Psychological Review
1989, Vol. 96, No. 4, 658–673

Copyright 1989 by the American Psychological Association, Inc.
0033-295X/89/$00.75

Reasoning by Model: The Case of Multiple Quantification

P. N. Johnson-Laird and Ruth M. J. Byrne
MRC Applied Psychology Unit
Cambridge, England

Patrizia Tabossi
University of Bologna
Bologna, Italy

A theory of deductive reasoning is presented for a major class of inferences that has not been investigated by psychologists: inferences that depend on multiply-quantified premises (e.g., "*None* of the Princeton letters is in the same place as *any* of the Dublin letters"). It is argued that reasoners construct mental models based on their knowledge of the meanings of quantifiers (and other terms, including relational expressions). Three experiments corroborate the model theory's prediction that inferences that require the construction of only 1 model will be easier than those that require more than 1 model. The model theory assumes that the logical properties of quantifiers emerge from their meanings and are not mentally represented in rules of inference. How such a semantic process can occur *compositionally* (i.e., guided by the syntactic analysis of sentences) is described.

Deductive reasoning is a process of thought that yields new information from old and aims to establish valid conclusions, that is, conclusions that are necessarily true given the truth of the initial premises or observations. The study of its underlying mental mechanisms is almost as old as experimental psychology, but remains a matter of controversy. There are three main views that have been proposed in both cognitive psychology and artificial intelligence. First, the reasoning mechanism depends on formal rules of inference; second, it depends on content-specific rules of inference; and, third, it depends on semantic procedures that search for interpretations (or models) of the premises that are counterexamples to conclusions. The principal goal of this article is to establish a theory of deductive reasoning for a major class of inferences that has not been investigated before by psychologists: those that depend on multiply-quantified premises. There are theories of relational reasoning and of syllogistic reasoning (i.e., from singly quantified premises), but multiple quantification is more powerful, and its logical analysis calls for the full resources of that branch of logic known as the *first-order predicate calculus*. In this article we develop a theory based on the manipulation of models and report experimental evidence that confirms this theory.

An example of a multiply-quantified assertion is as follows: *None* of the artists is taller than *any* of the beekeepers. Such assertions contain a relational expression—here, a two-place relation, "taller than"; its arguments are quantified by such expressions as "all," "some," "none," and "any." These quantifiers behave in ways that are similar to the quantifiers of the first-order predicate calculus, but there are other "nonstandard" quantifiers, such as "most" and "few," that do not. Our concern here is solely with standard quantifiers, and we begin by considering them in the light of the three major classes of psychological theories of reasoning.

Three Theories of Reasoning

Formal Rules of Inference

For many years, it was taken for granted that the human inferential mechanism is based on formal rules of inference (see, e.g., Braine, 1978; Inhelder & Piaget, 1958; Osherson, 1975; Rips, 1983; Wason & Johnson-Laird, 1972). According to these theories, the first step in reasoning is to make manifest the *logical form* of the premises by representing them in an internal language that reveals this syntactic structure. Next, formal rules of inference are used (by definition, in a purely syntactic way) to derive conclusions. The theories have usually adopted the philosophy of so-called *natural deduction*, in which each connective has its own associated rules of inference (see, e.g., Braine, 1978; Johnson-Laird, 1975; Osherson, 1975; Rips, 1983). For example, the rule of *modus ponens* governs implications of the form, if p then q, which can be symbolized as $p \rightarrow q$. Hence, given assertions of the form

$$p$$

$$p \rightarrow q,$$

the rule permits the derivation of the conclusion: q.

Different formal theories vary in the details of the language in which they express logical form and in their rules of inference. Braine and Rumain (1983, p. 296) observed of quantified inferences of the following sort:

> All psychologists are either clinicians or experimenters
>
> Therefore, all psychologists who are not clinicians are experimenters

that the conclusion seems immediate. They wrote, "Adults we

We are grateful to Rachel Falmagne, Sam Glucksberg, Mark Keane, Christopher Longuet-Higgins, and to two anonymous "mental logicians" for helpful criticisms of an earlier draft of this article, and we thank Ian Nimmo-Smith for statistical advice.

Correspondence concerning this article should be addressed to P. N. Johnson-Laird, MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge, CB2 2EF, England.

have asked say that the conclusion follows directly. This behaviour suggests that they have an inference schema that takes them from premise to conclusion in a single step." They, accordingly, proposed that mental logic contains a rule of the following form:

a's are either F or G.

