Cognition, 16 (1984) 1–61

Syllogistic inference

P.N. JOHNSON-LAIRD*

MRC Applied Psychology Unit, Cambridge

BRUNO G. BARA

University of Milan

Abstract

This paper reviews current psychological theories of syllogistic inference and establishes that despite their various merits they all contain deficiencies as theories of performance. It presents the results of two experiments, one using syllogisms and the other using three-term series problems, designed to elucidate how the arrangement of terms within the premises (the 'figure' of the premises) affects performance. These data are used in the construction of a theory based on the hypothesis that reasoners construct mental models of the premises, formulate informative conclusions about the relations in the model, and search for alternative models that are counterexamples to these conclusions. This theory, which has been implemented in several computer programs, predicts that two principal factors should affect performance: the figure of the premises, and the number of models that they call for. These predictions were confirmed by a third experiment.

Introduction

The ability to make deductions that depend on quantifiers is a prerequisite for everyday thinking and for mathematics and science. Quantifiers include

*We thank Stefania Bandini, M. Caterina Gallo, Michele Neri, Giuliano Geminiani and Alison Black for their technical assistance, and Patrizia Tabossi for carrying out a replication of Experiment 2. We are also grateful to Steve Isard, Antonella Carassa and Marco Colombetti for their help in devising the computer programs, to A.R. Jonckheere for stimulating statistical advice, and to Martin Braine, Earl B. Hunt, Jane Oakhill and Russ Revlin for discussion and criticisms of earlier versions of this paper. Part of our research was supported by grants from the Social Science Research Council and the European Training Programme. Reprint requests should be sent to P.N. Johnson-Laird, MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, U.K.

0010-0277/84/$18.80

such expressions as 'all', 'most', 'some', 'few', 'more than half', 'finitely many', 'uncountably many', and so on. A subset of these quantifiers give rise with simple predicates to syllogisms, which were originally analysed by Aristotle. A syllogism consists of two premises and a conclusion, which can each occur in one of four 'moods', which we state here together with their customary mnemonics:

All X are Y	(A: a universal affirmative premise)
Some X are Y	(I: a particular affirmative premise)
No X are Y	(E: a universal negative premise)
Some X are not Y	(O: a particular negative premise)

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

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

There are accordingly 64 possible logical forms for the premises of a syllogism (4 moods for each premise × 4 figures). Twenty-seven of these premise pairs yield valid conclusions interrelating the end terms provided that one bears in mind that a conclusion may take the form A – C or C – A. If one restricts the conclusions to just one form, say C – A, which is the format favoured by medieval Scholastic logicians, but not Aristotle, then only 19 pairs of premises yield valid conclusions interrelating the end terms. We shall often speak of the other premises as not yielding valid conclusions, but this is a *façon de parler* since, in fact, any set of premises yields an infinite number of valid conclusions, e.g. conclusions that consist of conjunctions or disjunctions of the premises. The point is that these other premises do not yield a valid conclusion interrelating the end terms, and people hardly ever draw valid conclusions from them—they either draw an invalid conclusion that does interrelate the end terms or they claim that nothing follows from the premises. (The set of syllogistic premises that yield valid conclusions interrelating the end terms can be found in the Appendix.)

Since the turn of the century, the majority of psychological studies of reasoning with quantifiers have concerned syllogisms (see Evans, 1982), and this concentration of effort is sensible, because they lie on the borderline of human competence. Some syllogisms are very easy and nearly everyone gets them right, but others are very hard and nearly everyone gets them wrong. For example, given premises of the form:

Some of the artists are beekeepers
All the beekeepers are chemists

the majority of subjects readily infer the valid conclusion:

Some of the artists are chemists

(see Johnson-Laird and Steedman, 1978). But, given the premises:

None of the archers are boxers
All the boxers are clerks

hardly anyone draws the correct conclusion:

Some of the clerks are not archers.

Syllogisms therefore make an excellent test case for cognitive science: they are at the centre of quantificational reasoning; they have only a small number of different forms; and performance with them varies considerably but has yet to be adequately explained. If psychologists prove unable to account for how people make these deductions, then they are unlikely to succeed in explaining any complex cognitive functions.

In some areas of cognition, it is possible to develop a theory of competence that specifies what the mind has to compute and a theory of performance that outlines a plausible algorithm for carrying out these computations. Thus, for example, linguists propose theories of competence in the guise of grammars, and psycholinguists develop theories of performance in the form of parsing algorithms. Unfortunately, in the case of deductive reasoning, no-one has ever formulated a theory of competence. The resulting theoretical gap has been filled by the largely tacit assumption that since logic encompasses the set of valid deductions, it characterizes ideal human competence. This assumption leads naturally to the view that performance is based on logical techniques for the computation of valid deductions. We will argue in due course that it is a mistake to base a theory of competence on logic, and a mistake to base a theory of performance on logical techniques. The majority of theories of the syllogism, however, have adopted one or other of these expedients.

Our aim is to present a general theory of deductive reasoning that provides an account of syllogistic inference as a special case. Our theory has grown out of a series of earlier studies carried out by the first author and his colleagues, and we will briefly review them in order to set the present paper in its context. The first finding was a striking but problematical clue to performance with syllogisms: the figure of the premises exerts a response bias on the valid conclusions produced by subjects (see Wason and Johnson-Laird, 1972). The effect was shown to apply equally to invalid conclusions in an experiment carried out with Janellen Huttenlocher, and a hypothesis accounting for this so-called 'figural bias' was proposed as part of a general information-processing theory of syllogistic inference (see Johnson-Laird, 1975). The experiment and a refinement of the theory, which was modelled in a computer program, were presented by Johnson-Laird and Steedman (1978). The two fundamental assumptions of this theory are that the premises are represented as 'mental models' and that the figural response bias is a reflection of how information about the premises is represented in these models. Subsequent work explored the nature of mental models and led to the proposal that they are a general representation used for all sorts of inference, not just syllogisms, and for all sorts of discourse (see Johnson-Laird, 1980; 1983). We also embarked on the present series of experiments, which, as we shall argue, cast doubt on the earlier explanation of the figural bias. The present paper reports these experiments and presents a new theory of syllogistic inference based on the concept of mental models. During the several years that the theory has been under development, various 'snapshots' of its current state have been published (see Johnson-Laird, 1982; 1983), but these accounts have been both partial and defective in matters of detail. We have now modelled the new theory in a suite of computer programs, and we will present the first complete account of its definitive form in this paper.

The notion of a mental model is a subtle one. Its crucial characteristics as far as inference is concerned are that a mental model is finite, computable, and contains tokens in relations that represent entities in a specific state of affairs. A premise that describes a particular situation can be represented by a single mental model even if the description is incomplete or indeterminate. The initial model can be thought of as constructed by a procedure that makes plausible assumptions on the basis of general knowledge and even makes arbitrary assumptions if there is no relevant information. If these assumptions turn out to be wrong in the light of subsequent discourse, then the procedures can revise the model, if possible, so as to be consistent with the discourse as a whole. For example, if you are told that all the women in the room are feminists, you may build a model on the basis of the arbitrary assumption that there are three women in the room. If you subsequently learn that there are five women in the room, then you can revise the model appropriately. The content captured in a model is therefore a function of both the model and the processes that can revise and evaluate it. In effect, a single model can stand for an infinite number of possible states of affairs—all those that are compatible with the description on which it is based. There are obviously limits on the revision of a mental model: people forget the original description, the process of revision may place too great a cognitive load on the system, and so on. Nevertheless, it is possible to advance a psychological theory of inference based on this idea of manipulating mental models.

There is a general semantic principle that governs all valid deductions: *an* *inference is valid if its conclusion is true in every possible interpretation of its premises.* All logical calculi are designed to capture the set of inferences that meet this semantic criterion, though, as we shall see, not all calculi can completely succeed in this task. What the principle means is that in theory any deduction can be made using the following general procedure:

Step 1: construct a mental model of the premises, i.e. of the state of affairs they describe.

Step 2: formulate, if possible, an informative conclusion that is true in all models of the premises that have so far been constructed. An informative conclusion is one that, where possible, interrelates terms not explicitly related in the premises. If no such conclusion can be formulated, then there is no interesting conclusion from syllogistic premises.

Step 3: if the previous step yields a conclusion, try to construct an alternative model of the premises that renders it 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.

If a *given* conclusion has to be evaluated, then all that is required is a simplified version of step 3: try to construct an alternative model of the premises that renders the conclusion false. If there is such a model, then the conclusion is invalid; but if there is no such model, then the conclusion is valid.

There can, of course, be no general decision procedure for the first-order predicate calculus (see e.g. Boolos and Jeffrey, 1980). But, where a model is finite and there are only a finite number of alternatives to it, then the procedure above yields an effective decision about the validity of any deduction. It even works for deductions from the following sort of premises:

More than half the artists are beekeepers
More than half the artists are chemists.

These premises can be represented by the following model:

chemist =	artist =	beekeeper
	artist =	beekeeper
chemist =	artist	
chemist		beekeeper

which can be constructed from a knowledge of the truth conditions of the premises. The procedures for constructing the model make a number of arbitrary assumptions, e.g. that there are three artists, since the meaning of 'more than half' calls only for a plural number of individuals. The model supports an informative conclusion establishing a relation not explicitly stated in the premises:

At least one chemist is a beekeeper.

Any attempt to build a model in which there is no chain of identities leading from a chemist to a beekeeper violates the meaning of the premises. Hence, the deduction is valid. What is interesting about this simple inference is that it cannot even be expressed within the standard first-order predicate calculus, because there is no way to express the quantifier, 'more than half', in terms of quantification over individuals. To capture this quantifier, one needs to quantify over sets, that is, one needs the second-order quantificational calculus (see Barwise and Cooper, 1981). This calculus, however, is 'incomplete' in that there is no way to specify formal rules of inference for it that enable the complete set of valid deductions to be derived. Doubtless, a subset of the calculus can be formalized that allows the conclusion above to be derived (cf. Keisler, 1970), but the fact that the problem arises at all with such a simple inference casts some doubt on the utility of the first-order predicate calculus as an implicit model of competence.

We shall argue that the mental model account of deductive competence yields a theory of syllogistic performance that is more psychologically plausible than any other theory. The first section of our paper accordingly assesses existing theories and establishes that they all have some shortcomings. The second section presents the results of two experiments designed to elucidate the figural bias. These results lead us to reject our earlier theory of syllogistic inference. The third section of the paper outlines the new theory of syllogistic inference, which accounts for all the effects of figure and for the systematic errors made by subjects. In the fourth section, we report an experiment designed to test the main predictions of the new theory. Finally, in the fifth section, we draw some general conclusions about the status of our theory.

1. Theories of syllogistic inference

The 'atmosphere' hypothesis

The early experimental studies of syllogisms concerned the sources of error in performance. Woodworth and Sells (1935) proposed that reasoners tend to accept a conclusion that is congruent with the mood created by the 'atmosphere' of the premises. The initial hypotheses about atmosphere were clumsy, but subsequently a succinct formulation was proposed by Woodworth and Schlosberg (1954, p. 846) and Begg and Denny (1969):

A negative premise creates a negative atmosphere, even when the other premise is affirmative.

A particular ("some") premise creates a particular atmosphere even when the other premise is universal.

Since the effect is apparently stronger for valid than for invalid conclusions, there must be, as Woodworth and Sells allowed, an independent inferential mechanism.

The main evidence against the atmosphere hypothesis is that when subjects are asked to state in their own words what follows from syllogistic premises, they often respond that there is no valid conclusion (Johnson-Laird and Steedman, 1978). Such responses contravene the hypothesis, because there are always two possible conclusions congruent with the atmosphere of the premises. Moreover, subjects even make the "no valid conclusion" response when there is a valid conclusion, and indeed one that is in accordance with the atmosphere. For example, given premises of the form:

Some B are A
No C are B

there is a valid conclusion in accordance with the atmosphere:

Some A are not C

yet 60% of subjects responded that there was no valid conclusion, and only 10% of them drew the valid conclusion. Such evidence casts doubt on the generality of the atmosphere effect (see also Dickstein, 1978); we will show later that responses that seemingly corroborate its existence can be given an alternative explanation.

The 'conversion' hypothesis and the figural effect

Another potential source of error is the alleged tendency, emphasized by Chapman and Chapman (1959) and Revlis (1975), to make illicit conversions of premises. There is evidence (see e.g. Woodworth and Schlosberg, 1954) that subjects do fall into the trap of converting symbolic assertions of the form:

All the x's are y's

into:

All the y's are x's.

