. Recently, Walter Schaeken and the present author have corroborated both these predictions: subjects took reliably longer (about 2 sec) to read an indeterminate premise, and made reliably more errors (about 10%) with such problems (see Schaeken, Johnson-Laird, & d'Ydewalle, 1993; and for similar results with spatial inferences, Byrne & Johnson-Laird, 1989).

Systems based on formal rules of inference make exactly the wrong predictions about these domains. The one-model problem above calls for a transitive inference to establish the relation between a and c, which is a precursor to establishing the relation between d and e. In contrast, the multiple-model problem does not need to derive the relation between b and c, because it is directly asserted by the second premise. Hence, the one-model problem has a longer formal derivation than the multiple-model problem, and so according to rule theories the one-model problem should be harder than the multiple-model problem. The irrelevant premise in the multiple-model problem cannot be responsible for its greater difficulty, because one-model problems with an irrelevant premise remain reliably easier than multiple-model problems, and no harder than one-model problems without irrelevant premises. In short, multiple models do cause problems for the human inferential system.

MODELS AND SENTENTIAL REASONING

If a speaker tells you:

Both the alarm light came on and the bell sounded.

you can imagine the two events in a single model:

a   b

where "a" denotes a representation of the alarm light coming on and "b" denotes a representation of the bell sounding, and the two are combined in a single model in which the left–right axis has no temporal significance. Suppose that instead of a conjunction, the speaker asserts:

The alarm light came on or the bell sounded.

how are you likely to represent this assertion? An obvious answer is that you construct two models, one for each possibility:

a

    b

The further assertion:

In fact, the alarm light did not come on.

rules out the first model to leave only the second, from which it follows:

The bell sounded.

There are three points to note about this inference. First, it was made without relying on a syntactic rule of inference, such as:

p or q, or both
not q
$\therefore$ p

Secondly, it was made without any decision about whether the disjunction was inclusive (the alarm light came on or the bell sounded, or both) or exclusive (the alarm light came on or the bell sounded, but not both). The advantage of moving straight to the deduction is that it saves time, and here it does not matter which interpretation is correct. Indeed, the speaker may not have had a particular interpretation in mind. Thirdly, the models are partial in that the first model does not represent explicitly that "not b" holds in this situation, and the second model does not represent explicitly that "not a" holds in this situation.

Now let us consider how this approach might be extended to deal with the most puzzling of sentential connectives, "if ". Imagine that the speaker asserts:

If the alarm light came on then the bell sounded.

Despite many pages of philosophical and linguistic analysis, there is no consensus about the meaning of conditionals. Yet children appear to master them without too much difficulty. So what does this conditional mean, and how is it likely to be mentally represented? Granted that the conditional

is true, it allows that one possible state of affairs is that the alarm light came on and so the bell sounded too. This possibility calls for a model akin to the one above for a conjunction:

a   b

But, clearly, the conditional differs in meaning from the conjunction. The "if" clause refers to just one possibility, and another possibility is that the alarm light did not come on. Hence, the representation must allow for this other possibility. But, if the alarm light did not come on, what then? Did the bell sound, or not? Before you get too embroiled in this question, bear in mind that individuals may not have to answer it in order to understand the rest of the discourse. They need not make the possibilities explicit if they do not matter. In short, the following models of the conditional can serve as an initial representation of its meaning:

a   b
   . . .

where the ellipsis represents the other possibilities. Of course, individuals ought to make a mental note that one event which cannot occur as an alternative possibility is that the alarm light came on, because this possibility is already represented in the explicit model. How this mental note is actually represented is by no means certain.

There is a well-known indeterminacy in the meaning of conditionals. An assertion such as:

If you eat your semolina, then you can have a chocolate.

naturally suggests that if you don't eat your semolina then you won't get a chocolate. This interpretation is equivalent to a bi-conditional:

If you eat your semolina, then you can have a chocolate, and if you don't eat your semolina, then you cannot have a chocolate.

Or, more succinctly:

If, and only if, you eat your semolina, then you can have a chocolate.

In contrast, the following conditional:

If you eat your semolina, then you will stay healthy.

has an interpretation that is not so strong. You may stay healthy even if you abstain from semolina. This interpretation is equivalent to a one-way conditional:

If you eat your semolina, then you will stay healthy, and if you don't eat it you may or may not stay healthy.

In daily life, people often seem to understand conditionals without worrying about whether they are one-way or bi-conditionals — just as they interpret disjunctions without worrying about whether they are inclusive or exclusive. In the laboratory, experimental subjects similarly vacillate in their interpretations of conditionals and disjunctions, though a specific content or context can bias their interpretations (see e.g. Legrenzi, 1970).

The initial models of the conditional, "If the alarm light came on then the bell sounded", are:

a   b
  . . .

and it may not be clear whether the conditional should be interpreted as a one-way or bi-conditional. But it may not matter. If you learn as a definite categorical fact:

The alarm light came on.

then you can forget about the implicit alternatives. They are eliminated to leave only the explicit model that supports the conclusion:

∴ The bell sounded.

Once again, you have made a deduction without fixing the precise interpretation of a premise and without using the relevant formal rule, which is known as *modus ponens*:

if p then q
p
∴q.

Other inferences may cause reasoners to make a fuller interpretation of a conditional of the form "if a then b". Its initial models, as we have seen, take the form:

       a   b
       . . .