Therefore, a's that are not F are G.

Three stages of inference that are separated by logicians are thereby collapsed into one: the elimination of quantifiers, deduction based on connectives, and the restoration of appropriate quantifiers.

We accept that in simple inferences the use of separate rules to eliminate and to restore quantifiers seems highly implausible. But, there is a cost to be paid for doing away with these rules and building their effects into the rules for connectives. The resulting system is not complete; that is, there are valid inferences that cannot be derived within it, including those multiply-quantified deductions that are the topic of this article. There is, in fact, no existing psychological theory based on formal rules that is sufficiently powerful to accommodate them.

Formal theories have been successful in predicting some empirical observations. Osherson (1975) discovered that the number of steps in a derivation correlates with subjects' performance in judging the validity of propositional inferences. Similarly, the number of steps required for propositional inferences according to Braine, Reiser, and Rumain's (1984) theory predicts the latencies and rated difficulty of subjects' evaluations of conclusions. Rips (1983) has developed a computer simulation that also models a number of empirical phenomena in the protocols of subjects who think aloud while proving propositional inferences. Despite these successes, formal theories run into several problems.

One problem is that the content of premises can have striking effects on the conclusions that subjects draw from them (e.g., Byrne, 1989; Byrne & Johnson-Laird, 1989; Griggs & Cox, 1982; for reviews, see Evans, 1982; Wason & Johnson-Laird, 1972). Yet formal rules by definition are supposed to apply regardless of content. Theorists argue that these effects arise during the stage of understanding the premises and retrieving their logical form. An alternative hypothesis is that there are content-specific rules of inference.

Content-Specific Rules of Inference

The idea of content-specific rules of inference was first proposed in artificial intelligence (Hewitt, 1971) and is related to the development of production systems (Anderson, 1983; Newell, 1973). Variations on it have been explored by psychologists seeking to account for the effects of content (Johnson-Laird & Wason, 1977, p. 353). Cheng and Holyoak (1985) have also argued that people are guided by pragmatic reasoning schemas, that is, general rules that apply to particular classes of goals. An example is the *permission* schema: If action A is to be taken, then precondition B must be satisfied; if precondition B is not satisfied, then action A must not be taken.

A content-specific rule of inference—couched, for instance, in a production system—expresses a general assertion, such as that all psychologists are experimenters, in the form of a procedure:

Condition: (Psychologist x)

Action: (Experimenter x),

so that whenever its antecedent condition is matched by an assertion about a specific individual, for example,

(Psychologist Alicia),

the action is automatically triggered and makes the further assertion,

(Experimenter Alicia).

This variable-binding mechanism is thus similar to Braine and Rumain's (1983) collapsing of rules for quantifiers and connectives into a single step.

General assertions, such as "All psychologists are experimenters," can also be expressed in a different procedure that works in a backward chain from the goal of establishing that someone is an experimenter to the subgoal of establishing that the same individual is a psychologist. In other words, whenever the subgoal is satisfied, the main goal is also satisfied.

Formal and content-specific rules both run into a dilemma. On the one hand, people can make some deductions that depend solely on the properties of quantifiers and connectives; and so content-specific rules alone cannot account for all reasoning. On the other hand, people are sometimes affected by content, and so a uniform procedure for extracting formal structure and applying formal rules to it cannot account for all reasoning, either. Both theories can, of course, be adapted so as to avoid the dilemma. Another way out is provided by the third, and radically different, class of theories, in which inference depends not on matching sentences to rules of inference but on procedures that operate on the semantic interpretations of premises.

Reasoning by Model

Logicians long ago recognized that in addition to a purely formal system of inference—the proof-theoretic methods that exploit essentially syntactic procedures—it is also possible to formulate a semantic system based on providing a model-theoretic interpretation of the formal language. A proof-theoretic system defines formal derivations of proofs; a model-theoretic system defines valid inferences (i.e., those in which the conclusion is true in any interpretation, or model, of the premises). A major part of metalogic is devoted to proving certain relations between the formal and the semantic systems. A formal calculus is thus said to be "complete" if any theorem that is valid in the semantics is derivable within the calculus. A major discovery in 20th-century logic is that there are incomplete logics. For example, there is no consistent way to formalize the *second-order predicate calculus* (in which quantified variables can take sets as their values) that is guaranteed to enable all valid conclusions to be derived. Formal systems are therefore not trivial variants of the semantics of logical systems (see, e.g., Jeffrey, 1981).