There is also evidence that subjects reason more accurately with premises that yield the same valid conclusion even if a premise is converted (Revlis, 1975; Revlin and Leirer, 1978) and with premises that explicitly prevent conversion (Ceraso and Provitera, 1971). But, these phenomena are a very different matter from the spontaneous conversion of *all* premises in the process of syllogistic inference. If subjects were automatically to convert every premise—an assumption once adopted by Revlis (1975), though he now holds to it less strongly (personal communication)—then premises in the figure:

A – B
B – C

should be just as likely to elicit a conclusion of the form C – A as premises in the figure:

B – A
C – B

In fact, as we have already mentioned, there is a very marked figural bias. Premises in the figure:

A – B
B – C

tend to yield conclusions of the form, A – C, whereas premises in the figure:

B – A
C – B

tend to yield conclusions of the form, C – A (see Johnson-Laird, 1975; Johnson-Laird and Steedman, 1978). For sixty years, students of the syllogism did not observe the figural effect, because they assumed that medieval Scholastic logic characterized human competence, and because their experimental procedures relied on the evaluation of one or more *given* conclusions. Scholastic logic recognizes only conclusions of the form, C – A, and experimenters failed either to make systematic comparisons with the other form of conclusion, A – C, or to determine what conclusions subjects draw spontaneously.

Syllogistic theories based on Euler circles

Complete theories of performance with syllogisms have only recently begun to be proposed. They have been based on well-known logical techniques for syllogistic inference, and have been intended to account only for the evaluation of given conclusions. The best known logical technique is the method of Euler circles—a geometrical analogy that the mathematician, Leonhard

Euler, used to teach logic to a German princess. The technique was in fact invented by Leibniz; it is often confused with a superior technique known as the method of Venn diagrams (see below). The basic idea of the Euler method is to use circles drawn in the Euclidean plane to stand for sets of entities. Hence, each of the four moods of syllogistic premises can be represented diagrammatically. A premise of the form 'All A are B', requires 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; and, 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' requires four different diagrams: A intersecting B, A included in B, B included in A, and A co-extensive with B. A premise of the form, 'No A are B', calls for one diagram in which the two circles are wholly separate and do not intersect. A premise of the form 'Some A are not B', requires three diagrams: A intersecting B, B included in A, and A and B wholly separate.

In order to make a valid deduction, it is necessary to consider all the different ways in which the respective diagrams for the two premises can be combined: a conclusion is valid if it holds for all the different combinations. The process of checking all the combinations is by no means trivial, because there is no simple algorithm for carrying it out and the total number of combinations is generally greater than the product of the numbers of separate diagrams for the two premises. The reader may care to try to construct the complete set of combinations for premises in the form of the easy problem presented earlier:

Some of the A are B
All the B are C

We have yet to encounter anyone who succeeds in this task. There are 16 different combinations (see below), yet even when one knows the number it is still difficult to find them all.

Wason and Johnson-Laird (1972) made an informal theoretical use of Euler circles, but Erickson (1974) presented a comprehensive theory of syllogistic performance based on them. He assumed that reasoners form representations that are isomorphic to Euler circles, and that they base their conclusions on combined representations of the premises. Certain aspects of the theory are plausible, but it suffers from one severe problem: the large number of different ways in which the diagrams can be combined. Erickson accordingly explored three versions of the theory. The first version assumes that subjects construct all possible representations; this is obviously implausible since it predicts perfect performance. The second version assumes that subjects construct only one of the many possible combinations of the diagrams representing premises. This version is also implausible since it predicts that subjects will always draw a conclusion interrelating the end terms, and never respond "no valid conclusion". The third, and most successful, version accordingly assumes that subjects construct some but not all possible combinations. Erickson does not formulate any comprehensive principles that determine which combinations are constructed. The theory fits his data only on the basis of the estimates of many probabilities concerning both the representation of individual premises and the production of combined representations (see also Erickson, 1978). The theory is also forced to assume that subjects are prey to the atmosphere effect, because there are occasions where an overlap between the circles representing A and C needs to be interpreted as 'Some A are C', and other occasions where the overlap needs to be interpreted as 'Some A are not C'. To invoke the atmosphere effect in order to save subjects' rationality is a paradoxical remedy indeed.

One way in which to make the Eulerian method more tractable is to abandon circles in favour of strings of symbols, and this motive perhaps lies behind the theory developed by Sternberg and his colleagues (Guyote and Sternberg, 1981). Corresponding to a diagram of circle A included within circle B, these theorists postulate a representation equivalent to:

$$
\begin{aligned}
a1 &\subset B \\
a2 &\subset B \\
b1 &\subset A \\
b2 &\subset \bar{A}
\end{aligned}
$$

We have used the standard set-theoretic notation instead of Guyote and Sternberg's own symbols. The first two assertions represent set A included in set B: a1 and a2 are disjoint but exhaustive subsets of A. The second two assertions represent one subset, b1, of B as included within A, and the other subset, b2, as included in its complement, not-A. (In fact, the authors are not entirely clear about the status of the tokens, a1, a2, b1, b2: they refer to them both as subsets and as individual members of the upper-case sets. Since their notation is unequivocally interpreted to mean that a1 is a subset of A, we have followed this usage, in which case, a1 and a2 denote subsets of A, not members of it.) The numbers of different symbolic representations for the four moods are the same as the numbers of different Euler diagrams. The procedure for combining the representations is complicated and calls for four separate steps, which depend upon two rules of inference:

(1) If one set (or subset) is included in a second set, and the second set is included in a third set, then the first set (or subset) is included in the third set.

(2) If one set (or subset) is included in the complement of a second set, and the second set itself is included in a third set, then the first set (or subset) is either included or not included in the third set. (Since this indeterminate conclusion is true of any two sets, it is not clear why the elaborate antecedent conditions are specified.)

The first step is to construct transitive chains of symbols that represent partial combinations of the representations of the premises, which are then condensed into a summary of the possible relations between A and C. The second step is designed to eliminate some of these partial combinations by adding each of them, again using the same two rules, onto the end of the representation of one of the premises. If the result is inconsistent with the other premise, then the partial combination is eliminated. The third step adds the remaining partial combinations onto the *front* of the representation of one of the premises, compares the result with the other premise, and if there is an inconsistency eliminates the partial combination. The fourth step combines the surviving partial combinations into complete ones, and selects a matching conclusion, if any, from the list of those presented.

What is strange about the theory is that having proposed a complicated account of the combination of representations, it does not locate the causes of error in those processes. The theory assumes that subjects represent premises accurately, and that the rules of inference are accurately applied in all four steps. The theory does assume, however, that subjects may go wrong in selecting a conclusion that matches the final combined representation(s). It also assumes that subjects are able to consider at most four of the possible combinations of representations. Guyote and Sternberg are thereby obliged, like Erickson, to invoke the atmosphere effect to save subjects' rationality.

The theories based on Euler circles have problems as accounts of either competence or performance. On the one hand, the theories deny rationality to human reasoners. If people never consider more than four combinations (or some lesser number), then they are irredeemably irrational in making even the simple deduction:

> Some of the A are B
> All the B are C
> Therefore, Some of the A are C.

They may reach the right conclusion but they cannot do so for the right reasons, because they are unable to consider all 16 of the possible combinations. In other cases, they will be unable to deduce the valid conclusion, or will draw a conclusion where none is warranted. On the other hand, the theories are unable to explain actual performance. To take a typical but striking pair of examples, the previous problem is very easy and nearly everyone gets it right, yet it requires 16 different Euler combinations. The following premises from which hardly anyone draws the correct conclusion:

None of the A are B
All the B are C

require only six different combinations. Even if we assume that premises of the form 'All the X are Y' are represented by only one Euler diagram in which X is wholly included within Y (see Erickson, 1978), the position is little better: the easy problem requires six different diagrams and the hard problem requires five. The number of combinations ought to correlate with the difficulty of the problem, but evidently it does not. Similarly, the Eulerian theories contain no machinery accounting for the figural effect on responses. Indeed, Guyote and Sternberg claim that there are never more than two conclusions consistent with the final combinations of the premises. This claim is true only if the subjects' responses are restricted to Scholastic conclusions—a restriction that indeed applies to the materials used in their experiments. A simple counterexample to their general claim is provided by the premises:

All A are B
No B are C

which yield validly any one of the four conclusions:

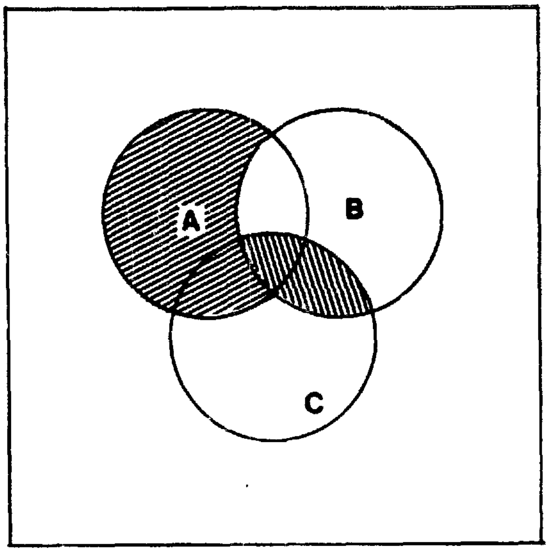
No A are C
No C are A
Some A are not C
Some C are not A

The theories also fail to account for the error of responding 'no valid conclusion' when there is a valid conclusion and, moreover, one that is congruent with the atmosphere of the premises.

Syllogistic theories based on Venn diagrams

Proponents of human rationality will rightly reject any theory of syllogistic inference that is incompatible with valid reasoning. One alternative that they can adopt is a more tractable logical technique, such as the method of Venn diagrams. These diagrams again rely on representing sets as circles, but a single comprehensive diagram is used for the syllogism as a whole (see Fig. 1). The diagram contains three overlapping circles within a square representing the universe of discourse; the circles represent the sets A, B, and C respectively. A premise of the form, 'All A are B', is interpreted by shading out those parts of the A circle that do not overlap B so as to indicate that

Figure 1. *A Venn diagram appropriately shaded to represent the premises 'All A are B', 'No B are C'.*



there are no A's outside the set of B's. A premise of the form, 'Some A are B', is interpreted by drawing crosses in the parts of A that overlap with B: one cross in the part of A that is both B and C, and one cross in the part of A that is B but not C. The crosses indicate that at least one of these two areas is not empty. The premise, 'No A are B', is interpreted by shading out the overlap between A and B to indicate that it is empty. The premise, 'Some A are not B', is interpreted by putting crosses in the parts of A that do not overlap with B: one cross in the part of A that is not B but is C, and one cross in the part of A that is neither B nor C. The crosses indicate that at least one of these two areas is not empty. Once both premises have been interpreted, any valid relation between A and C can be read off from the resulting diagram. Figure 1 presents the Venn diagram for premises of the form:

All A are B
No B are C

The shaded portions represent subsets whose existence is ruled out by the premises. The diagram establishes the validity of the conclusion: No A are C, or its converse.

Newell (1981) has proposed a theory which, in fact, uses strings of symbols to represent the different areas of a Venn diagram and formal rules o combine the strings corresponding to the two premises. The premise, 'All A are B', is accordingly represented by the following string:

Nec A+B+, Pos A−B+, Pos A−B−

which means that there are necessarily A's that are B's (Newell follows Aristotle in taking universals to establish existence), that possibly there are B's that are not A's, and that possibly there are entities that are neither A's nor B's. The fact that the premise renders it impossible for there to be A's that are not B's is represented by the absence of any string concerning A+B−. As Newell points out, this latter convention makes the notational system vulnerable to errors of omission. Newell employs heuristic rules that combine strings to form new strings and that compare the result with a given conclusion. The theory is clearly intended to account for underlying competence in evaluating given conclusions rather than to provide a theory of performance. It makes no predictions about errors, and it is plainly aimed at illustrating how a theory of reasoning can be developed within the framework for studying problem solving devised by Newell and Simon (1972).

There is an alternative and perhaps simpler theory on the same general lines, which depends on the close relation between Venn diagrams and truth tables. If a table of the contingencies concerning sets A, B, and C, is constructed, using a '+' in the table to indicate a set, and a '−' to indicate the complement of the set, then each row in the table corresponds to a different area in a Venn diagram. Syllogistic premises can then be interpreted by eliminating or establishing rows appropriately. For example, the premises:

1. Some A are B
2. All B are C

yield the following interpretation in the table of contingencies for A, B, and C:

	A	B	C	
Premise 1 establishes one or other of these rows {	+	+	+	
	+	+	−	Premise 2 eliminates this row
	+	−	+	
	+	−	−	
	−	+	+	
	−	+	−	Premise 2 eliminates this row
	−	−	+	
	−	−	−	

Thus, the first premise establishes that the overlap of A and B is not empty, and the second premise establishes that the parts of B outside C are empty. The first row in the table is therefore definitely established, and so the conclusion, 'Some A are C', is valid. (We can now establish directly the number of Euler diagrams required to represent these premises. If one ignores the last contingency in the table, which represents those entities that are not members of any of the three sets, there remain four rows that are open, i.e. neither established nor eliminated by the premises. There are $2^4 = 16$ joint possibilities for them, and hence there are 16 different Euler diagrams representing the premises: one diagram for each possibility.)

Syllogistic inference and the predicate calculus

As theories of competence, the trouble with Euler circles, Venn diagrams, and their symbolic equivalents, is that they cannot be generalized to other forms of quantificational inference, such as deductions that depend on a multiple quantification of a single relation within a premise. They simply cannot represent both quantifiers in such deductions as:

> Not all voters hate all politicians.
> Therefore, some politicians are not hated by some voters.