If reasoners bear in mind that the implicit model corresponds to the situations in which the antecedent, a, does not hold, then there are several ways in which they could represent this information. For instance, they could indicate that the antecedent is exhaustively represented in the explicit model—i.e. it cannot occur in the implicit model (see Johnson-Laird and Byrne, 1991). But, they could do the same job by storing a footnote with the implicit model. The explicit model captures what happens when the antecedent is true, and so the footnote (shown here in parentheses) represents that the antecedent does not occur:

       a   b
     {¬a} . . .

where "¬" signifies negation. Similarly, the models of a bi-conditional are as follows:

       a   b
     {¬a   ¬b} . . .

The computations required to combine models are simple with footnotes, and they could be treated as merely a notational variant of the method using exhaustion.

   A computer program recently devised by the author implements several levels of expertise in sentential reasoning according to the model theory. Here, we will consider three such levels. At a simple level (level 1), the program represents a conditional premise:

   if p then q

as having the following models:

       p   q
     {¬p} . . .

The explicit model captures what happens when the antecedent occurs, and the footnote in parentheses represents that the antecedent does not occur in the implicit model (the ellipsis). At this level, the program similarly represents an inclusive disjunction:

p or q, or both

by the following models:

p

    q

p   q

in which the first model does not make explicit that q does not occur, and the second model does not make explicit that p does not occur. The computations required to interpret compound premises are simple with footnotes. Thus, given the premise:

if p then q, *and* p

the program combines each of the models for the conditional with the model for p using a semantic procedure for "and". According to this procedure, two explicit models combine to yield an explicit model that avoids unnecessary duplications:

p   q   *and* p yield p   q

If one explicit model is inconsistent with another, or with the content of a footnote on an implicit model, then no new model is formed from them—i.e. the output is the null model:

{¬p} ... *and* p yield nil

where "nil" represents the null model, which is akin to the empty set. The procedure for conjoining two sets of models thus multiplies each model in one set by each model in the other set according to the following principles:

1. If the two models are explicit, then they are joined together eliminating any redundancies, unless one model contradicts the other—i.e. one represents a proposition and the other represents its negation, in which case the output is the null model. When people combine two separate premises, they tend to drop propositions that they know categorically. For example, if they know that p is the case, then the model:

    p   q

combines with the model of the categorical premise:

p

to yield the following model:

q

2. If one model is explicit and the other model is from the footnote on an implicit model, then the result is the explicit model unless its content contradicts the model in the footnote, in which case the output is the null model.
3. If both models are implicit, then the result is an implicit model, which conjoins the footnotes on the two implicit models unless the two footnotes contradict one another, in which case the output is the null model.

As an illustration, consider the conjunction of two conditionals, "if a then b", and "if b then c", with the following respective models:

```
   a    b
 {¬a} . . .
```

and:

```
   b    c
 {¬b} . . .
```

The conjunction of the two sets of models yields:

```
   a      b    c      (rule 1)
          b    c      (rule 2)
 {¬a    ¬ b} . . .    (rule 3)
```

The conjunction of:

```
 a     b    and    {¬b} . . .
```

yields the null model according to rule 2. Although conjunction is guided by the content of footnotes, their content does not surface in any explicit models. Hence, at this stage, it is still impossible to make certain inferences, such as *modus tollens*.

In contrast, at the next level up in competence (level 2), the content of footnotes is made explicit in any models resulting from the conjunction of implicit and explicit models. The previous premises accordingly yield the models:

```
    a       b       c
   ¬a       b       c
  {¬a      ¬b} . . .
```

One consequence is that at this level the program is able to make a *modus tollens* deduction. Given the premises:

    if a then b
    not b

the conjunction of the two sets of models proceeds as follows:

```
    a    b   and  ¬b   yields  nil
  {¬a} . . .   and  ¬b   yields  ¬a
```

from which it follows:

    not a.

Performance at this level also yields negated models in disjunctions, and indeed is probably as accurate as possible given the use of implicit models.

At its highest level of performance (level 3), the program fleshes out the contents of implicit models wholly explicitly. For example, fleshing out the implicit model of a conditional:

```
    p    q
  {¬p} . . .
```

calls for ¬p to occur in every new model, whereas separate models need to be made for q and ¬q because the footnote does not constrain them. The result is accordingly:

```
    p    q
   ¬p    q
   ¬p   ¬q
```

With such wholly explicit models, the program needs only rule 1 to combine sets of models. When the models of a set of premises are wholly explicit, there are as many of them as there are rows that are true in a truth table of all the premises. Footnotes are accordingly a device that allows the inferential system to represent certain information implicitly—it can be made explicit but at the cost of fleshing out the models. The notation can

be used recursively—as it is in the computer program—to accommodate propositions of any degree of complexity.

How, in fact, do logically naive individuals reason? The data do not allow us to pin down a precise answer. What we do know, however, is that the principal predictions of the model theory are correct. When individuals reason, the greater the number of explicit models that they have to construct, the harder the task is—they take longer, they make more errors, and their erroneous conclusions tend to be consistent with the premises. Thus, for example, reasoning from inclusive disjunctions is harder than reasoning from exclusive disjunctions, reasoning from disjunctions is harder than reasoning from conditionals, and reasoning from conditionals is harder than reasoning from conjunctions (see e.g. Johnson-Laird, Byrne, & Schaeken, 1992; Bauer & Johnson-Laird, 1993). Conclusions that are merely consistent with the premises are not valid deductions, but the theory predicts such errors. They occur because reasoners construct some, but not all, of the models of the premises.