There are various psychological theories of reasoning based on the notion of constructing and manipulating models of assertions. Some theorists have argued that inferences based on singly-quantified premises, such as "All psychologists are experimenters," are interpreted in the form of Euler circles (Erickson, 1974) or by equivalent strings of symbols (Guyote & Sternberg, 1981). Others have explored systems based on Venn diagrams or equivalent strings of symbols (Newell, 1981). However, the standard systems of Euler circles and Venn diagrams apply only to singly-quantified assertions and have no way of representing, for example, "Some of the psychologists are taller than some of the linguists" (see, e.g., Gardner, 1958). Hence, we need another sort of theory that assumes that mental models have the same structure as the situations that they represent (Johnson-Laird, 1983). In this theory, a finite set of individuals is represented, not by a circle enscribed in Euclidean space, but by a finite set of mental tokens. These tokens may occur in the form of a visual image, or they may not be directly accessible to consciousness. What matters is not the subjective experience—most people claim to be unaware of how they reason—but the *structure* of the hypothesized models.

Theories based on these models have been successful in accounting for the patterns of performance that occur in reasoning from singly-quantified premises, such as syllogisms (see Byrne & Johnson-Laird, in press-b; Johnson-Laird & Bara, 1984a), in reasoning about two-dimensional spatial relations (see Byrne & Johnson-Laird, in press-c), and in reasoning propositionally (see Johnson-Laird, Byrne, & Schaeken, 1989). The number of models constructed for any particular premises depends on the precise details of the reasoning algorithm. Johnson-Laird and Bara (1984a) described two computer programs that construct differing numbers of models for syllogistic reasoning: One program never constructs more than two models for any syllogism; the other program constructs three models for certain cases. Perhaps because people differ in the procedures they use, it has so far proved impossible to decide between the two accounts. What is common to them both, however, is the set of syllogisms that require only one model. Hence, the crux of the theory is whether it is necessary to construct one model or more than one model in order to draw a valid conclusion. Problems requiring only one model offer no choice to the reasoner, but where there is a choice, the problem will be more difficult because ordinary reasoners evidently have no simple deterministic procedure for searching for counterexamples. We now propose to extend this theory to inferences based on multiply-quantified assertions, but we need first to outline the requisite formal logic.

The Logic of Multiply-Quantified Assertions

An assertion such as

Some artists are not in the same place as all beekeepers,

contains two quantified arguments of a two-place relation. There are 12 logically distinct ways of quantifying a two-place relation:

1. $(\forall\text{A})(\forall\text{B})(\text{A}R\text{B})$
2. $(\text{EA})(\forall\text{B})(\text{A}R\text{B})$
3. $(\text{EB})(\forall\text{A})(\text{A}R\text{B})$
4. $(\forall\text{B})(\text{EA})(\text{A}R\text{B})$
5. $(\forall\text{A})(\text{EB})(\text{A}R\text{B})$
6. $(\text{EA})(\text{EB})(\text{A}R\text{B})$

and their respective negations. There are many ways of expressing each of the resulting propositions in natural language. Sentences with quite different superficial forms can express the same underlying proposition. Thus, each of the following sentences has an interpretation in common:

Some artists are not in the same place as all beekeepers.

Not all artists are in the same place as all beekeepers.

Some artists are not in the same place as some beekeepers.

The predicate calculus captures the senses of these sentences by the following logically equivalent expressions in which "$\neg$" symbolizes negation; "E" symbolizes the existential quantifer "at least some"; "$\forall$" symbolizes the universal quantifier "any"; and, for simplicity, the apparatus of restricting quantification to particular sets has been taken for granted:

(E artist) $\neg$ ($\forall$ beekeeper) (In-same-place artist beekeeper)

$\neg$ ($\forall$ artist)($\forall$ beekeeper) (In-same-place artist beekeeper)

(E artist)(E beekeeper) $\neg$ (In-same-place artist beekeeper).