An obvious way in which to capture multiple quantification and a much wider range of inferences is to assume a theory of competence based on the standard first-order predicate calculus. Linguists such as Chomsky (1977), and philosophers such as Davidson (1967), have implicitly adopted such a view for some time. They assume that sentences have a logical form that calls for the apparatus of quantifiers and variables posited by the calculus.

The major difficulty for the predicate calculus as the basis of a theory of competence is the counterintuitive nature of its rules of inference. Of course there are many different ways in which to formalize the calculus, but none of them uses rules of inference that are immediately obvious to a naive individual. The basic function of the rules is to eliminate the quantifiers so that deductions can be made by employing the rules of inference from the propositional calculus (which is a part of the predicate calculus). It is a straightforward matter to eliminate universal quantifiers: the rule of 'universal instantiation' allows a universal quantifier to be replaced by any constant denoting an individual. This rule merely formalizes the notion that if a predicate applies to everything in a universe of discourse, then it applies to any individual in that universe. The problem arises with the rule for eliminating the existential quantifier, 'for at least some x'. The rule of 'existential instantiation' allows the quantifier to be replaced by an arbitrary constant provided that this constant has not occurred elsewhere in the argument. The idea is that if a predicate applies to someone or something, then it applies to a particular entity, and one can assume that it applies, say, to Fred provided that Fred has not already been referred to elsewhere. The concept of an arbitrary constant can be problematic even to students of logic, but the only obvious way to avoid the rule of existential instantiation is to opt for a formalization based on the so-called 'resolution' rule of inference. Unfortunately, such systems require the premises to be translated into a very unnatural uniform disjunctive form in which existential quantifiers are represented by a special function.

Although no psychological theories have been based directly on the predicate calculus, Braine and Rumain (1983) have constructed an ingenious set of inferential schemata that build quantifiers into the required set of propositional rules of inference. Braine and O'Brien (1983) have proposed two rules of inference that could be used to make syllogistic inferences. The first rule, like Guyote and Sternberg's, yields transitivity:

(1) If all (some) of A are B, and all B are C, then all (some) A are C.

The second rule applies to negatives:

(2) If all (some) A are B, and no B are C, then all (some) A are not −C.

We have here expressed the rules informally rather than in Braine and O'Brien's notation. Their theory is primarily an account of rational competence, but it does make some systematic predictions about errors in performance. It accounts for the figural bias in responses on the assumption that the rules apply directly to premises in the first figure since the terms occur in the rules in the same order as they occur in that figure, whereas they do not apply directly to premises in the second figure. Even the adoption of the first-order predicate calculus, however, does not suffice for an adequate account of competence. There are simple inferences that cannot be accommodated within it. For example, the deduction, which we discussed in the introduction:

More than half the artists are beekeepers
More than half the artists are chemists
Therefore, at least one chemist is a beekeeper

cannot be captured in the calculus.

The theories that we have reviewed rely either on formal logic or recognized logical techniques for syllogistic inference. Despite their many virtues, they all have some shortcomings as theories of competence and as theories of performance. A theory of competence should at the very least account for deductions with singly quantified assertions, with multiply quantified assertions such as "not all voters hate all politicians", and with unorthodox quantifiers such as "more than half". Many deductions of these sorts are within the competence of ordinary individuals, but, as we have shown, none of the existing theories can cope with all of them. A theory of syllogistic performance should at the very least account for the the relative difficulty of different forms of syllogism, for the figural response bias, and for the nature of erroneous responses, including those of the type, "there is no valid conclusion (interrelating the end terms)". Not surprisingly given its goals, the atmosphere hypothesis can account only for some errors and not for those of the form "no valid conclusion"; it was not intended to cope with relative difficulty or with the figural effect (which was unknown at the time). The conversion theories certainly account for some errors and for some aspects of the relative difficulty of syllogisms, but they cannot explain either the figural bias or the erroneous "no valid conclusion" responses. The Euler circle theories can account for some aspects of relative difficulty, for some erroneous conclusions, including in Guyote and Sternberg's case certain erroneous "no valid conclusion" responses, but these theories offer no explanation for the figural bias. The Venn diagram theories are not intended to account for performance and thus are mute on questions of error and response bias. The predicate calculus theory is still under active development; it has yet to be used to explain either the relative difficulty of syllogisms or the "no valid conclusion" errors. Since the reader may well have lost track of the details, Table 1 summarizes our review of the strengths and weaknesses of the theories: it shows which aspects of competence and performance each class of theories explains and fails to explain. Before we can develop a better theory, we must obviously elucidate the causes of the 'figural effect'.

2. Experiments on the figural effect

The effect of figure on the form of syllogistic conclusions is extremely reliable and sufficiently robust to be readily demonstrated in lectures and laboratory classes. The crucial question, of course, is: what causes it? In an earlier theory, Johnson-Laird and Steedman (1978) ascribed the effect to a built-in directional bias in the mental representations of premises, which they argued were easy to scan in one direction but difficult to scan in the opposite direction. These authors recognized, however, that the bias might alternatively be a consequence of the mental processes required to form an integrated representation of the premises. The aim of our first experiment was to determine whether the effect would still occur when subjects were given only a relatively short time in which to draw conclusions. If it is created by the processes of constructing a model of the premises, then it should certainly occur in these circumstances. If it only arises as a result of prolonged cogitation, then it should be considerably reduced. In the final analysis, the motivation of the

Table 1. *Theories of syllogistic inference categorized in terms of their basic mechanism and evaluated as accounts of rational competence and syllogistic performance. A '+' indicates that at least one theory within the category copes with the phenomenon, a '–' indicates that no theory within the category copes with it, a '+/–' indicates that at least one theory copes with some of the phenomena, and a blank indicates that the status of existing theories is uncertain.*

	Category of theory				
	Atmosphere hypothesis	Conversion theory	Euler circles	Venn diagrams	Predicate calculus
Competence with:					
1. Syllogisms	–	–	+	+	+
2. Multiple quantifiers	–	–	–	–	+
3. Quantifiers like 'more than half'	–	–		–	–
Performance:					
1. Relative difficulty of syllogisms	–	+/–	+/–	–	
2. Figural response bias	–	–	–	–	+
3. Erroneous responses, including 'no valid conclusion'	+/–	+/–	+	–	

experiment is not crucial, because it revealed an unexpected phenomenon that enabled us to reject one of the potential causes of the effect.

Experiment 1

Method

A subject's task in this experiment was to draw a conclusion from each of a series of sensible syllogistic premises. Each pair of premises was presented twice to every subject. On the first occasion, the subject had 10 sec in which to draw a conclusion or to state that there was no valid conclusion that could be drawn. After the subject had responded to all of the syllogisms in this way, they were re-presented one at a time, together with the subject's initial conclusions. For each problem, the subject now had 1 min in which to revise the earlier judgement, if need be. The main point of this second phase of the experiment was to encourage subjects to make a rapid and intuitive response during the initial phase: they could feel relaxed about these 10 sec responses because they knew they would have an opportunity to revise them. But, we were also interested in whether or not the revisions would yield a greater figural effect.

Each syllogistic premise consisted of a sensible everyday statement of one of the four following sorts:

All of the x's are y's
Some of the x's are y's
None of the x's are y's
Some of the x's are not y's.

There were consequently four choices of form for each premise, and four choices of figure for the premises as a whole, yielding a total of 64 different pairs of premises. Each subject received all 64 possible pairs in a different random order.

Procedure and materials

The subjects were tested individually. They were told that they were going to take part in an experiment on how people combine information in order to draw conclusions from it. They would be given a series of pairs of statements about different groups of people, whom they should imagine as assembled in a room, and they would have to write down what, if anything, followed necessarily from these premises about the occupants of the room. The purpose of this instruction was to isolate the contents of the problems from the subjects' attitudes and expectations. They were also told that their conclusions should be based solely on the information in the premises, and not on plausible suppositions or general knowledge. In the first part of the experiment, they had to draw a conclusion from the presented pair of premises within 10 sec: they were not to worry if this initial response was wrong, because they would have a chance to correct it later. If they considered that there was no conclusion that followed necessarily from the premises, they simply had to write down, 'nothing'. The subjects were not given any instructions about the interpretation of quantifiers, formal logic, or how to reason. There were five three-term series problems as practice trials to familiarize the subjects with the timing of the procedure so as to ensure that there were no late responses. The subjects then received the 64 test trials. There was a pause of 5 min. Finally, each pair of premises was presented again in the same order as before, together with the subjects' previous conclusion. They were allowed up to 1 min in which to revise their earlier response.

The experiment was conducted with native speakers of Italian. The contents of the syllogism were devised so as to minimize semantic relations between the terms within each premise pair while retaining plausibility for any possible conclusion, valid or invalid. This end was achieved by choosing occupations for the two end terms in each problem and an interest or preoccupation for the middle term, for example:

> Nessun magistrato e ornitologo (None of the judges is an ornithologist)
> Tutti i matematici sono ornitologi (All the mathematicians are ornithologists)

and:

> Qualche architetto e vegetariano (Some of the architects are vegetarians.)
> Qualche vegetariano non e notaio (Some of the vegetarians are not notaries.)

Each pair of premises was mimeographed on a separate sheet of paper, and the subjects wrote their responses on these sheets.

Subjects

Twenty volunteers, who were students at the University of Milan, were paid 1500 lire to take part in the experiment. None of the subjects had received any formal training in logic. One subject was rejected half way through the experiment because she claimed she could not make any more inferences since she did not know the particular individuals referred to in the premises. She made it clear that she had been attempting to use her personal experience as the basis for her conclusions (cf. Scribner's (1977) study of syllogistic inference among the Kpelle of Liberia). This subject was replaced by another.

Results

The figural bias in responses was highly reliable in the 10 sec presentation, equally reliable in the subsequent 60 sec presentation, and there was no significant difference between the conditions. Since the results of the first phase of the experiment—the 10 sec presentation—are likely to have strong residual effects on the second phase, we will concentrate on what happened in the first phase. These results for each of the 64 problems are presented, along with other information, in the four tables of the Appendix (Tables 9–11). Table 2 shows the percentages of A – C and C – A conclusions that

Table 2. *The effect of figure on the form of conclusions in Experiment 1: the percentages of A – C and C – A conclusions for each figure. The data in the left-hand columns are from the 10 sec presentation, and those in parentheses from the subsequent 60 sec presentation. The balance of the percentages consist of 'no valid conclusion' responses and the small proportion of erroneous conclusions that failed to contain both end terms.*

	Figure of premises			
	A – B B – C	B – A C – B	A – B C – B	B – A B – C
Form of conclusion				
A – C	53 (65)	17 (19)	21 (31)	21 (26)
C – A	4 (5)	32 (42)	16 (19)	8 (11)

were drawn in 10 sec for the four figures. All 20 subjects showed the expected bias towards A – C conclusions for A – B, B – C premises, and towards C – A conclusions for B – A, C – B, premises ($p = 0.5^{20}$), and the effects were equally marked both for valid and invalid premises. They were also reliable in an analysis by materials: 28 of the 32 relevant problems yielded the predicted bias at both the 10 sec and the 60 sec presentations, and there were only two problems at the 10 sec presentation and three problems at the 60 sec presentation that yielded contrary results (Sign tests, $p < 0.0001$).

Apart from the figural bias, the most obvious phenomenon in the results was the high proportion of 'no valid conclusion' responses that occurred in the initial phase of the experiment. Table 3 presents the percentages of these responses both where they were correct and where they were incorrect. Surprisingly, the proportion of these responses increases significantly over the four figures, with the A – B, B – C figure producing the fewest and the B – A, B – C figure producing the most. This trend over the four figures is reliable both for the 10 sec results (Kendall's $W = 0.447$, $p < 0.01$) and the 60 sec results (Kendall's $W = 0.449$, $p < 0.01$). The data also suggest that the proportion of these responses declines considerably when subjects have the opportunity to revise their responses in the 60 sec condition: 66% of the occasions on which subjects changed their minds were shifts from 'no valid conclusion' to a conclusion. This shift is reliable: 17 out of the 20 subjects yielded fewer 'no valid conclusion' responses in the 60 sec condition and only one subject yielded contrary results (Sign test, $p < 0.0001$), and likewise 52 out of the 64 problems yielded this pattern of results and only two problems yielded contrary data (Sign test, $p < 0.00001$).

Table 3. *The effect of figure on 'no valid conclusion' responses in Experiment 1: the percentages for premises with a valid conclusion relating the end terms (incorrect responses) and for premises without such a valid conclusion (correct responses). The data in the left-hand columns are from the 10 sec presentation, and those in parentheses from the subsequent 60 sec presentation.*

	Figure of premises				
	A−B B−C	B−A C−B	A−B C−B	B−A B−C	Overall
Premises with valid conclusions (incorrect)	30 (20)	38 (24)	51 (38)	59 (48)	46 (34)
Premises without valid conclusions (correct)	55 (36)	57 (46)	70 (58)	86 (80)	67 (54)

In the case of the two symmetrical figures:

A−B B−A
C−B B−C

there was an interesting tendency: where the conclusion was in the same mood as just one of the premises, the end term of the premise tended to play the same grammatical role in the conclusion as it did in the premise itself. For instance, with premises of the form:

All the A are B
Some of the C are B

a conclusion containing the quantifier 'some' tended to take the form:

Some of the C are A

whereas with premises of the form:

Some of the A are B
All the C are B

a conclusion containing 'some' tended to take the form:

Some of the A are C.