WHEN "OR" MEANS "AND": AN UNEXPECTED PREDICTION

Readers may find the account of temporal reasoning plausible, but have doubts about the extension of the model theory to sentential connectives. The apparatus of implicit models and footnotes seems implausible at first, yet, the underlying intuition is simple. Given a conditional, such as:

If the alarm light came on then the bell sounded.

individuals grasp that both events may have occurred, but defer a detailed representation of what happens if the alarm light did not come on. One striking and unexpected piece of evidence in favour of the theory was discovered from observing the performance of the computer program outlined in the previous section. Before I describe this phenomenon, however, readers might like to think for themselves about the following question: in what sorts of assertions does "or" tend to mean the same as "and"? Many people suppose the answer has to do with the specific lexical content of the assertions, but the phenomenon in question applies to neutral content about letters and numbers. In testing the program, I noticed that at level 1, an inclusive disjunction of the form:

If A then 2, *or* if B then 3.

elicited the following set of models:

```
A      2
              B      3
A      2      B      3
       . . .
```

But, the program also produced exactly the same set of models for a conjunction of the two conditionals:

If A then 2, *and* if B then 3.

It seemed at first sight that there might be a "bug" in the program (or the theory), but a closer inspection revealed that the interpretations were a consequence of an interaction between two components of the theory: the use of an implicit model in representing conditionals, and the use of partial models of disjunctions. A further test revealed that a disjunction of two conjunctions:

A and 2, *or* B and 3.

also yielded the same explicit models as those above (though no implicit model).

In order to test these predictions, we have recently carried out two experiments (Johnson-Laird & Barres, 1994). For each assertion, the subjects wrote down a list of the possible circumstances in which the assertion would be true—i.e. pairings of letters and numbers, such as "A 2". The results reliably corroborated the predictions. Many individuals do interpret a disjunction of conditionals in the same way as a conjunction of conditionals (and in the same way as a disjunction of conjunctions). Which assertions do subjects erroneously interpret? The answer is that the disjunction of the conditionals is true in many more cases than the models above allow, the conjunction of the conditionals is true in some more cases than the models above allow, but the disjunction of conjunctions is accurately represented by the models above (though they do not enumerate the different ways in which each conjunction can be false). Individuals untrained in logic therefore do seem to represent conditionals using implicit models and disjunctions using partial models.

SOME APPARENT DIFFICULTIES FOR THE THEORY OF CONDITIONALS

Despite the successes of the theory, there are certain aspects of it to which sceptics take exception (see the commentaries accompanying Johnson-Laird and Byrne, 1993). They object to the theory's account of the meaning of one-way conditionals, which it treats as equivalent to what logicians refer to as "material implication"—i.e. to the following truth table:

p	q	p implies q
T	T	T
T	F	F
F	T	T
F	F	T

The first two rows in this truth table are obvious: if the antecedent is true, the consequent has to be true, and if it is false then so too is the conditional as a whole. But the second two rows raise problems. In daily life, one does not treat a conditional as true merely because its antecedent is false. Indeed, subjects in experiments often judge that a conditional is "irrelevant" in such cases (see Johnson-Laird & Tagart, 1969; Evans, Newstead, & Byrne, 1993). Conditionals, the sceptics say, seem to go beyond a mere relation between the truth values of their antecedents and consequents—they convey a connection between the two events, such as a cause or a reason.

In fact, these objections to the model theory are not decisive. People judge a conditional as irrelevant when its antecedent is false, because such cases do not correspond to anything that they have made explicit in their initial models of the conditional. If they do flesh their models out explicitly, however, then they will realise that a conditional is certainly not falsified by a case where the antecedent is false. Such cases are entirely consistent with the conditional. In other words, the following possibilities are all allowed by the respective interpretations of the conditional:

One-way conditional		*Bi-conditional*	
p	q	p	q
¬p	q	¬p	¬q
¬p	¬q		

Hence, a one-way conditional is true in cases where the antecedent is false, and in cases where the consequent is true. These facts lead on to the so-called "paradoxes of material implication". They are not real paradoxes, but merely valid inferences that seem strange at first:

The alarm light did not come on.
∴ If the alarm light came on then the bell sounded.

and:

    The bell sounded.
∴ If the alarm light came on then the bell sounded.

The reason they are odd is because they throw semantic information away—i.e. the premises convey more information than the conclusions (see Johnson-Laird & Byrne, 1991). Similar oddities can be constructed using disjunctions—e.g.

    The alarm light came on.
∴ The alarm light came on *or* the bell sounded, or both.

This deduction is also valid, but logically untrained individuals balk at it, finding it improper. It too throws semantic information away.

    While it is true that the antecedent and consequent of a conditional are often connected causally or in some other way, the conditional *per se* does not convey the connection. It makes sense to conjoin a conditional with a denial of any such over-arching connection—e.g.

    If the alarm light came on then the bell sounded, though there was no causal or other connection between the two events.

This case contrasts with the use of the connective "because". An analogous assertion with this connective is manifest nonsense:

    Because the alarm light came on, the bell sounded, though there was no causal or other connection between the two events.