In the predicate calculus, unlike natural language, the universal quantifier makes no claim about the existence of members in the corresponding set, whereas "all" often carries such an implication (see Boolos, 1984; Johnson-Laird & Bara, 1984b). Likewise, in natural language, unlike the predicate calculus, there are ambiguities in scope. An assertion such as

Every artist is in the same place as some beekeeper,

has two distinct interpretations:

($\forall$ artist)(E beekeeper)(In-same-place artist beekeeper),

which means that every artist is in the same place as some beekeeper or other, and

(E beekeeper)($\forall$ artist)(In-same-place artist beekeeper),

which means that there is at least some beekeeper who is in the same place as every artist. The predicate calculus distinguishes between these two interpretations in terms of two different formulas. Scope ambiguities, however, are not invariable in natural language. Thus, the assertion

No artist is in the same place as some beekeepers,

strongly suggests the interpretation in which "some" has the larger scope; that is, there are some beekeepers for whom no artist is in the same place:

(E beekeeper) $\neg$ (E artist)(In-same-place artist beekeeper).

Linguistic theories of the interpretation of quantified sentences have generally operated in ways that are similar to the predicate calculus, although they differ in the theoretical paradigm within which they are framed (see Harman, 1972; Johnson-Laird, 1970; Keenan, 1971; Lakoff, 1972; May, 1985; Partee, 1975). Thus, for example, Hornstein (1984) captured differences in scope by postulating distinct underlying logical forms, which are transformationally derived from the surface syntax of sentences, whereas Cooper (1983) handled them within the semantic component of an analysis based on Montague grammar.

Relations differ quite widely in their logical properties, and logicians distinguish between several major families of them. We are concerned here with just two of their logical properties. First, a relation is transitive if it yields a transitive conclusion; for example,

If x is in the same place as y,

and y is in the same place as z,

then x is in the same place as z.

Second, a relation is symmetric if it yields a symmetric conclusion, for example,

If x is in the same place as y,

then y is in the same place as x.

There are other logical properties, and, as the examples should make clear, one can ask what is the status of any relation with respect to any property. Thus, the relation "in the same place as" is transitive and symmetric. Whatever logical properties of a relation are needed to derive a conclusion in the predicate calculus, they must be stated as separate assumptions (or "meaning postulates") along with the premises.

There are various ways in which to formalize the predicate calculus. The usual strategy for proofs depends on a three-stage process:

1. The quantifiers in the premises are eliminated by special rules (rules of instantiation).
2. Reasoning then occurs on the basis of propositional connectives, such as disjunction and conjunction, and accordingly uses only propositional rules.
3. Quantifiers are restored to the resulting conclusion, using special rules to reintroduce them (rules of generalization).

Because the calculus recognizes two quantifiers, the universal quantifier "any" and the existential quantifier "some," each has its own rules of instantiation and generalization. The rule for instantiating an existential quantifier is straightforward. Given an existential assertion, such as

(Ex)(Philosopher x) & (Psychologist x),

the rule says, in effect, because the premise applies to at least someone in the domain of discourse, then at any point in a proof, a hypothetical individual can be introduced as the value of the variable, provided that the same individual has not been instantiated earlier:

(Philosopher Dan) & (Psychologist Dan).

The premise applies to some hypothetical individual—here, assumed to be Dan. Hence, this hypothetical individual is ultimately going to have to be replaced by an existential quantifier. Given a universally quantified implication, "If anyone is a psychologist, then he or she is an experimenter":

$(\forall x)$(Psychologist x) $\rightarrow$ (Experimenter x),

the rule of universal instantiation says, in effect, because the premise applies to everyone in the domain of discourse, then at any point in a proof anyone can be freely introduced as the value of the variable, for example,

(Psychologist Dan) $\rightarrow$ (Experimenter Dan).

Once the quantifiers have been eliminated from expressions, the second stage, which concerns only the connectives, can take place. The third and final stage reintroduces quantifiers. Where an individual is an instantiation of an existential quantifier, then this quantifier must be restored; otherwise, a universal quantifier can be restored.

Table 1 shows a summary of a formal derivation of the following valid inference:

None of the Princeton letters are in the same place as any of the Cambridge letters.

All of the Cambridge letters are in the same place as all of the Dublin letters.

Therefore, None of the Princeton letters are in the same place as any of the Dublin letters.

In order to derive the conclusion, it is necessary to state explicitly (in meaning postulates) that the relation "in the same place as" is transitive and symmetric. Stage 1 eliminates the quantifiers, Stage 2 uses the rules of inference governing propositional connectives, and Stage 3 reintroduces quantifiers. Because the hypothetical individuals were originally introduced by universal instantiation, the rule of universal generalization can be used to restore universal quantifiers.