The effects of such a bias tend to cancel out over the figure as a whole because of the existence of such complementary pairs. Nevertheless, there was evidence for the effect in both the 10 sec and the 60 conditions. Since the two conditions yielded similar data, we will consider only the 10 sec case: 64% of the conclusions were in accordance with the bias, 24% went against it, and the remaining conclusions were in moods that did not correspond to either premise (only one subject yielded results that ran counter to the bias, and there were two ties, Sign test, $p = 0.0001$). The bias was, in fact, reliable only for the A – B, C – B figure (with no subject yielding results contrary to it). In the case of the B – A, B – C figure, there was an overriding tendency to draw conclusions of the form A – C: half the subjects drew no more than one C – A conclusion for the 16 problems in this figure.

Finally, there was a type of error that occurred sufficiently often to be worth reporting. If one of the premises was of the form, 'Some of the X are not Y', then several subjects would draw an affirmative conclusion containing the quantifier 'some' (see Appendix). This response can be readily explained as the result of the subjects taking 'Some of the X are not Y' to imply that some of the X *are* Y—an invited inference, which though logically unwarranted, is highly plausible in everyday life (see Grice, 1975).

Discussion

The experiment confirmed the existence of a figural bias even when premises are presented only briefly. The most striking finding, however, was that the proportion of trials on which the subjects initially drew no conclusion was affected by the figure of the premises (see Table 3). This effect was unexpected. It is worth noting that the symmetric figure B – A, B – C has more premises yielding valid conclusions than any other figure, and yet it elicited the greatest proportion of 'no valid conclusion' responses. A natural explanation of these responses is that subjects are having difficulty in constructing any model of the premises and that they therefore respond that there is no valid conclusion. The fact that the difficulty increases over the four figures forces us to reject Johnson-Laird and Steedman's (1978) hypothesis that the effect of figure arises from a directional asymmetry in the mental representations of the premises. This hypothesis cannot explain the relative difficulty of forming initial models, since it postulates effects that can come into play only after a model has been constructed.

The results suggest an alternative hypothesis: the figural effect arises from the process of integrating the premises within working memory. It follows from this hypothesis that effects of figure should not be unique to syllogisms but should also occur with other sorts of inference, including relational inferences and three-term series problems, such as:

Anna is taller than Bertha
Bertha is taller than Carol.

Our second experiment was designed to test this prediction.

Experiment 2

Previous studies of three-term series problems have invariably employed techniques in which subjects are either given a specific question to answer or else asked to evaluate a specific conclusion (see e.g. Hunter, 1957; De Soto *et al.*, 1965; Huttenlocher, 1968). We therefore decided to investigate what happens when subjects have to draw conclusions in their own words. In order to obviate the problem of the 'markedness' of certain comparative terms (see Clark, 1969), we elected to use problems employing a single relational expression, 'is related to', denoting kinship. This relational expression serves as its own converse. Since these problems are very much easier than syllogisms, we did not expect figure to play so powerful a role in inference because subjects should usually be able to integrate the information from the second premise into their model of the first premise (cf. Hunter, 1957). Hence, there should be an overall bias towards conclusions of the form A – C, particularly with the symmetrical figures. But, with premises such as:

A is related to B
B is related to C

there should be an increased bias towards the conclusion:

A is related to C

whereas with premises such as:

B is related to A
C is related to B

the bias should be reduced and subjects should more often tend to conclude:

C is related to A.

We also systematically manipulated the mood of the premises (affirmative or negative) in order to detect whether, as in the previous experiment, when only one premise was in the same mood as the conclusion drawn by a subject, its end term played the same grammatical role in the conclusion as in the premise, e.g.:

B is not related to A
B is related to C
Therefore, C is not related to A.

Design

The subjects drew spontaneous conclusions from 16 pairs of simple relational premises: 4 figures × 4 moods. The figures consisted of the four possible arrangements of the terms in the premises:

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

The moods consisted of the four possible combinations of affirmative and negative premises:

Affirnative	Affirmative	Negative	Negative
Affirmative	Negative	Affirmative	Negative

Each subject received the 16 problems in a random order with the constraint that each block of four problems contained one instance of each figure and one instance of each mood.

Materials and procedure

Twenty-four male and 24 female first names, all of two syllables, were selected from a list of commonly used names. They were divided into triplets of names of the same gender with no names within a triplet having the same initial letter. These triplets were then assigned at random to the 16 basic sorts of inference in which each premise has the form, 'X is related to Y', or 'X is not related to Y'.

The subjects were told that the experiment was about the way in which people combine separate pieces of information. Their task was to say what, if anything, followed necessarily from pairs of statements. All the statements would be about how people are related to one another, and the subjects were to imagine that simple relations, such as brother or sister, were involved. Finally, the subjects were told that they could take as long as they liked to make their responses.

Subjects

Ten students at the University of Sussex were paid 50p to participate in the experiment.

Results

Nearly all the responses to the problems in the mood with two negative premises were of the form, 'no valid conclusion', and our analysis of the results is accordingly confined to the 12 problems in the other three moods. All but 8% of the responses to them took the form, 'X is (not) related to Y' or 'X and Y are (not) related', and for the purposes of scoring we have simply used the order in which the two names occurred as the dependent variable. Table 4 presents the percentages of the forms of conclusion for each figure of the premises. It is evident that there was a general bias towards A – C conclusions, and that A – B, B – C premises enhanced the bias, whereas B – A, C – B premises eliminated it. Likewise, there was the expected effect of mood in the symmetric figure A – B, C – B: there was a bias towards A – C conclusions except when the second premise was negative, and therefore the same mood as the conclusion, in which case there were 60% C – A conclusions and only 30% A – C conclusions. The results from the other symmetric figure, B – A, B – C, however, failed to yield the appropriate switch from A – C to C – A conclusions when the first premise was negative—a pattern that was similar to the results of Experiment 1. Overall, 59% of the conclusions were in accordance with the biases predicted by the theory and

Table 4. *The effect of figure on the form of conclusions to the three-term series problems in Experiment 2: the percentages of A – C and C – A conclusions for the 12 problems in each figure that yielded conclusions interrelating the end terms.*

	Figure of premises			
	A–B B–C	B–A C–B	A–B C–B	B–A B–C
Form of conclusions				
A–C	77	47	50	70
C–A	23	43	33	27

33% were not, and the difference was reliable both by subjects (Wilcoxon's $T = 2.5$, $n = 8$, $p < 0.025$) and by materials (Wilcoxon's $T = 13.5$, $n = 12$, $p < 0.025$). This pattern of results has been confirmed in a further study carried out by our colleague, Patrizia Tabossi (personal communication).

Discussion

The results confirmed that figural effects do occur in three-term series problems. There was an overall bias towards A – C conclusions, which suggests that subjects were generally able to build up a model of the premises in the order in which they occurred. This bias was enhanced in the A – B, B – C figure, but eliminated in the B – A, C – B figure.

The figural effect has implications for the various theories of relational inference. It provides, for example, an alternative explanation for some of the phenomena allegedly caused by the preference for working downwards in the mental construction of vertical arrays (cf. De Soto *et al.*, 1965). A theory of three-term series problems must allow for the relative difficulty of combining information from premises in the different figures—a principle that was anticipated by Hunter (1957). However, we shall not pursue these implications here, since our immediate goal is to construct a theory of syllogistic performance that explains the figural effect.

3. A theory of syllogistic inference

Syllogisms call for a special case of the deductive theory based on mental models. In this section, we will spell out a theory of the mental processes underlying syllogistic inference and the likely causes of error within them. This theory of performance extends and significantly modifies its precursors (see Johnson-Laird, 1975; Johnson-Laird and Steedman, 1978; Johnson-Laird, 1983); we will describe it in terms of an algorithm implemented in LISP-80. The general theory of deductive competence, which was described in the introduction, is based on three main steps: the interpretation of the premises as a mental model, the formulation of an informative conclusion, and the search for an alternative model of the premises that refutes the conclusion. We will deal with each of these steps in detail and with how errors may arise in each of them.

Step 1: The interpretation of premises

Mental models represent finite sets by finite sets of mental tokens, and they can accordingly be mapped one-to-one onto the states of affairs that they represent. However, Euler circles, Venn diagrams, semantic networks, and formulae in the predicate calculus, have no such direct mapping onto the states of affairs that they represent. In particular, Euler circles and Venn diagrams represent finite sets in terms of non-denumerable infinities of points in the Euclidean plane, and semantic networks and the predicate calculus represent descriptions in a way that is very remote from the structure of states of affairs, i.e. they are mapped into syntactically structured strings of symbols. A mental model of the assertion:

All the accountants are pianists

contains a set of tokens that corresponds to the set of accountants, a set of tokens that corresponds to the set of pianists, and a set of identity relations between the tokens that corresponds to identities between the entities:

```
accountant =  pianist
accountant =  pianist
             0 pianist
```

Beyond this isomorphism, we make no strong assumptions about the way in which the information is specifically represented. It is doubtful whether anyone will ever know much about such matters. In our computer program, each line in the diagram above corresponds to a list, and a list of all the lists represents the premise. In the interests of legibility, we omit here and throughout the paper the parentheses demarcating the lists. The numbers of tokens are, of course, arbitrary in that they can be recursively revised if need be. The zero sign is a symbol used to indicate that it is uncertain whether or not the relevant individual—here, a pianist who is not an accountant—exists. There is no need for subscripts, since different tokens can be taken to represent different individuals unless there is an identity link between them.

As the example illustrates, it is parsimonious to represent the fact that there may be certain individuals—pianists who are not accountants—by introducing special tokens to stand for them, since in this way *one* model captures the content of the assertion in contrast to the need for two separate Euler diagrams, one representing set A included in set B and the other representing the two sets as co-extensive. The mental model does not lead to a combinatorial explosion, because anything that is predicated of pianists will apply to pianists in the uncertain category, too, and will not increase the number of possibilities. Mental models can therefore directly represent these referential indeterminacies in a way that is computationally tractable, i.e. there is not an exponential growth in complexity.

The theory assumes that reasoners construct mental models of the states of affairs described by premises. These models may take the form of vivid images or they may be largely outside conscious awareness. What is important about them, as far as the theory is concerned, is their underlying structure. Each list corresponds to a separate individual, and the order of elements in a list plays no semantic role but is relevant to the order in which the elements are processed. A model of a universal affirmative assertion, such as the one above, has the following structure:

```
All of the X are Y:   x =  y
                      x =  y
                          0y
```

where the numbers of tokens are arbitrary, and the zero represents an entity that may or may not exist. The presence of the definite article in this assertion implies that X's definitely exist. An assertion of the form, 'All X are Y', is often taken to have no such existential implication, e.g. 'All deserters will be shot' can be true even if there are no deserters. Such an assertion can be represented by the following model:

```
All X are Y:              0x =  y
                          0x =  y
                               0y
```

The representations of the other sorts of assertion that occur in syllogisms are straightforward:

```
Some of the X are Y:       x =  y
                          0x   0y

None of the X are Y:       x
                           x
                          --------
                                y
                                y
Some of the X are not Y:   x
                           x
                          --------
                          0x    y
                                y
```

The notation for negation is simple: it can be interpreted as a barrier between tokens that prevents them from being identified. The model for 'None of the X are Y' accordingly ensures that no x is identical to any y, and *vice versa*, and that no revision of the model can occur in which such an identity is established. The optional x in the representation of 'Some of the X are not Y' is on the same side of the barrier as the y's and could accordingly be identical to one of the y's. Where a premise asserts definite information about an entire set of entities (i.e. there are no optional tokens), as in the Y term of 'Some X are not Y', the relevant term is 'distributed', to use the traditional terminology. For instance, in the premise, 'Some of the pupils are not team members', the term 'team members' is distributed because there are *no* team members who are identical to certain pupils.

Once one premise of a syllogism has been interpreted, it is possible to interpret the other premise and to form an integrated model of both premises. There are at least two ways in which this process could occur: separate models of the two premises could be constructed and then combined to form a single integrated model; or, alternatively, a model of one premise could be constructed and then the information from the other premise could be added directly to it. In either case, the hinge on which the integration depends is, of course, the 'middle' term which occurs in both premises. There are no obvious empirical consequences that distinguish between these two procedures. We shall assume that information from one premise is directly added to a model of the other premise, but the whole of the following theory could be re-expressed in terms of the combination of two separate premises.