    Sceptics also object to our account of reasoning with conditionals. They argue that it makes the wrong predictions about certain conditional inferences. Thus, the initial models of a conditional allow an inference of the form known as "affirmation of the consequent":

    If a then b
    b
    ∴ a

that is valid only if the "if" premise is interpreted as a bi-conditional. But the initial models do not allow an inference of the form known as "denial of the antecedent":

    If a then b
    not a
    ∴ not b

that is also valid only if the "if" premise is interpreted as a bi-conditional. The critics say that both these inferences occur with comparable frequencies (Evans, 1993; O'Brien, personal communication). In fact, the only safe generalisation is that the frequencies of the two inferences are remarkably labile. Some studies do indeed report that affirmation of the consequent occurred more often than denial of the antecedent. There is now known to be a marked "figural effect" in propositional reasoning—that is, individuals tend to frame conclusions in the same order as the information in them entered working memory (see Bauer & Johnson-Laird, 1993). This effect may depress the frequency of affirmation of the consequent (and *modus tollens*). Once reasoners flesh out their models more explicitly, they will affirm the consequent and deny the antecedent only if they make the bi-conditional interpretation.

Another objection concerns inferences of the following form:

If p and either q or r then s or both t and u
p and either q or r
∴s or both t and u.

This deduction calls for many models, and so it ought to be difficult. Clearly, it is not difficult, and so the model theory seems to be wrong. In fact, the theory allows that reasoners can construct a representation of the meaning of premises—indeed, the theory needs such a representation so that the manipulation of models can occur without losing track of the meaning of the premises. This representation enables reasoners to notice the match between one proposition and another. Indeed, just such a process was part of the first attempt to devise a semantic account of reasoning (see Johnson-Laird, 1983, p. 46 *et seq.*). If reasoners notice such a match, as in the premises above, they do not have to construct detailed models of the premises. They need only make an inference of the form:

If A then B
A
∴B

where A = "p and either q or r", and B = "s or both t and u". There is no reason to suppose that they will use the formal rule of *modus ponens*. The standard model-theoretic machinery will do as well. What would challenge the model theory is a case where an easy inference can be made only by constructing many models. So far, no such cases have been forthcoming.

THE PHENOMENA OF THE
SELECTION TASK

The aim of this section is to describe one particular reasoning task, which has probably attracted more competing explanatory hypotheses than any other. The task was invented by Peter Wason (1966), investigated by Wason and the present author, and has inspired many subsequent investigations (for a review, see Evans et al., 1993). This section will describe five of the task's main phenomena, and the next section will show how the model theory accounts for them.

In an early version of the selection task, Wason laid four cards on the table:

A   B   2   3

and the subjects knew that every card had a letter on one side and a number on the other side. Their task was to select those cards that needed to be turned over to find out whether the following conditional rule was true or false:

If there is an "A" on one side of a card, then there is a "2" on the other side.

The materials do not engage any existing knowledge about the relation at stake. With such "neutral" materials, most subjects selected the "A" card alone, or the "A" card and the "2" card. Only a few subjects selected the "3" card. Yet, if this card were turned over to reveal an "A" on the other side, it would falsify the rule. The failure to select the card corresponding to the false consequent of the conditional is robust, and it is the first phenomenon of the selection task.

An experimental manipulation pioneered by Evans (e.g. Evans & Lynch, 1973) is to use conditionals with negative constituents in the selection task. Given the conditional:

If there is an "A" on one side, then there is *not* a "2" on the other side.

subjects perform much better. They tend to select the "A" card and the card that falsifies the consequent—i.e. the "2" card. But, this gain in perform-ance may not reflect a better insight into the task. Evans (1989) argues that subjects are not really reasoning at all, but are guided by two heuristics that lead them merely to select whatever card makes the

antecedent true (the "if" heuristic), and whatever card is mentioned in the consequent, whether or not it is negated (the "matching" heuristic). Thus, given the conditional:

If there is *not* an "A" card on one side, then there is a "2" card on the other side.

the subjects select the card that makes the antecedent true (i.e. the "B" card), and the card mentioned in the consequent (i.e. the "2" card). The tendency to ignore negations in the consequent of conditionals and to select whatever card is mentioned in them is the second phenomenon of the selection task.

Anyone with a syntactic conception of deductive reasoning is likely to be puzzled by the subjects' failure to select a potentially falsifying card. That arch-formalist, Jean Piaget, wrote that "reasoning is nothing more than the propositional calculus itself" (Inhelder & Piaget, 1958, p. 305), and that faced with verifying whether x implies y, a subject "will look in this case to see whether or not there is a counter-example x and non-y" (Piaget, in Beth & Piaget, 1966, p. 181). It seems that adult subjects in the selection task have not reached the Piagetian level of formal operations. Yet they are supposed to have attained it around the age of 12. There is a still more striking effect. By changing the content of the selection task, subjects' performance was strikingly enhanced. They tended to make the correct selections with a rule such as "Every time I go to Manchester I travel by train" (Wason & Shapiro, 1971), and with the deontic task of checking for violations to a postal regulation, such as: "If an envelope is sealed, then it must have a 5d stamp on it" (Johnson-Laird, Legrenzi, & Legrenzi, 1972). These results were not so robust, however, and some studies failed to replicate them. One sort of content did yield a reliable improvement in performance. As Griggs and Cox (1982) showed, deontic conditionals, such as:

If a person is drinking beer then the person must be over 18.

tended to elicit correct selections (the card corresponding to a beer drinker, and the card corresponding to someone less than 18). This finding is the third robust phenomenon of the selection task. It has led some theorists to postulate a special role for knowledge of the relevant situation. Others argue that deontic contents trigger special "pragmatic reasoning schemas" (Cheng & Holyoak, 1985), such as:

If the precondition is not satisfied (e.g. not over 18 years) then the action (e.g. drinking beer) must not be taken.

And still others argue for the relevance of social contracts and particularly for a procedure for checking for cheaters that has evolved as a result of natural selection (see Cosmides, 1989; Gigerenzer & Hug, 1992).