The premises of the following inference differ only in one of the quantifiers in the second premise,

None of the Princeton letters are in the same place as any of the Cambridge letters.

All the Cambridge letters are in the same place as some of the Dublin letters.

Therefore, None of the Princeton letters are in the same place as some of the Dublin letters.

or, equivalently,

Some of the Dublin letters are not in the same place as any of the Princeton letters.

The formal derivation of the inference is almost identical to the one in Table 1: It uses exactly the same propositional rules in Stage 2, and the derivation as a whole is exactly the same length. The only difference is that the instantiation of the existentially quantified variable must include a tag to ensure that the final process of generalization restores an existential quantifier.

Although there are psychological theories of reasoning based on formal rules of inference, these theories are primarily concerned with propositional reasoning (see Braine, 1978; Braine,

Table 1
A Formal Derivation in the Predicate Calculus

The premises can be symbolized as	
1. $(\forall P)(\forall C)\neg(PSC)$	[None of the P are in the same place as any of the C]
2. $(\forall C)(\forall D)(CSD)$	[All of the C are in the same place as all of the D]
3. $(\forall X)(\forall Y)(\forall Z)(XSY \& YSZ \rightarrow XSZ)$	[Transitivity of 'in the same place as']
4. $(\forall X)(\forall Y)(XSY \rightarrow YSX)$	[Symmetry of 'in the same place as']
Stage 1: Instantiation of quantifiers	
5. $(\forall C)\neg(pSC)$	[universal instantiation of P in premise 1]
6. $\neg(pSc)$	[universal instantiation of C in 5]
7. $(\forall D)(cSD)$	[universal instantiation of C in premise 2]
8. (cSd)	[universal instantiation of D in 7]
9, 10, 11. $(pSd \& dSc) \rightarrow pSc$	[universal instantiation of X, Y, and Z in 3]
12, 13. $(cSd) \rightarrow (dSc)$	[universal instantiation of X and Y in 4]
Stage 2: Propositional reasoning	
14. $\neg(pSd \& dSc)$	[modus tollens from lines 6 and 11]
15. $\neg(pSd)$ or $\neg(dSc)$	[equivalent to 14 by de Morgan's law]
16. (dSc)	[modus ponens from 8 and 13]
17. $\neg(pSd)$	[disjunctive rule from 15 and 16]
Stage 3: Reintroduction of quantifiers	
18. $(\forall D)\neg(pSD)$	[Universal generalization of d in 17]
19. $(\forall P)(\forall D)\neg(PSD)$	[Universal generalization of p in 18]
The conclusion corresponds to: None of the P are in the same place as any of the D.	

Note. $\forall$ = universal quantifier "all"; $\neg$ = symbol for negation; S = the relation "in the same place as." For simplicity, we have adopted the apparatus of restricting quantified variables to particular sets: P = Princeton letters, C = Cambridge letters, and D = Dublin letters.

Reiser, & Rumain, 1984; Osherson, 1975; Rips, 1983). Unfortunately, as we mentioned earlier, there is no such theory that is powerful enough for multiply-quantified inferences. We will consider the possibility of constructing such a theory later; but now we turn to a theory based on mental models.

A Theory of Multiply-Quantified Reasoning Based on Mental Models

According to the model theory, a valid inference depends on three stages of thought. First, the premises are understood, that is, interpretative procedures construct a mental model representing their content. This process depends on the meaning of the premises and on general knowledge. Second, if possible, a conclusion is formulated that asserts something that is not explicitly stated in any single premise. Third, to test the validity of the conclusion, a search is made for alternative models of the premises that refute the conclusion. If there is no such alternative model, the conclusion is valid. If there is such an alternative model, then the reasoner should return to the second stage to try to determine whether there is any conclusion true in all of the models so far constructed. If it is uncertain that there is such an alternative model, then the reasoner can draw the conclusion in a tentative or probabilistic way.