In his *Prior Analytic*, Aristotle argued that syllogisms of the form:

All A are B
All B are C
Therefore, All A are C

are perfect, because their validity is intuitively obvious and requires no further argument (see Kneale and Kneale, 1962, p. 67; Lear, 1980, p. 3). Aristotle dealt with other syllogisms and other figures by showing how they could be transformed into perfect syllogisms using logical procedures that included the conversion of premises. William James (1890) made a similar point, arguing that it is easy to integrate two relational assertions of the form:

A is related to B
B is related to C

because the two occurrences of the middle term are contiguous; and Hunter

(1957) introduced two mental operations—the conversion of premises, and their mental re-ordering—to bring the two occurrences of the middle term into contiguity for premises in other figures. This idea can be extended to deal with the integration of information in mental models. Our theory makes three main assumptions about these processes.

First, working memory constrains inferential performance. It has a limited processing capacity and there is consequently some difficulty in forming within it a representation that integrates the two premises. The problem is not simply one of remembering what the premises are: it is not solved even if the subjects have the premises in front of them throughout. The crux is that both premises must be represented in working memory simultaneously in order to integrate them (unless the subjects use paper and pencil or some other external aid). The mental model corresponding to one premise must accordingly be retained in memory long enough to allow information from the other premise to be incorporated within it. Information in memory, however, tends to fade away—whether as a result of interference, decay, or some other factor, is a matter that need not concern us here. If part of a mental model does disappear, then it can be refreshed only by re-reading and re-interpreting the premise.

Second, the first information into working memory tends to be the first information out of it. This 'first in, first out' principle explains why, for example, it is easier to recall a list of digits in the order in which they were presented than to recall them in the opposite order (see Broadbent, 1958, p. 236). The same principle should apply to making inferences, and the natural order in which to state a conclusion is the order in which its terms entered memory.

Third, there is likely to be a preference for constructing a mental model of the first premise to be presented and then integrating within it the information from the second premise. Where the figure of the premises makes it difficult to effect an immediate integration, other operations have to be carried out, and premises that require these operations will be harder to integrate because of the increased load on working memory. There are two operations that are carried out to aid the integration of premises:

(1) The interpretation of a premise can be renewed. In this way, it is possible to build a mental model of the second premise and then by renewing the interpretation of the first premise to add its information to the model. This manoeuvre, in effect, re-orders the premises.

(2) The interpretation of a premise can be switched round in the 'cognitive workspace' of working memory so that a premise of the form B – A takes on the form A – B.

The second operation should not be confused with the verbal conversion of a premise. The converse of 'Some A are B' is 'Some B are A' and they are equivalent in that when one is true the other is true; the converse of 'All A are B' is 'All B are A' and they are not equivalent. If reasoners invariably formed the converse of premises, they would often fall into error (see Chapman and Chapman, 1959; Revlis, 1975). The notion of switching round an interpretation concerns only the order of the accessibility of information in working memory. The interpretation of 'All A are B' takes the form:

```
a =  b
a =  b
    0b
```

where the a's are the first items into working memory. If this interpretation is switched round, it takes the form:

```
 b =  a
 b =  a
0b
```

which is logically accurate, but the b's will be accessible prior to the a's. We assume, however, that there is some possibility that the optional b that is not an a may be forgotten in the process of switching round an interpretation. This omission yields the representation:

```
b =  a
b =  a
```

which is inaccurate and may lead reasoners to assume that 'All B are A'. Such errors should occur primarily when reasoners are forced to switch round an interpretation of a premise in order to form a unified model.

We can illustrate the two sorts of operation by considering inferences in each of the four figures.

With premises in the A – B, B – C figure, the two instances of the middle term occur one after the other, and it is a straightforward matter to construct a model of the first premise and then to integrate within it the information from the second premise. Since the a's will precede the c's into working memory, the 'first in, first out' principle leads to a conclusion of the form A – C. For example, given premises of the form:

Some of the A are B
All the B are C

a reasoner can construct a model of the first premise:

```
 a =  b
0a   0b
```

and then add the information conveyed by the second premise:

```
 a =  b =  c
0a   0b =  c
          0c
```

At this point, step 1 is complete, and a mental model representing the pre-ises has been constructed.

With premises in the B – A, C – B figure, the two instances of the middle term do not occur one after the other but are separated in time, and cannot be immediately integrated. The simplest way to proceed is to construct a model based on the second premise, C – B, to renew the interpretation of the first premise, B – A, and then to effect the integration. The 'first in, first out' principle leads to a conclusion of the form C – A. Since the procedure is slightly more complex than that for the previous figure, it should take slightly longer, place a slightly greater load on memory, and therefore lead to an increase in errors.

With premises in the A – B, C – B figure, there are two possible routes to integration. Subjects can construct a model of the first premise, A – B, switch round their interpretation of the second premise from C – B to B – C, and then integrate its information about C. Alternatively, they can construct a model of the second premise, C – B, renew their interpretation of the first premise, A – B, switch it round to B – A, and then integrate its information about A. On the plausible assumption that switching round an interpretation is a more complex operation than merely renewing an interpretation of a premise, then the present procedures are more complex than those required for the previous figure, since both of the present routes call for switching round an interpretation.

The difficulty should be still greater with the final figure, B – A, B – C, since an interpretation has to be switched round in order to construct the *initial* model, i.e. a complete model of a premise has to be manipulated. There are again two possible routes. Subjects can construct a model based on the second premise, switch it round from B – C to C – B, renew their interpretation of the first premise, B – A, and integrate its information. Alternatively, having read both premises, they can renew their interpretation of the first premise, B – A, switch round its interpretation to A – B, renew their interpretation of the second premise, B – C, and integrate its information.

The complexity of the operations required to integrate the premises increases over the four figures. Hence, the proportion of correct valid conclusions should decline over the four figures with a correlated increase in the number of "no valid conclusion" responses and errors.

Step 2: The formation of informative conclusions

An important though often neglected fact is that ordinary reasoners spontaneously attempt to formulate conclusions that maintain the semantic content of premises and that establish relations between terms not explicitly linked in them. If the premises establish a relation between A and B, and a relation between B and C, then they try to draw a conclusion that relates the 'end' terms, A and C. Thus, competence goes beyond logic, since logic sanctions any valid conclusion including conclusions that are not informative in this way. In the case of syllogisms, there are only a limited number of relations that can hold between the end terms, and subjects drawing spontaneous conclusions hardly ever depart from them (see Johnson-Laird and Steedman, 1978). The formulation of an informative conclusion can be explained by two principles. First, in forming an initial model, reasoners are guided by the heuristic of trying to maximize the greatest number of different roles on the fewest number of individuals. Second, they derive a conclusion by scanning the model and establishing what relation, if any, holds between each of the end tokens. The middle term therefore tends not to be referred to in the conclusion.

The theory distinguishes four possible relations in a mental model between a token of one end term and a token of the other end term:

(1) There are positive links between them:

a = b = c.

(2) They are completely separated by one or more 'impenetrable' negative barriers, as in these examples:

```
a = b        a
a = b        a
   0b        ---------
-----------        b
         c         b
         c   ---------
                      c
                      c
```

(3) They are separated by a 'penetrable' negative barrier, i.e. one that has members of the same class (either an end term or the middle term) on both sides of it:

```
 a               a
-----------    ----------
0a   b = c     0a    b
     b = c     ----------
        0c                c
                          c
```

The consequences of penetrability as defined in this way will be explained presently.

(4) They are in an indeterminate relation in that they are neither linked positively nor separated by a negative barrier.

These four possibilities are mutually exclusive and exhaustive, though obviously there may be different relations between different pairs of end terms. The formulation of a conclusion depends on the nature of the links between all the end terms, but the principles are intuitively obvious:

(1) If all the links from tokens of A to tokens of C are positive, the conclusion has the form:

All the A are C.

(2) Otherwise, if there is at least one positive link from a token of A to a token of C, the conclusion has the form:

Some of the A are C.

(3) If all the tokens of A are separated by at least one impenetrable negative barrier from the tokens of C, the conclusion has the form:

None of the A are C.

(4) If the negative barrier is penetrable (as defined above), then the conclusion has the form:

Some of the A are not C.

(5) Finally, if there are only indeterminate relations between the end tokens, then there is no conclusion that can be drawn interrelating them.

These principles yield, as the output of step 2, the maximally informative conclusions consistent with the models. The process of formulating an informative conclusion should be relatively error-free in the case of many mental models. But, where there is a load on working memory, then we may expect errors to occur. In particular, subjects may fail to consider conclusions that run counter to the 'first in, first out' principle of working memory. This failure to scan models in both directions will lead to erroneous "no valid conclusion" responses in cases where there is in fact a valid conclusion of the form, 'Some of the C are not A'.

Step 3: The search for alternative models of the premises

The theory assumes that reasoners make deductions, not by employing mental rules of inference, but rather by searching for alternative models of the premises that render putative conclusions false. Some premises are compatible with only one model. Our earlier example of premises of the form:

Some of the A are B
All the B are C

yields the model:

```
 a =  b =  c
0a   0b =  c
          0c
```

Step 2 produces the conclusion:

Some of the A are C

if the model is scanned in the 'first in, first out' direction. It produces the conclusion:

Some of the C are A

if the model is scanned in the opposite direction. Any attempt to sever the positive link from a to c violates the meaning of the premises, and hence there is no alternative model of the premises. The reader might suppose that there are alternative models, such as:

a =	b = c		a =	b = c
	0b = c		a =	b = c
0a	= c		0a	0b = c
			0a	= c

Such models, however, are not genuine alternatives to the original, because they still support the same conclusions. These examples bring out the important point that the search for alternatives aims to establish that a putative conclusion is false. The process is accordingly based on an initial model, and seeks either to break positive identities or to bypass negative barriers. There is no point in merely changing the number of tokens of a particular sort: a model is only a genuine alternative if it falsifies a current conclusion.

Quite how people search for alternative models is difficult, if not impossible, to determine. We do not suppose that they search either randomly or in a totally systematic way. A random search could indeed be highly effective. For example, you could randomly reorganize the model, check to determine that it was still consistent with the premises, and, if so, then check whether it was still consistent with the conclusion. If it was not, then you would know that the inference was invalid. If it was consistent with the conclusion, then you could make another random reorganization, and so on, until you had tried all possible reorganizations consistent with the premises. Such a procedure would work, and would yield only valid conclusions if it was pursued until all possible rearrangements of the model had been sampled: there are only a finite number of them. This procedure, however, would be very wasteful. At the other extreme, you could follow a totally systematic set of rules, perhaps akin to the semantic methods postulated by Beth (1971) in his formal system. However, the variation in performance, both between and within subjects, suggests that ordinary individuals follow no such regime. In fact, what is crucial to our theory is not the nature of the search for counterexamples, but the number of genuinely different models that have to be constructed in order to draw the correct conclusion. The principal consequence of failing to carry out the search exhaustively, and thus of failing to consider all of these models, is of course an invalid conclusion of a predictable type.

In developing our computer implementation of the theory, we have used a method that depends on five procedures for constructing alternative models:

(1) A procedure that can break positive links. If a model contains an isolated optional token of the middle term, 0b, as in:

```
a =  b =  c
a =  b    0c
     0b
```

the procedure severs the positive link between the end terms and establishes a new link:

```
a =  b
a =  b    0c
     0b = c
```

(2) A procedure that shifts isolated optional tokens of end items round negative barriers. For example, the premises, 'Some of the A are B, None of the B are C' produce the model:

```
 a =  b
0a    0b
-----------
          c
          c
```

which yields the conclusion, 'None of the A are C' or its converse. The procedure shifts the isolated 0a round the barrier:

```
 a =  b
      0b
-----------
0a        c
          c
```

to yield a model that is still consistent with premises but that no longer supports the same conclusions. It yields instead: 'Some of the A are not C' or its converse.

(3) A procedure that adds more optional end tokens of the same type to be found in a model. Thus, given the previous model, this procedure adds a further 0a to yield:

```
  a =  b
      0b
 ----------
0a        c
0a        c
```

and this refutes the conclusion, 'Some of the C are not A', though the converse conclusion, 'Some of the A are not C' remains unscathed.

(4) A procedure that swaps optional and non-optional middle tokens around a negative barrier. Hence, given the model for the premises 'All of the A are B, Some of the B are not C':

```
  a =  b
 ----------
0a =  b   c
     0b   c
```

this procedure yields the model:

```
     0b
 ----------
0a =  b   c
 a =  b   c
```

(5) Finally, if a model contains two negative barriers, there is a procedure that shifts the tokens of one end term round to the same side as the tokens of the other end term. This procedure applied to the model:

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

yields:

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

Obviously, we do not wish to defend the detailed psychological reality of these procedures. The crucial point is the number of models that have to be constructed by them in order to draw the correct conclusion. Some premises require only one model to be constructed: the test procedures cannot yield any other model of the premises. Other premises require the construction of more than one model. For example, premises of the form:

All of the A are B
some of the B are C

are interpretable in two distinct models:

```
    (1)               (2)
a =  b =  c       a =  b
a =  b    0c      a =  b    0c
     0b                0b =  c
```

The first model yields the conclusion, 'Some of the A are C', or its converse. The second model rules out both of these conclusions, and so there is no valid conclusion interrelating A and C. Still other premises require three models to be constructed, e.g. premises of the form, 'Some of the A are B, None of the B are C'. There are no syllogisms that require more than three models to be constructed.