The hypothesis about checking for cheaters has led to the discovery of quite different patterns of selections in the task. Cosmides (1989) told subjects a story about a fictitious tribe that culminated in a statement of the condition governing the right to eat cassava root:

If a man has a tattoo on his face then he eats cassava root.

The subjects tended to check for those who might be cheating—i.e. they selected the card corresponding to the *false* antecedent (man with no tattoo) and a *true* consequent (eats cassava root). Given the following conditional used by Manktelow and Over (1991):

If you tidy your room then you may go out to play.

the subjects' selections depended on whether they were asked to take the point of view of the mother, who laid down the condition, or the point of view of the child, who was on the receiving end of it. The mother is presumably concerned that her child does not cheat, and subjects who take her point of view tended to select the following cards:

did not tidy   (false antecedent)
went out to play   (true consequent)

The child is presumably concerned that the mother does not renege on the deal, and the subjects who took the child's point of view tended to select the following cards:

tidied the room   (true antecedent)
did not go out to play   (false consequent)

Politzer and Nguyen-Xuan (1992) demonstrated similar effects with a conditional of the form:

If the purchase exceeds 10,000 francs, then the salesman must stick on the back of the receipt a voucher gift for a gold bracelet.

The selections depended on whether the subjects took the point of view of a manager checking for cheating customers or the point of view of a customer checking the deal. Subjects with a neutral point of view tended to select all four cards. Manktelow and Over argue that subjects assess the utilities of the outcomes in ways that go beyond social contracts or checking for cheaters. Politzer and Nguyen-Xuan defend a revised version of pragmatic reasoning schemas. The selection of false antecedent and true consequent cards, and sometimes of all four cards, is the fourth phenomenon of the selection task.

There are many other findings on the selection task, but one other phenomenon is often lost sight of in the controversies over pragmatic reasoning schemas, social contracts, and utilities: the task can be made easier by changing the form of the rule. Wason and Johnson-Laird (1969) observed an improved performance with a disjunction instead of a conditional. Likewise, Wason and Green (1984) showed that subjects performed more accurately with a rule concerning a single sort of entity and its property:

All the circles are black.

than with a rule interrelating two entities. Such improvements are the fifth phenomenon of the selection task.

The five phenomena are accordingly:

1. With neutral conditionals of the form, if p then q, subjects tend to select the p card alone, or the p and q cards.
2. With neutral conditionals containing negations, subjects tend to select the card that makes the antecedent true and the card mentioned in the consequent, whether or not the consequent is negated—e.g. with "if p then not q", subjects tend to select the p card and the q card.
3. With a realistic content, subjects carry out the task more accurately.
4. Depending on the subjects' point of view about a deontic rule of the form if p then q, they tend to select the p and not-q cards, or the not-p and q cards, or all four cards.
5. A change in the form of the rule can yield a greater number of correct selections.

One final word of warning before we turn to the explanation of these phenomena. It is tempting to suppose that an experimental manipulation either yields insight into the task or not—indeed, papers on the selection task often adopt this dichotomous point of view. In fact, the range of

performance covers the entire spectrum. Experimenters typically report reliable improvements in performance above the level with a neutral conditional, but the degree of the improvement varies considerably from one manipulation to another. No manipulation yields perfect selections in one condition and wholly erroneous selections in another.

HOW THE MODEL THEORY EXPLAINS THE SELECTION TASK

Some authors doubt whether the selection task engages any process of reasoning at all: it may just elicit biases or trigger relevant knowledge. Or, say others, it may engage a process of decision making in which subjects compute the expected utilities of different outcomes. Or it may engage special mental procedures with which evolution has equipped human beings to enable them to live a rich social life. The diversity of phenomena is matched by the diversity of explanations. Many of the explanations, however, have little psychological justification from outside the selection task; and none of them explains all five of the phenomena. According to the model theory, however, subjects *do* reason in the selection task—albeit not always successfully, and the theory does provide a unified explanation of the five phenomena (Johnson-Laird & Byrne, 1991).

When people reason, as we have seen, they can do so only on the basis of what is explicit in their models of the premises. This principle applies to the selection task too: people reason only about what is explicit in their models of the rule. The task requires subjects to select those cards that they need to turn over to determine whether an indicative rule is true or false, or whether a deontic rule has been violated. Hence, the model theory predicts that subjects will apply these constraints but only to those cards that they have explicitly represented in their models of the rule. A one-way interpretation of the neutral conditional:

If there is an "A", then there is a "2".

yields the models:

    A   2
    {¬A} . . .

and a bi-conditional interpretation yields the models:

    A   2
    {¬A ¬2} . . .

Subjects should therefore select the "A" card alone in the one-way case, and the "A" card and the "2" card in the bi-conditional case. The theory accordingly accounts for the first phenomenon of the selection task: these are the predominant selections for neutral conditionals. The theory goes beyond the phenomenon to make a further prediction: there should be a correlation between the interpretation of the conditional (as a one-way or bi-conditional) and the pattern of selections. This correlation has been recently confirmed by Francesco Cara and Stefana Broadbent (personal communication).