The first stage of reasoning, accordingly, is the construction of a model of the premises. The theory postulates that mental models represent (a) finite sets of individuals by finite sets of mental tokens and (b) properties of individuals and relations between them by corresponding mental tokens. For example, the premise

> None of the Princeton letters is in the same place as any of the Cambridge letters.

has the following sort of model,

| p p p | c c c |

which represents three Princeton letters and three Cambridge letters—the numbers of letters are arbitrarily selected. The relation "in the same place as" is here represented by mental tokens corresponding to spatial locations; thus, the model represents entities in places, and the vertical barriers signify boundaries (i.e., the Princeton letters are *not* in the same place as the Cambridge letters). The information in a second premise

> All of the Cambridge letters are in the same place as all of the Dublin letters.

can be added to the model to yield the following,

| p p p | c c c d d d |

The second stage of reasoning is to formulate a putative conclusion that conveys something that is not explicit in any premise. A procedure that scans the model for such a relation will yield the conclusion

None of the Princeton letters is in the same place as any of the Dublin letters.

The third stage calls for a search for alternative models that refute the putative conclusion. The fundamental distinction in the theory is between those premises that require only one model to be constructed and those that require more than one model. In our current example, there is no alternative model and so the conclusion is valid.

We will now consider an inference that does depend on the construction of multiple models. Consider the premises

None of the Princeton letters is in the same place as any of the Cambridge letters.

All of the Cambridge letters are in the same place as some of the Dublin letters.

The first premise yields the same sort of model as before,

| p p p | c c c |

When the information from the second premise comes to be added to this model, then, unlike the previous example, there is a choice between more than one possible model. One of these models is

| p p p | c c c d d | od |

in which the "o" in front of the rightmost "d" indicates that this item is optional (i.e., it may or may not be present in the situation represented by the model). This model supports the same conclusion as before:

None of the Princeton letters is in the same place as any of the Dublin letters.

Another possible model, however, is

| p p | c c c d d | p od |

which falsifies the previous conclusion. Taken together, the two models support the conclusion:

Some of the Princeton letters are not in the same place as any of the Dublin letters.

But, again, this conclusion can be refuted by a further model,

| p p p od | c c c d d |

These three models exhaust the set of possibilities, ignoring as irrelevant, of course, the precise numbers of tokens. Hence, the only valid conclusion, which holds over all of the models, is

None of the Princeton letters is in the same place as some of the Dublin letters,

or, equivalently:

Some of the Dublin letters are not in the same place as any of the Princeton letters.

There are, of course, weaker conclusions that also follow, such as

Some of the Dublin letters are not in the same place as some of the Princeton letters.

or in one of its senses,

None of the Princeton letters is in the same place as all of the Dublin letters.

There are no further models of the premises that refute these conclusions.

In short, the first example offered no choice about how to combine the two premises, but the present case offers such a choice and so it has more than one model. Choice, here, concerns not trivial variations in the particular number of tokens representing a set, but only those matters that directly affect what valid conclusions can be drawn from the premises, that is, the place into which a letter is put. The theory, accordingly, assigns any set of premises to one of three categories: (a) those that yield one model and, therefore, validly support a novel conclusion; (b) those that yield more than one model but validly support a novel conclusion because it holds across all of the models; and (c) those that yield more than one model and do not validly support a novel conclusion because none holds across all of the models. An informal procedure for discovering the status of a problem is to construct a model of the first premise and then to determine whether there is any choice in adding the information from the next premise to the model. If not, then the premises support one model and have a valid conclusion. If there is a choice, then it is necessary to determine whether the resulting set of models supports a conclusion that interrelates the end terms and that holds across all of the models.

With multiple-model *syllogisms*, it is clear that people are not equipped with a simple deterministic algorithm for searching for alternative models; they produce a wide variety of responses to the same problem and are not particularly consistent from one presentation of it to the next (see Johnson-Laird & Bara, 1984a). A similar lack of a deterministic search procedure for *problem solving* has been reported by Newell and Simon (1972, chap. 1). Hence, granted that working memory has a limited processing capacity, the present theory predicts that people should perform more accurately with one-model problems than with multiple-model problems with valid conclusions. It also predicts that they should perform more accurately with one-model problems than with multiple-model problems with no valid conclusions, although here the comparison is confounded by the qualitatively distinct responses to the two sorts of problems. (If people are biased toward responding that there is no valid conclusion, then they may perform quite well on such problems for spurious reasons.) The aim of Experiment 1 was to determine whether these predictions were correct.

Experiment 1

Because there had been no previous study of multiply-quantified inferences, we needed to find out whether ordinary individuals were able to reach a reasonable proportion of correct conclusions for a sample of problems. There is a large number of doubly-quantified inferences. Each premise can express 1 of 12 underlying propositions, and there are also often many ways, as we have seen, to express one of these propositions. The order of the terms in the two premises can be in one of four so-called *figures*.