If reasoners carry out all three inferential steps correctly, then their responses will be entirely rational. However, since the search for alternative models is likely to place a considerable load on working memory, the greater the number of models to be considered, the harder the task should be.

We have now specified the interpretations of the four sorts of premises as mental models, the way in which the information from the second premise is incorporated within the models, the four principles for assessing the nature of the link between a pair of end tokens, the five principles for drawing an informative conclusion on the basis of these links, and the five principles for forming alternative models of the premises. It is not obvious that these principles suffice to generate the set of alternative models for any syllogistic premises, and still less is it obvious *what* the set of models is for any arbitrary premises. However, the principles have been implemented in our most recent computer program (written in LISP-80) and the models that the program produces for each of the 64 forms of syllogistic premises are presented in the tables in the Appendix. The models are shown together with the correct conclusions that they should yield if they are all constructed and with the incorrect conclusions that they should yield if some of them are neglected. The tables also include the frequencies with which each conclusion was drawn in Experiment 1 (10 sec condition) and in Experiment 3, to which we now turn. The reader will note that nearly all of the predicted errors are made by subjects, and that very few of the responses in the two experiments fall outside the predicted ones.

4. An experimental test of the theory of performance

Two principal factors should affect the difficulty of a syllogism: the figure of the premises, which can make it hard to construct an initial model (in step 1) and can bias the order in which a conclusion is formulated (in step 2), and the number of mental models that have to be constructed, which can place an additional load on working memory (in step 3). Experiment 3 was designed to test these two predictions and the detailed predictions about erroneous conclusions summarized in the tables in the Appendix.

Experiment 3

Method

The subjects attempted to make an inference from each of the 64 possible pairs of premises presented with a sensible content similar to the materials used in Experiment 1. The subjects were tested individually and received the problems in a random order. Their responses were timed but, unlike Experiment 1, there was no time pressure and the subjects could take as long as they liked to draw a conclusion.

Subjects

Twenty volunteers, who were students at the University of Milan, were paid 1500 lire to participate in the experiment. None of them had received any formal training in logic, and none of them had taken part in such an experiment before.

Results

The results for each of the 64 problems are presented in the tables in the Appendix. The figure of the syllogism once again exerted a massive bias on the form of the conclusions. Table 5 shows the percentages of conclusions that were drawn in the A – C and C – A forms for the four figures. All twenty subjects showed the predicted bias towards A – C conclusions for the A – B, B – C figure, and towards C – A conclusions for the B – A, C – B figure ($p = 0.5^{20}$), and the effects were equally marked for valid and invalid conclusions. They were also reliable in an analysis by materials: 31 of the 32 relevant syllogisms yielded the predicted bias and there was one tie (Sign test, $p < 0.0001$). In the case of the symmetrical figures, if only one premise agreed with the mood of the conclusion, then, as before, there was a reliable tendency for its end term to play the same role in the conclusion as it did in the premise. For instance, with premises of the form:

All the A are B
Some of the C are B

if a subject drew a conclusion containing the quantifier, 'some', then it would tend to take the form:

Some of the C are A.

Fifty-five percent of conclusions in the symmetric figures conformed to the bias and 25% ran counter to it (Wilcoxon's $T = 14.5$, $n = 18$, $p < 0.005$).

Table 5. *The effect of figure on the form of conclusions in Experiment 3: the percentages of A – C and C – A conclusions for each figure. The balance of the percentages consist of 'no valid conclusion' responses and the small proportion of erroneous conclusions that failed to include both end terms.*

	Figure of premises			
	A – B B – C	B – A C – B	A – B C – B	B – A B – C
Form of conclusions				
A – C	78	10	28	28
C – A	6	63	23	17

Table 6. *The percentages of valid conclusions drawn in Experiment 3 as a function of the figure of the premises and the number of models to be constructed (n = the number of problems in each condition).*

	Figure of premises				
	A – B B – C (n = 6)	B – A C – B (n = 6)	A – B C – B (n = 6)	B – A B – C (n = 9)	Overall percentage (n = 27)
Number of models to be constructed					
1 model (n = 11)	90	83	72	43	72
2 models (n = 4)			30	20	25
3 models (n = 8)	30	30	3	8	12
3 models contrary to figure (n = 4)	3	3			3
Overall percentage (n = 27)	51	48	35	22	37

The effect was more pronounced for the A – B, C – B figure: 61% of conclusions conforming to it and 19% running counter to it (Wilcoxon's $T = 3.5$, $n = 16$, $p < 0.005$). But, although the effect was smaller for the B – A, B – C figure, it was reliable (in contrast to Experiment 1): 48% of conclusions conforming to it and 33% running counter to it (Wilcoxon's $T = 23.5$, $n = 15$, $p < 0.025$).

The theory predicts that two main factors should affect drawing *valid* conclusions: figure and number of models. There are three main sorts of predicted response to premises that yield valid conclusions: correct valid conclusions, incorrect conclusions arising from a failure to consider all possible models of the premises, and erroneous responses of the form, 'No valid conclusion' arising from a failure to scan models in both directions. Table 6 shows the percentages of valid conclusions as a function of figure and number of models. The predicted decline as the number of models to be constructed increases is large and extraordinarily reliable (Page's $L = 274.5$, $z = 5.45$, p is less than one in a million). The decline in valid conclusions over the four figures is also of a considerable size and again extraordinarily reliable (Page's $L = 567$, $z = 5.19$, p less than one in a million). There is no apparent interaction between these two variables, but it cannot be assessed because of the necessarily empty cells and the unequal numbers of problems in the cells.

Table 7 presents the percentages of 'no valid conclusion' responses to

Table 7. *The percentages of erroneous 'no valid conclusion' responses in Experiment 3 (to premises permitting valid conclusions) as a function of the figure of the premises and the number of models to be constructed (n = the number of problems in each condition).*

	Figure of premises				
	A – B B – C (n = 6)	B – A C – B (n = 6)	A – B C – B (n = 6)	B – A B – C (n = 9)	Overall percentage (n = 27)
Number of models to be constructed					
1 model (n = 11)	0	5	25	28	14
2 models (n = 4)			30	23	26
3 models (n = 8)	0	20	48	45	27
3 models contrary to figure (n = 4)	13	15			14
Overall percentage (n = 27)	4	11	34	34	22

problems that in fact have a valid conclusion. As predicted, the percentages show a tendency to increase over the four figures and also with the need to consider more models. These data are obviously not completely independent of the percentages of valid conclusions. For the purposes of analysis, however, we can take into account all the responses to the premises with valid conclusions. It was impossible to use latency as a general measure, because there were too many erroneous conclusions. We therefore used a dependent variable that combined both accuracy and latency. For each subject, we ranked valid conclusions first, then invalid conclusions, and finally 'no valid conclusion' responses (on the grounds that they are further from the truth for these problems than an invalid conclusion). Within these three categories, we ranked the responses according to their latencies. This procedure enabled us to rank each subject's complete performance, which was an important consideration because a subject attempted each problem only once. The effects of the number of models to be constructed were reliable for each of the four figures (Page's L ranged from 263 to 270, with probabilities ranging from $p < 0.0002$ to 0.000002). Likewise, the effect of figure was reliable for one-model problems (Page's $L = 560.5$, $p < 0.00001$), it was impossible to assess for two-model problems because they only occur in two figures, and it was reliable for three-model problems (Page's $L = 539.5$, $p < 0.002$).

Table 8. *The mean latencies (in sec) for the correct valid conclusions for ten of the one-model premises in Experiment 3.*

	Figure of premises			
	A – B B – C	B – A C – B	A – B C – B	B – A B – C
Mood of premises				
AA	9.50	12.21		–
IA	11.45			21.56
AI		13.72		22.59
EA		12.44	17.92	
AE	13.82		19.62	
Means	11.55	12.88	18.74	22.07

Only the one-model problems yielded enough correct conclusions for their latencies to be assessed, and they are summarized in Table 8. The variances were large and correlated with the means, some subjects failed to make a correct response in some figures, and there was one problem that no-one got right. However, we were able to rank the mean correct latencies of 14 subjects for 10 of the problems as a function of figure. The mean ranks for the four figures were: 1.7, 2.3, 2.7, and 3.4 (Page's $L = 387$, $p < 0.0005$). The one problem that defeated everyone was in the figure that was predicted to be most difficult:

All the B are A
All the B are C

The majority of subjects drew a conclusion of the form:

All the A are C

or its converse, which suggests that in swapping round an interpretation of a premise, they dropped an optional item, i.e. they constructed a model such as:

```
a = b =  c
a = b =  c
         0c
```

instead of:

```
   a = b = c
   a = b = c
  0a      0c
```

Premises that do not yield valid conclusions produced an increasing proportion of correct 'no valid conclusion' responses over the four figures: 16%, 33%, 50%, and 66% respectively (Page's L = 573.5, z = 5.09, $p <$ 0.0000002), and this trend bears out our earlier assumption that it becomes progressively harder to form an integrated model as the number of mental operations increases over the four figures. Although in the current version of the theory these problems all require two models to be constructed, they fall into three classes: problems in which both premises are negative (52% correct), problems in which both premises contain the quantifier 'some' (41% correct: the four problems that fall into both of these first two categories have been assigned to the present category), and the remaining problems that have moods with valid conclusions in other figures (18% correct). The greater difficulty of the last category is explicable on the grounds that errors will be made to them if optional items are forgotten as a result of switching round an interpretation.

5. General discussion

The theory of syllogistic performance predicts that errors can occur in any of the three stages of reasoning: in interpreting the premises, in formulating a conclusion, and in searching for counterexamples to it. The overwhelming impression made by our results is that these predictions are corroborated. First, the figure of the premises can create a bias towards formulating one form of conclusion rather than another, and affects the speed and accuracy of interpretation. When a figure requires additional operations in order to form an integrated model of the premises, then it is harder to draw a conclusion, and concomitantly the likelihood of responding that there is no valid conclusion is greater. Second, the greater the number of models that have to be constructed in the search for counterexamples, the harder the task is. Moreover, although we have presented no detailed analyses, the erroneous conclusions that the subjects drew can be largely accounted for in terms of the conclusions that would follow if one or more of the possible models of the premises are neglected or if a model is scanned in only one direction. Thus, for example, the following erroneous conclusion:

> All A are B
> Some B are C
> Therefore, Some A are C

might be taken as evidence in support of the atmosphere effect. Yet it is plainly explicable on the grounds that subjects consider the following model of the premises:

```
 a =  b =  c
0a =  b   0c
     0b
```

and overlook the model that refutes the conclusion:

```
 a =  b
0a =  b   0c
      b =  c
     0b
```

Our theory of deductive competence copes with more than just syllogisms. It accommodates arguments based on multiple quantification, and on quantifiers such as 'more than half' which go beyond the capacity of the first-order predicate calculus. Its instantiation as a theory of syllogistic performance explains the figural response bias and the effect of figure on accuracy and latency. It accounts for the relative difficulty of syllogisms (as a function of figure and number of models), and for the pattern of errors, including conversion errors (optional tokens are dropped in the process of switching round an interpretation) and erroneous 'no valid conclusion' responses (failure to scan models in a direction contrary to the 'first in, first out' principle). In short, it would receive a '+' in every row of Table 1. What, then, are its deficiencies? And what remains to be discovered about syllogistic inference?

In our view, there are two problems for our theory of syllogistic inference. First, in the case of the symmetric figures we need to discover what causes the quantified term in the conclusion to tend to match the term in the one premise, if any, that has the same mood. This 'matching' effect is not large but it is reasonably reliable. Second, we need to improve our understanding of the detailed nature of the processes governing the formulation of conclusions and the search for counterexamples.

These two problems may be related, as can be shown by the following crucial example. The premises:

All the A are B
Some of the C are not B

have two models:

```
 a = b             a = b   0c
0a = b   0c        a = b   0c
    0b                0b
------------      ------------
         c                 c
```

According to our present theory, when the first model is scanned from a's to c's, it yields the conclusion:

Some of the A are not C

and when it is scanned from c's to a's, it yields the conclusion:

Some of the C are not A.

When the second model is scanned from a's to c's it yields the erroneous response:

No valid conclusion

whereas scanned from c's to a's it still yields:

Some of the C are not A

which is the correct conclusion. None of the subjects in either Experiment 1 or Experiment 3 ever drew the conclusion that violates the 'matching' effect, i.e.

Some of the A are not C

and this omission is entirely representative of other premises of the same general sort. Hence, the current procedure for formulating conclusions may not be correct.

The next step of the argument is complicated and may be ignored by readers interested only in the general nature of the theory. Consider how the procedure for formulating conclusions operates when tokens of both end items are not separated by a negative barrier. If there are more tokens of the first end item scanned, say a, than of the second end item, c, then the procedure responds 'Some of the A are not C'. The motivation for this procedure is simple: there are not enough c's on the same side of the barrier to be matched up with all the a's. Only when further optional c's are added in constructing an alternative model will the 'no valid conclusion' response be forthcoming. Thus, in the previous example, the first response (scanning from a's to c's) is 'Some of the A are not C', and the second response (after a further optional c has been added) is 'no valid conclusion'.