When do people use negative assertions? The answer according to Wason (1965) is in order to correct misconceptions—e.g. "A spider is not an insect". Hence, the proposition to be corrected is ordinarily one that the listener or reader has in mind—that is, it should be separately represented in a mental model. This assumption also applies to the interpretation of conditionals with negative constituents. With the conditional:

If there is an "A", then there is not a "2".

if subjects tend to represent both the consequent and the proposition that it negates, they will construct the following models for a one-way interpretation:

$$A \quad \neg 2$$
$$2$$
$$\{\neg A\} \ldots$$

Likewise, for a one-way conditional with a negated antecedent:

If there is not an "A", then there is a "2".

they will tend to construct the following models:

$$\neg A \quad 2$$
$$A$$

Reasoners may even omit the negative propositions from these models. In any case, their models of the first of these conditionals, though not the second, should include the negated consequent, and so they should consider this card in carrying out their selections. It bears on the truth or falsity of the conditional, and so they should tend to select it. The card corresponding to the true antecedent, which is also explicitly represented, bears on the truth or falsity of the conditional and so subjects should also tend to select it. The model theory thus accounts for the second phenomenon of the

selection task. But, it makes an additional prediction: negating the antecedent of a conditional should yield a proposition that can be paraphrased by a disjunction, whereas negating the consequent should not. Thus, the second of the two conditionals above, but not the first, can be paraphrased as:

Either there is an "A" or there is a "2".

There are no experimental data, but the equivalence was noted by Stoic logicians long ago (see Kneale & Kneale, 1962, p. 162).

When should subjects get the selection task right? One necessary condition according to the model theory is that they represent the negated consequent in their models of the conditional. But, as Johnson-Laird and Byrne (1991, p. 80) write: an insightful performance may further depend on an explicit representation of what is not possible—i.e. the real impossibility given the rule [if "A" then "2"] of:

A   ¬2

This idea relates to the model theory's treatment of counter-factual conditionals, such as:

If there had been an "A" then there would have been a "2".

which should normally elicit the following models:

Actual state: ¬A   ¬2
Counter-factual states: A   2
                            {¬A} . . .

The epistemic status of the different models has to be represented, as the labels show.

The models of an indicative conditional, such as:

If there is an "A" then there is a "2".

are represented as real possibilities. This apparatus also allows a model to be represented as impossible. And, in order to test the truth of a conditional, it may be necessary to focus on what it rules out as impossible. It follows that any experimental manipulation that leads individuals to flesh out their models of the conditional should tend to improve performance in the selection task. One such manipulation is the use of a content that is likely to make violations of the rule salient, either by triggering specific memories

or by eliciting a framework in which such violations are highlighted. The model theory therefore accounts for the third phenomenon of the selection task—the enhanced performance with realistic materials and especially those of a deontic sort—without invoking permission schemas (Cheng & Holyoak, 1985) or special procedures for checking for cheaters (Cosmides, 1989). However, the theory goes beyond these findings to make a general prediction. Any manipulation that draws attention to counter-examples should improve performance even if the materials are not deontic (Johnson-Laird & Byrne, 1991, p. 80–81). This prediction has been corroborated recently by two unpublished studies. Daniel Sperber and Vittorio Girotto (personal communication) have established the point using several domains. For example, they told their subjects that a certain machine generated cards according to the rule:

If a card has an "A" on one side, then it has a "2" on the other side.

The machine went wrong and ceased to obey the rule, but it has been repaired, and the subjects have to check that the job has been done properly. They are thus likely to represent the machine's potential error explicitly: A ¬2. The subjects in this condition and other similar ones carried out the selection task more accurately.

Roberta Love and Claudius Kessler (personal communication) have obtained similar results. For example, in a science fiction domain, they used the conditional rule:

If there are Xow then there must be a force field.

where Xow are strange crystal-like living organisms who depend for their existence on a force field. In a context that suggested the possibility of counter-examples—mutant Xows who can survive without a force field— the subjects carried out the selection task more accurately than in a control condition that did not suggest such counter-examples. As Fillenbaum (1977) and others have argued, many conditional promises in daily life call for a bi-conditional interpretation. For example, the conditional:

If you tidy your room then you can go out to play.

is taken to imply:

If you don't tidy your room then you cannot go out to play.

When the models for such a bi-conditional are fleshed out in a fully explicit way, they are as follows:

```
 t      p
¬t     ¬p
```

where "t" denotes tidying your room and "p" denotes going out to play. There are two potential counter-examples in this case:

```
 t     ¬p
¬t      p
```

Hence, a proper test of the conditional calls for selecting all four cards (as a neutral point of view elicits from subjects). If subjects focus on the first counter-example, they will select the t and not-p cards; if they focus on the second counter-example, they will select the not-t and p cards.

One way that experiments have manipulated the focus is by asking subjects to take a particular point of view. In conditional promises, there is usually an asymmetry between the violations. On the one hand, for the mother the salient interpretation of the conditional is:

If you don't tidy your room, then you cannot go out to play.

and the salient violation is that you don't tidy your room but nevertheless go out to play:

```
¬t     p
```

On the other hand, for the child the salient interpretation is:

If I tidy my room, then I can go out to play.

and the salient violation is, I tidy my room and yet I don't get out to play:

```
 t     ¬p
```

The model theory accordingly explains the fourth phenomenon of the selection task: when deontic rules are interpreted as bi-conditionals, instructions can make salient one or other (or both) of the counter-examples.

Once again, the theory goes beyond the obtained results to make a further prediction. With a factual conditional that strongly suggests a bi-conditional interpretation, such as:

If candidates are qualified, then they passed the exam.

the same phenomena should occur as with the deontic conditionals. When counter-examples are made salient, subjects with a neutral point of view should tend to select all four cards. The suggestion that some candidates may have qualified improperly should make salient the counter-example:

    qualified   ¬passed exam

In contrast, the suggestion that some candidates may have been improperly debarred should make salient the counter-example:

    ¬qualified   passed exam

Such a result would show that a deontic content is not essential to the phenomenon.