It is a fact that whenever there is one isolated optional item anywhere in a model, the procedures for revising the model can always add others. If this principle were recognized by the procedure for formulating conclusions, it would merely respond 'no valid conclusion' whenever it scanned one end item and then found an isolated optional token of the other end item on the same side of a negative barrier. This slight change to the procedure for formulating conclusions has a radical effect, as we discovered by implementing it in a program. The premises in the example still, of course, yield the first model. But, when the procedure scans it from a's to c's, it responds:

No valid conclusion.

When the procedure scans from c's to a's, it responds:

Some of the C are not A

because there is a c separated by a negative barrier from all the a's. The first response, 'no valid conclusion', plainly does not call for the search for an alternative model; and there is no alternative model that falsifies the second response. In short, the premises cease to call for two models, since one model will suffice.

The general effect of the revised procedure for premises with valid conclusions is to reduce all three-model problems to two-model problems, and to reduce all two-model problems to one-model problems. There is not necessarily a concomitant reduction in the explanatory power of the revised account, since the former two-model problems should remain harder than the original one-model problems as a result of the need to scan them in both directions. However, the revised procedure has an effect that we did not anticipate. It goes some way to solving the first of our problems. In the case of our example, the revised procedure predicts the response:

Some of the C are not A

where C has the same grammatical role as in the premise in which it occurs. The converse response, which violates this matching relation and which we failed to observe, is no longer predicted.

There are undoubtedly differences from one individual to another in the way in which they make syllogistic inferences. Our alternative implementations of the theory suggest a way in which some of these differences might be explained. Certainly, the cause of individual differences is a major problem that remains to be solved. Another outstanding issue is the development of syllogistic ability in childhood. Students of reasoning have perhaps been unduly influenced by Piaget's claim that young children are unable to cope with 'all' and 'some' (see Inhelder and Piaget, 1964). His treatment of quantifiers was always somewhat peripheral, since his notion of competence was so closely tied to the propositional calculus.

Deductive competence can, in fact, be subsumed under one general principle: validity depends on the absence of counterexamples. Actual deductive performance, as out results suggest, can also be subsumed under one general principle: the search for counterexamples depends on working memory, and anything that increases the load on its processing capacity is likely to affect accuracy. The latter point has emerged from other studies of inference (see Johnson-Laird and Wason, 1970; Baddeley and Hitch, 1974; Hitch and Baddeley, 1976; Oakhill and Johnson-Laird, 1983). In a recent unpublished study, Oakhill has shown that a simple measure of the processing capacity of an individual's working memory correlates reliably with performance in syllogistic inference (rho = 0.7).

An earlier theory of performance (Johnson-Laird, 1975) explained the figural response bias in terms of an asymmetry in the structure of mental representations, but the possibility that it arose from the *process* that constructed models was a plausible alternative, and spelt out as such by Johnson-Laird and Steedman (1978). This principle, which has been vindicated by our results, is embodied in the theory of performance. The notion of re-ordering and switching round the interpretation of premises is quite explicit in Hunter's (1957) paper on three-term series problems. But the grand ancestor of this line of thought is undoubtedly Aristotle. Where the present theory departs from tradition is in assuming that the operations occur, not on linguistic entities, but on their interpretations as mental models.

The success of the theory of performance strengthens our case that logical competence depends, not on mental rules of inference, Euler circles, or Venn diagrams, but on the ability to interpret premises as mental models, and to search for alternatives that violate putative conclusions. Mental models, unlike some of the logical techniques, can cope with any ordinary sort of deductive inference. Moreover, since we would not wish to deprecate Leibniz's achievement in inventing Euler circles (and failing to get the credit for it), Venn's ingenuity in devising his technique, or Frege's genius in formulating the predicate calculus, we would argue that few people are likely to succeed in re-inventing these methods for themselves without being aware of it. Psychological theories based on these methods accordingly run into the great difficulty of explaining how, without explicit instruction, a person could acquire the complex procedures required to set up and to run them. Logic, in particular, can be thought of as providing a systematic formal procedure that guarantees validity. Our evidence suggests, however, that there is no mental logic for quantifiers: ordinary reasoners appreciate the overriding need to search for counterexamples to putative conclusions, but they have no machinery for making the search in a systematic way, and consequently often lapse into error.

Appendix

Tables 9 to 11 present all 64 pairs of syllogistic premises, the predicted mental models, and the results from Experiment 1 (10 sec condition) and from Experiment 3. Each of the four tables deals with one figure; each cell in a table corresponds to a particular pair of premises and presents the models and predicted responses produced by a computer program implementing the theory. To help the reader, impenetrable negative barriers are drawn as solid lines, and penetrable negative barriers are drawn as dotted lines, though no such explicit representation is used by the program. Correct responses are in capital letters; responses that are contrary to the figural effect are in parentheses; and responses depending on Gricean implicatures are italicized. A '?' indicates a response not produced by the program that was made by more than two subjects; these responses include those that are predicted by the theory if an optional element is dropped. This aspect of the theory has been implemented in a separate program, because, except with AA premises, its main effect is to make certain otherwise predicted responses impossible. We have therefore not included this fe ture of the theory in these tables. The data in the left-hand columns are the numbers of subjects (out of 20) making the responses in the 10 sec condition of Experiment 1, and the data in the right-hand columns are the numbers of subjects (out of 20) making the responses in Experiment 3. Where the totals in a cell fail to sum to 20, the missing data are idiosyncratic conclusions, e.g. the inclusion of a middle term instead of an end term as in 'Some A are B'.

Table 9

Figure:
A–B
B–C

Second premise	First premise: A				First premise: I			
A	a = b = c a = b = c 0b = c 0c	ALL A ARE C (SOME C ARE A)	19 –	19 –	a = b = c 0a 0b = c 0c	SOME A ARE C (SOME C ARE A) ? No valid conclusion	16 – 4	18 – –
I	a = b = c a = b 0a = b 0c 0a = b 0c 0b b = c 0b	Some A are C (Some C are A) NO VALID CONCLUSION	13 3 3	13 1 4	a = b = c a = b 0c 0a 0b 0c 0a b = c 0b	Some A are C (Some C are A) NO VALID CONCLUSION	12 – 8	18 – 2
E	a = b a = b 0b ——— c c	NO A ARE C (NO C ARE A) ? No valid conclusion	11 1 6	14 3 –	a = b a = b a = b 0a = b 0b 0b ——— - - - - - - - - - - - - - - - - c 0a c 0a c c c 0a c	No A are C (No C are A) SOME A ARE NOT C (Some C are not A) No valid conclusion	3 3 1 – 12	8 3 4 – –
O	a = b 0b - - - - - - - - - - - - - - - - 0a = b = c 0a = b c 0b c a = b c	Some A are not C (Some C are not A) NO VALID CONCLUSION *Some A are C*	10 – 7 –	15 – 1 4	a = b 0b - - - - - - - - - - - - - - - - 0a 0b c 0a 0b c a = b c a = b c	Some A are not C (Some C are not A) NO VALID CONCLUSION *Some A are C*	7 – 13 –	13 – 2 2

First premise

E | **O**

E

```
a          a          a      0c
a          a      0c  a      0c
________   - - - - -  - - - - -
   b = c      b = c      b = c
   b = c      b = c      b = c
      0c
```

No A are C	13	16
(No C are A)	2	2
Some A are not C	–	–
(SOME C ARE NOT A)	–	–
No valid conclusion	4	1

O

```
a           a       0c
- - - - -   - - - - - -
0a   b = c  0a   b = c
     b = c  0a   b = c
        0c
```

Some A are not C	5	9
(Some C are not A)	1	1
NO VALID CONCLUSION	8	1
Some A are C	5	4

E

```
a          a          a      0c
a          a      0c  a      0c
________   - - - - -  - - - - -
   b = c      b = c      b = c
  0b  0c     0b         0b
```

No A are C	10	13
(No C are A)	–	1
Some A are not C	–	–
(SOME C ARE NOT A)	–	1
No valid conclusion	9	4

O

```
a           a       0c
- - - - -   - - - - - -
0a   b = c  0a   b = c
     0b  0c      0b
```

Some A are not C	3	8
(Some C are not A)	–	–
NO VALID CONCLUSION	13	4
Some A are C	4	6

E

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

No A are C	11	11
(No C are A)	–	1
NO VALID CONCLUSION	9	5

O

```
a           a       c
- - - - -
0a   b      a       c
            - - - - - -
     b      0a   b
________
        c        b
        c
```

No A are C	2	4
(No C are A)	–	1
NO VALID CONCLUSION	11	2
Some A are C	5	9
? Some A are not C	2	4

E

```
a          a       c
a          a  0b   c
________   - - - - -
   b          b
- - - - -
  0b   c
       c
```

No A are C	6	4
(No C are A)	–	–
NO VALID CONCLUSION	13	8
Some A are C	1	4

O

```
a           a       c
- - - - -
0a   b      a  0b   c
- - - - -   - - - - - -
     0b  c  0a   b
         c
```

Some A are not C	4	15
(Some C are not A)	–	–
NO VALID CONCLUSION	15	1
Some A are C	1	–

Table 10

Figure:
B–A
C–B

Second premise	First premise: A	First premise: I
A	c = b = a c = b = a 0b = a 0a ALL C ARE A 4 15 (SOME A ARE C) – – ? All A are C 13 2	c = b = a c = b 0c = b 0c = b 0b 0a b = a 0b 0a Some C are A 9 15 (Some A are C) 2 4 NO VALID CONCLUSION 8 1
I	c = b = a 0c 0b = a 0a SOME C ARE A 10 19 (SOME A ARE C) 4 1	c = b = a c = b 0c 0b 0a b = a 0c 0b 0a Some C are A 6 14 (Some A are C) 4 1 NO VALID CONCLUSION 9 4
E	c c c 0a c c 0a c 0a ——— - - - - - - - - - - - - b = a b = a b = a b = a b = a b = a 0a No C are A 10 12 (No A are C) 2 6 Some C are not A – – (SOME A ARE NOT C) – – No valid conclusion 7 1	c c c 0a c c 0a c 0a ——— - - - - - - - - - - - - b = a b = a b = a 0b 0a 0b 0b No C are A 5 11 (No A are C) 3 – Some C are not A – – (SOME A ARE NOT C) – 1 No valid conclusion 12 5
O	c c 0a - - - - - - - - - - - - 0c b = a 0c b = a b = a 0c b = a 0a Some C are not A 7 7 (Some A are not C) 1 – NO VALID CONCLUSION 9 3 *Some C are A* 2 8	c c 0a - - - - - - - - - - - - 0c b = a 0c b = a 0b 0a 0b Some C are not A – 3 (Some A are not C) – 1 NO VALID CONCLUSION 17 10 *Some C are A* 1 5

First premise

E

c = b
c = b
0b
———
a
a

NO C ARE A	8	12
(NO A ARE C)	4	2
? No valid conclusion	7	3

c = b   c = b   c = b
0c 0b   0b   0b
———
a 0c   a 0c   a
a   a 0c   a

No C are A	2	3
(No A are C)	4	2
SOME C ARE NOT A	2	6
(Some A are not C)	–	–
No valid conclusion	12	4

c   c a
c   c a
———
b   b
b   b
———
a
a

No C are A	1	6
(No A are C)	7	4
NO VALID CONCLUSION	11	9

c   c a
- - -
0c b   c a
b   0c b
———
a   b
a

No C are A	1	2
(No A are C)	–	–
NO VALID CONCLUSION	15	9
Some C are A	2	5
? Some C are not A	1	4

O

c = b   0b
- - -
0c = 0b a   0c = b a
0b a   c = b a

Some C are not A	8	14
(Some A are not C)	–	–
NO VALID CONCLUSION	9	4
Some C are A	1	1

c = b   0b
- - -
0c 0b a   c = b a
0b a   c = b a
0c

Some C are not A	7	13
(Some A are not C)	1	–
NO VALID CONCLUSION	9	4
Some C are A	2	3

c   c a
c   c 0b a
- - -
b   b
- - -
0b a
a

No C are A	4	4
(No A are C)	–	2
NO VALID CONCLUSION	13	12
Some C are A	1	–

c   c a
- - -
0c b   c 0b a
- - -
0b a 0c b
a

Some C are not A	2	6
(Some A are not C)	1	1
NO VALID CONCLUSION	13	10
Some C are A	2	2