The explanation of the fifth phenomenon should now be obvious. The purpose of implicit models is to minimise the load on working memory. Manipulations that lead to explicit models, or that minimise load, should improve performance on the selection task. Hence, the task should be easier with disjunctions, which have no wholly implicit models, and with conditionals about single entities rather than two entities. Any manipulation that reduces the memory load should be effective, and indeed this prediction has also been corroborated (see e.g. Oakhill & Johnson-Laird, 1985). Most theories of cognition are likely to allow for such effects, but they are a natural consequence of the model theory with its emphasis on minimising explicit representations.

CONCLUSIONS

There is a gulf between formal and semantic methods in logic. The two sorts of psychological theories of reasoning similarly differ so much that experimental evidence should enable us to decide between them. In practice, it has not been easy. There are many different rule theories, and several of them are still under active development: theorists continue to modify the rules of inference and the strategies that deploy them. The model theory also continues to develop—the version for propositional reasoning described here differs from previous accounts. It postulates slightly different models and the principles for combining them. Hence, if the controversy is to be resolved, then empirical results need to contravene the fundamental principles of a theory rather than just a specific version of it.

Formal rule theories apply to only limited domains of deductive reasoning. They predict that the difficulty of a deduction depends on the

length of the derivation and the accessibility or ease of use of the particular rules on which it depends. The rules and the strategy differ from one version of a theory to another, and it is therefore hard to specify what could strike at the heart of rule theories—perhaps no experiment can do better than to test a specific rule theory. In contrast, the model theory is easy to refute in principle. It applies to all domains of deduction, and it makes two general predictions. The first prediction is that erroneous conclusions will tend to be at least consistent with the premises. This prediction can be tested without any account of the particular models for the domain: it is necessary to test merely that the erroneous conclusions tend to be consistent with the premises rather than inconsistent with them. The prediction has been corroborated in all the main domains of deduction (see e.g. Johnson-Laird & Byrne, 1991). The second prediction is that the greater the number of models on which an inference depends, the harder it will be—it will take longer, and it will be more likely to go wrong. In temporal reasoning, the number of models is easy to assess and, as shown in the "Theory of mental models" section above, it correctly predicts difficulty. In other domains, such as syllogisms, the prediction requires a detailed account of the models constructed from premises. Sentential reasoning stands somewhere between these two domains. As the "Models and sentential reasoning" section above showed, a conjunction (a and b) calls for one model; an exclusive disjunction (a or else b) calls for two models; and an inclusive disjunction (a or b, or both) calls for three models. No version of the model theory could realistically change these numbers, and they make the correct predictions about inferential difficulty (see Johnson-Laird et al., 1992).

What complicate matters are conditionals. The "Models and sentential reasoning" section defended the view that they tend to be interpreted by one explicit model and one implicit model with an attached mental footnote. The next section in this chapter ("When 'or' means 'and': an unexpected prediction") showed that this hypothesis leads to the correct prediction that logically untrained individuals will interpret a disjunction of conditionals:

If A then 2, or if B then 3.

as true in the same cases as the conjunction of the conditionals:

If A then 2, and if B then 3.

The next section ("Some apparent difficulties for the theory of conditionals") defended the theory against some potential objections, and the final two sections showed how it elucidates the main phenomena of Wason's selection task.

In conclusion, there are three crucial distinctions between formal rule theories and the mental model theory. First, the rule theories lack a decision procedure whereas the model theory has a simple one. In contrasting a semantic and a syntactic method in logic, Quine (1974, p. 75) wrote: "[The syntactic method] is inferior in that it affords no general way of reaching a verdict of invalidity; failure to discover a proof for a schema can mean either invalidity or mere bad luck". The same problem vitiates the psychological theories based on formal rules: "The 'search till you're exhausted' strategy gives one at best an educated, correct guess that something does not follow" (Barwise, 1993). A conclusion is valid if, and only if, it holds in all models of the premises, and so the model theory provides a simple decision procedure in many domains of deduction.

The second comparison concerns predictive power. Rule theories exist for spatial and sentential reasoning. Unlike the model theory, they have nothing to say about temporal or modal reasoning, or reasoning with quantifiers such as "most" or "more than half" that do not correspond to those of formal logic. Even within sentential reasoning, current formal rule theories give no account of several robust phenomena, including the nature of erroneous conclusions, or the results of Wason's selection task. The model theory explains these phenomena, and it goes beyond deduction to suggest a unified account of different modes of thought—for example, induction and creation depend on adding semantic information to models (e.g. Johnson-Laird, 1993). It also yields a foundation for probabilistic thinking and informal reasoning: the strength of such inferences depends on the proportion of possible states of affairs consistent with the premises in which the conclusion is true, and subjects can estimate this proportion by constructing models, which each correspond to an infinite set of possibilities (or, in some cases, a set of infinite sets of possibilities, Johnson-Laird, 1994).

The third point concerns falsifiability. There seem to be no experiments on verbal reasoning that could contravene the fundamental principles of formal rule theories, whereas experiments can in principle refute the fundamental principles of the model theory. One way in which formal rule theories might be refuted is to consider inferences based on visual perception. A rule theory posits the extraction of logical form from an internal description of the percept, whereas the model theory assumes that vision leads directly to a mental model (Marr, 1982). The two sorts of theories therefore diverge on the matter of diagrams. According to the model theory, it should be easier to reason from a diagram making explicit the alternative possibilities than from logically equivalent verbal premises. With a diagram, reasoners do not need to engage in the process of parsing and compositional semantic interpretation. Formal rule theories, however, have no grounds for such a prediction. With a diagram, reasoners have to

construct an internal description from which they can extract a logical form, and there is no reason to suppose that the process should be easier than the extraction of logical form from verbal premises. In fact, a recent study has shown that subjects reasoned reliably faster (about 30 sec) and considerably more accurately (30% increase in the number of valid conclusions) when the premises were diagrams rather than sentences (Bauer & Johnson-Laird, 1993). This result corroborates the model theory and perhaps contravenes the rule theories.

ACKNOWLEDGEMENTS

REFERENCES

Barwise, J. (1993). Everyday reasoning and logical inference. (Commentary on Johnson-Laird & Byrne, 1991.) *Behavioral and Brain Sciences, 16,* 337–338.

Bauer, M.I., & Johnson-Laird, P.N. (1993). How diagrams can improve reasoning. *Psychological Science, 4,* 372–378.

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

Braine, M.D.S., & O'Brien, D.P. (1991). A theory of If: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review, 98,* 182–203.

Byrne, R.M.J., & Johnson-Laird, P.N. (1989). Spatial reasoning. *Journal of Memory and Language, 28,* 564–575.

Cheng, P.N., & Holyoak, K.J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17,* 391–416.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31,* 187–276.

Evans, J.St.B.T. (1989). *Bias in human reasoning: Causes and consequences.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J.St.B.T. (1993). The mental model theory of conditional reasoning: Critical appraisal and revision. *Cognition, 48,* 1–20.

Evans, J.St.B.T, & Lynch, J.S. (1973). Matching bias in the selection task. *British Journal of Psychology, 64,* 391–397.

Evans, J.St.B.T., Newstead, S.E., & Byrne, R.M.J. (1993). *Human reasoning: The psychology of deduction.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Fillenbaum, S. (1977). Mind your p's and q's: The role of content and context in some uses of and, or, and if. In G.H. Bower. (Ed.), *Psychology of learning and motivation, Vol. II.* New York: Academic Press.

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43,* 127–171.

Griggs, R.A., & Cox, J.R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology, 73,* 407–420.

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

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness.* Cambridge: Cambridge University Press.

Johnson-Laird, P.N. (1993). *Human and machine thinking.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Johnson-Laird, P.N. (1994). Mental models and probabilistic thinking. *Cognition.*

Johnson-Laird, P.N., & Barres, P.E. (1994). When "or" means "and": A study in mental models. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc., pp. 475–478.

Johnson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Johnson-Laird, P.N., & Byrne, R.M.J. (1993). Precis of "Deduction", and authors' response: Mental models or formal rules? *Behavioral and Brain Sciences, 16,* 323–333, 368–376.

Johnson-Laird, P.N., Byrne, R.M.J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review, 99,* 418–439.

Johnson-Laird, P.N., Legrenzi, P., & Legrenzi, M.S. (1972). Reasoning and a sense of reality. *British Journal of Psychology, 63,* 395–400.

Johnson-Laird, P.N., & Tagart, J. (1969). How implication is understood. *American Journal of Psychology, 82,* 367–373.

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

Legrenzi, P. (1970). Relations between language and reasoning about deductive rules. In G.B. Flores D'Arcais, & W.J.M. Levelt (Eds.), *Advances in psycholinguistics.* Amsterdam: North-Holland.

Macnamara, J. (1986). *A border dispute: The place of logic in psychology.* Cambridge, MA: Bradford Books, MIT Press.

Manktelow, K.I., & Over, D.E. (1991). Social roles and utilities in reasoning with deontic conditionals. *Cognition, 39,* 85–105.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: W.H. Freeman.

Oakhill, J.V., & Johnson-Laird, P.N. (1985). Rationality, memory, and the search for counter-examples. *Cognition, 20,* 79–94.

Osherson, D.N. (1974–76). *Logical Abilities in Children, Vols. 1–4.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Quine, W.V.O. (1974). *Methods of logic*. Third edition. London: Routledge & Kegan Paul.

Politzer, G., & Nguyen-Xuan, A. (1992). Reasoning about conditional promises and warnings: Darwinian algorithms, mental models, and relevance judgements or pragmatic schemas. *Quarterly Journal of Experimental Psychology, 44,* 401–412.

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

Schaeken, W., Johnson-Laird, P.N., & d'Ydewalle, G. (1993). *Mental models and temporal reasoning*. Unpublished paper, Department of Psychology, Princeton University.

Simon, H.A. (1959). Theories of decision making in economics and behavioral science. *American Economic Review, 49,* 253–283.

Wason, P.C. (1965). The context of plausible denial. *Journal of Verbal Learning and Verbal Behavior, 4,* 7–11.

Wason, P.C. (1966). Reasoning. In B.M. Foss (Ed.), *New horizons in psychology*. Harmondsworth: Penguin.

Wason, P.C., & Green, D.W. (1984). Reasoning and mental representation. *Quarterly Journal of Experimental Psychology, 36A,* 597–610.

Wason, P.C., & Johnson-Laird, P.N. (1969). Proving a disjunctive rule. *Quarterly Journal of Experimental Psychology, 21,* 14–20.

Wason, P.C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology, 23,* 63–71.

Perspectives on Thinking and Reasoning

Essays in Honour of Peter Wason

Edited by

Stephen E. Newstead

and

Jonathan St.B.T. Evans
Department of Psychology,
University of Plymouth

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library