Table 11

Figure:
A–B
C–B

Second premise	First premise A			First premise I		
A	a = b = c a = b a = b = c a = b 0b b = c 0b b = c 0b 0b			a = b = c a = b 0a 0b = c 0a b = c 0b 0b = c 0b		
	All A are C	8	9	Some A are C	12	11
	All C are A	–	2	Some C are A	–	4
	NO VALID CONCLUSION	12	7	NO VALID CONCLUSION	6	3
I	a = b = c a = b 0a = b 0c 0a = b 0b b = c 0b 0c			a = b = c a = b 0c 0a 0b 0c 0a b = c 0b		
	Some A are C	1	–	Some A are C	4	7
	Some C are A	9	10	Some C are A	–	–
	NO VALID CONCLUSION	10	9	NO VALID CONCLUSION	15	11
E	a = b a = b 0b ——— c c			a = b a = b a = b 0a 0b 0b 0b - - - - - - - - - c 0a c 0a c c c 0a c		
	NO A ARE C	3	6	No A are C	–	–
	NO C ARE A	9	8	No C are A	6	6
	? No valid conclusion	8	6	SOME A ARE NOT C	1	–
				Some C are not A	–	–
				No valid conclusion	13	13
O	a = b a = b 0c 0a = b 0c a = b 0c 0b 0b - - - - - - - - - c c			a = b a = b 0a 0b 0c 0b 0c - - - - - - - - - c 0a c		
	Some A are not C	–	–	Some A are not C	–	1
	SOME C ARE NOT A	8	4	Some C are not A	1	5
	No valid conclusion	9	9	NO VALID CONCLUSION	17	6
	Some C are A	–	4	*Some C are A*	–	2

E	O
a a b = c b = c 0b **NO A ARE C** 11 9 **NO C ARE A** 3 5 ? No valid conclusion 6 4	a a 0a b = c 0a b = c b = c 0a b = c 0b 0b **SOME A ARE NOT C** 3 7 Some C are not A – – No valid conclusion 13 3 *Some A are C* 1 8
a a a 0c a a 0c a 0c b = c b = c b = c 0b 0c 0b 0b No A are C 5 7 No C are A – 3 Some A are not C – 1 **SOME C ARE NOT A** 2 1 No valid conclusion 12 5	a a 0c 0a b = c 0a b = c 0b 0c 0b Some A are not C – 4 Some C are not A – – **NO VALID CONCLUSION** 17 9 *Some A are C* 2 3
a a c a a c b b b b c c No A are C 1 1 No C are A – 1 **NO VALID CONCLUSION** 19 18	a a c 0a b a c b 0a b c b c No A are C 1 2 No C are A 5 4 **NO VALID CONCLUSION** 12 11 *Some A are C* 1 1
a a c a a c b b 0c b 0c b c No A are C 1 2 No C are A 2 1 **NO VALID CONCLUSION** 13 9 *Some C are A* 1 6	a a c 0a b 0c 0a b 0c b b c Some A are not C 1 1 Some C are not A – – **NO VALID CONCLUSION** 18 17 *Some A are C* 1 2

Table 12

Figure:
B–A
B–C

Second premise	First premise: A	First premise: I
A	a = b = c a = b = c 0a 0c SOME A ARE C – – SOME C ARE A – – ? All A are C 7 9 ? All C are A 1 4 ? No valid conclusion 12 7	a = b = c 0b = c 0a 0c SOME A ARE C 5 2 SOME C ARE A 4 10 ? No valid conclusion 9 5
I	a = b = c 0a = b 0c 0a SOME A ARE C 10 9 SOME C ARE A 1 4 ? No valid conclusion 9 5	a = b = c ; a = b 0c 0a 0b 0c ; 0a b = c 0b Some A are C 2 3 Some C are A 1 – NO VALID CONCLUSION 17 16
E	a = b / a = b / 0a / ——— / c / c a = b / a = b / - - - - / 0a c / c a = b / a = b / - - - - / 0a c / 0a c No A are C 10 10 No C are A 1 2 SOME A ARE NOT C – – Some C are not A – – No valid conclusion 9 8	a = b / 0a 0b / ——— / c / c a = b / 0b / - - - - / 0a c / c a = b / 0b / - - - - / 0a c / 0a c No A are C – 3 No C are A 3 2 SOME A ARE NOT C – 3 Some C are not A – 1 No valid conclusion 16 10
O	a = b / 0a / - - - - / 0a = b c / c a = b / - - - - / 0a = b c / 0a c SOME A ARE NOT C 7 9 Some C are not A – – No valid conclusion 12 5 *Some A are C* – 3	a = b / - - - - / 0a 0b c / 0b c 0b / - - - - / a = b c / 0a 0b c Some A are not C 1 7 Some C are not A – – NO VALID CONCLUSION 18 11 *Some A are C* – 1

First premise

E	O
a a a 0c a a 0c a 0c b = c b = c b = c b = c b = c b = c 0c No A are C 7 6 No C are A 4 6 Some A are not C – – SOME C ARE NOT A – – No valid conclusion 9 7	a a 0c a 0b = c a 0b = c b = c b = c 0c Some A are not C 2 1 SOME C ARE NOT A 2 – No valid conclusion 13 4 *Some C are A* – 9
a a a 0c a a 0c a 0c b = c b = c b = c 0b 0c 0b 0b No A are C 2 2 No C are A – 2 Some A are not C – – SOME C ARE NOT A – 2 No valid conclusion 17 11	a a 0c a 0b a b = c b = c 0b 0c Some A are not C – 2 Some C are not A – 4 NO VALID CONCLUSION 20 10 *Some C are A* – 2
a a c a a c b b b b c c No A are C 3 1 No C are A – 1 NO VALID CONCLUSION 17 16	a a c a 0b a 0b c b b c c No A are C 2 – No C are A 1 1 NO VALID CONCLUSION 14 16 *Some C are A* – –
a a c a a c b b 0b c 0b c No A are C 2 4 No C are A – – NO VALID CONCLUSION 14 11 *Some A are C* – –	a a c a 0b a 0b c b b 0b c c Some A are not C – 4 Some C are not A – – NO VALID CONCLUSION 20 15 *Some A are C* – –

References

Aristotle. (1968) *Analytica priora et posteriora.* W.D. Ross (ed.), Oxford, Clarendon Press.

Baddeley, A.D. and Hitch, G. (1974) Working memory. In G.H. Bower (ed.), *The Psychology of Learning and Motivation*, Vol. 8. New York, Academic Press.

Barwise, J. and Cooper, R. (1981) Generalized quantifiers and natural languages. *Ling. Phil.*, *4*, 159–219.

Begg, I. and Denny, J.P. (1969) Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning errors. *J. exper. Psychol.*, *81*, 351–354.

Beth, E.W. (1971) *Aspects of Modern Logic.* Dordrecht-Holland, Reidel.

Boolos, G.S. and Jeffrey, R.C. (1980) *Computability and Logic.* 2nd edition. Cambridge, Cambridge University Press.

Braine, M.D.S. and Rumain, B. (1983) Logical reasoning. In J. Flavell and E. Markman (eds.), *Carmichael's Manual of Child Psychology*, 4th edition: *Cognitive Development.* New York, Wiley.

Braine, M.D.S. and O'Brien, D.P. (1983) Categorical syllogisms: a reconciliation of mental models and inference schemas. Mimeo, New York University.

Broadbent, D. (1958) *Perception and Communication.* London, Pergamon.

Ceraso, J. and Provitera, A. (1971) Sources of error in syllogistic reasoning. *Cog. Psychol.*, *2*, 400–410.

Chapman, I.J. and Chapman, J.P. (1959) Atmosphere effect reexamined. *J. exper. Psychol.*, *58*, 220–266.

Chomsky, N. (1977) *Essays on Form and Interpretation.* Amsterdam, North-Holland.

Clark, H.H. (1969) Linguistic processes in deductive reasoning. *Psychol. Rev.*, *76*, 387–404.

Davidson, D. (1967) Truth and meaning. *Synthese*, *7*, 304–323.

De Soto, C.B., London, M. and Handel, S. (1965) Social reasoning and spatial paralogic. *J. Pers. Soc. Psychol.*, *2*, 513–521.

Dickstein, L.S. (1978) Error processes in syllogistic reasoning. *Mem. Cog.*, *6*, 537–543.

Erickson, J.R. (1974) A set analysis of behavior in formal syllogistic reasoning tasks. In R. Solso (ed.), *Theories in Cognitive Psychology: The Loyola Symposium.* Hillsdale, NJ., Erlbaum.

Erickson, J.R. (1978) Research on syllogistic reasoning. In R. Revlin and R.E. Mayer (eds.), *Human Reasoning.* Washington, D.C., Winston.

Evans, J.St.B.T. (1982) *The Psychology of Deductive Reasoning.* London, Routledge and Kegan Paul.

Grice, H.P. (1975) Logic and conversation. In P. Cole and J.L. Morgan (eds.), *Syntax and Semantics, Vol. 3: Speech Acts.* New York, Academic Press.

Guyote, M.J. and Sternberg, R.J. (1981) A transitive-chain theory of syllogistic reasoning. *Cog. Psychol.*, *13*, 461–525.

Hitch, G. and Baddeley, A.D. (1976) Verbal reasoning and working memory. *Q. J. exper. Psychol.*, *28*, 603–621.

Hunter, I.M.L. (1957) The solving of three-term series problems. *Brit. J. Psychol.*, *48*, 286–298.

Huttenlocher, J. (1968) Constructing spatial images: a strategy in reasoning. *Psychol. Rev.*, *75*, 550–560.

Inhelder, B. and Piaget, J. (1964) *The Early Growth of Logic in the Child.* London, Routledge and Kegan Paul.

James, W. (1890) *The Principles of Psychology, Vol. 2.* New York, Holt.

Johnson-Laird, P.N. (1975) Models of deduction. In R.J. Falmagne (ed.), *Reasoning: Representation and Process in Children and Adults.* Hillsdale, NJ., Erlbaum.

Johnson-Laird, P.N. (1980) Mental models in cognitive science. *Cog. Sci.*, *4*, 71–115.

Johnson-Laird, P.N. (1982) Thinking as a skill. *Q. J. exper. Psychol.*, *34A*, 1–29.

Johnson-Laird, P.N. (1983) *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness.* Cambridge, Cambridge University Press. Cambridge, MA.: Harvard University Press.

Johnson-Laird, P.N. and Steedman, M.J. (1978) The psychology of syllogisms. *Cog. Psychol.*, *10*, 64–99.

Johnson-Laird, P.N. and Wason, P.C. (1970) Insight into a logical relation. *Q. J. exper. Psychol.*, *22*, 49–61.

Keisler, H.J. (1970) Logic with the quantifier "there exist uncountably many". *Ann. Math. Logic*, *1*, 1–93.

Kneale, W. and Kneale, M. (1962) *The Development of Logic*. Oxford, Clarendon Press.
Lear, J. (1980) *Aristotle and Logical Theory*. Cambridge, Cambridge University Press.
Newell, A. (1981) Reasoning, problem solving and decision processes: the problem space as a fundamental category. In R. Nickerson (ed.), *Attention and Performance, VIII*. Hillsdale, NJ., Erlbaum.
Newell, A. and Simon, H.A. (1972) *Human Problem Solving*. Englewood Cliffs, NJ., Prentice-Hall.
Oakhill, J. and Johnson-Laird, P.N. (1983) The search for counterexamples. Mimeo, Laboratory of Experimental Psychology, Sussex University.
Revlin, R. and Leirer, V.O. (1978) The effect of personal biases on syllogistic reasoning: rational decisions from personalized representations. In R. Revlin and R.E. Mayer (eds.), *Human Reasoning*. Washington, D.C., Winston.
Revlis, R. (1975) Two models of syllogistic reasoning: feature selection and conversion. *J. verb. Learn. verb. Behav.*, *14*, 180–195.
Scribner, S. (1977) Models of thinking and ways of speaking: culture and logic reconsidered. In P.N. Johnson-Laird and P.C. Wason (eds.), *Thinking: Readings in Cognitive Science*. Cambridge, Cambridge University Press.
Wason, P.C. and Johnson-Laird, P.N. (1972) *The Psychology of Reasoning: Structure and Content*. London, Batsford. Cambridge, MA., Harvard University Press.
Woodworth, R.S. and Schlosberg, H. (1954) *Experimental Psychology*. Third edition. New York, Holt, Rinehart, and Winston.
Woodworth, R.S. and Sells, S.B. (1935) An atmosphere effect in formal syllogistic reasoning. *J. exper. Psychol.*, *18*, 451–460.

Résumé

Dans cet article sont passées en revue les théories psychologiques du traitement des syllogismes. On établit qu'en dépit de leurs mérites variés toutes sont en défaut en tant que théories de la performance. On présente les résultats de deux expériences, une utilisant des syllogismes et l'autre des problèmes avec des séries de trois termes conçues pour élucider comment l'arrangement des termes dans les premises ("figure" des premises) affecte la performance. Ces données sont utilisées pour construire une théorie fondée sur l'hypothèse que les gens construisent des modèles mentaux des premises, formulent des conclusions explicites sur les relations dans le modèle et cherchent des modèles qui seraient des contre-exemples pour leurs conclusions. Cette théorie, utilisée dans plusieurs programmes d'ordinateur, prédit que deux principaux facteurs affectent la performance: la figure des premises et le nombre des modèles qu'ils mettent en jeu. Cette prédiction est confirmée dans les trois expériences.