Table 2
Nine One-Model or Multiple-Model Problems Used in Experiment 1

One-model problems
1. All of the X are in the same place as all of the Y All of the Y are in the same place as all of the Z (Therefore, All of the X are in the same place as all of the Z.)
2. All of the X are in the same place as some of the Y All of the Y are in the same place as all of the Z (Therefore, All of the X are in the same place as all of the Z.)
3. All of the X are in the same place as all of the Y Some of the Y are in the same place as all of the Z (Therefore, All of the X are in the same place as all of the Z.)
Multiple-model problems with no valid conclusion
4. None of the X are in the same place as any of the Y None of the Y are in the same place as any of the Z (No valid conclusion interrelating X and Z.)
5. None of the X are in the same place as some of the Y None of the Y are in the same place as any of the Z (No valid conclusion interrelating X and Z.)
6. None of the X are in the same place as any of the Y Some of the Y are in the same place as none of the Z (No valid conclusion interrelating X and Z.)
Multiple-model problems with valid conclusions
7. None of the X are in the same place as any of the Y All of the Y are in the same place as some of the Z (Therefore, None of the X are in the same place as some of the Z.)
8. None of the X are in the same place as some of the Y All of the Y are in the same place as some of the Z (Therefore, None of the X are in the same place as some of the Z.)
9. None of the X are in the same place as any of the Y Some of the Y are in the same place as some of the Z (Therefore, None of the X are in the same place as some of the Z.)

Note. The conclusions (in parentheses) were not presented to the subjects.

Figure 1	Figure 2	Figure 3	Figure 4
A–B	B–A	A–B	B–A
B–C	C–B	C–B	B–C

Where a relation is symmetric, however, the difference between one order and another affects the logic of the problem only if the scope of the quantifiers changes.

It was clearly impossible to test subjects with the full set of problems, and so in Experiment 1, we used six one-model problems, six multiple-model problems with valid conclusions, and six multiple-model problems with no valid conclusions. The main aim of the experiment was to determine whether there was any difference between the one-model and the multiple-model problems with valid conclusions. If not, then we could immediately reject the model-based theory.

Our set of inferences was also chosen to answer a subsidiary question. There is always one set of individuals that is referred to in both premises—the so-called *middle term*—and in some of the inferences one reference to it concerned the whole set (e.g., "all of the Cambridge letters"), whereas the other reference did not (e.g., "some of the Cambridge letters"). Such an apparent clash of quantifiers should cause reasoners some difficulty in building a single model unless they treat "some" as signifying "at least some and possibly all." Hence, the model-based theory predicts that subjects would find such inferences slightly harder than those in which there was no clash of quantifiers.

Method

The experiment was carried out in Bologna, Italy, and the relation that was used in all of the problems was, "nello stesso posto," which is Italian for "in the same place as." This relation was chosen because it has the required logical properties (of transitivity and symmetry) to enable us to construct three one-model problems, three multiple-model problems with valid conclusions, and three multiple-model problems with no valid conclusions. These nine problems, which are shown in Table 2, are all in the first figural arrangement. We constructed a further nine problems by presenting the premises of these problems in the opposite order (i.e., in the second figural arrangement). The content of the terms consisted in the names of hobbies and occupations selected so that there was no obvious a priori relation among them. The triples of nouns were assigned to the 18 problems at random, in two different ways: One half of the subjects received one set of materials, and the other half of the subjects received the other set of materials. Each subject carried out the problems in a different random order.

The subjects were tested individually, and they were told to state what, if anything, followed from the premises. Each problem was printed on a separate piece of paper. The subjects had to read the problem aloud and then, working at their own pace, state their answer aloud. The experimenter recorded the answer. Subjects were told that in some cases there might not be sufficient information to draw a conclusion between the two end terms; in that case they were to say "no valid conclusion." The task was explained by way of an example that also served as a practice trial.

We tested 20 students at the University of Bologna (13 women and 7 men) whose ages ranged from 20 to 25 years. Two of the subjects were replaced during the course of the experiment because they did not follow the instructions. There are no selection procedures for entry into Italian universities, and so our subjects there come from a much broader population than do students in the United States, for example.

Results and Discussion

Table 3 shows the percentages of correct responses for the three sorts of problem. There were 68% correct solutions to the one-model problems and 13% correct solutions to the multiple-model problems with valid conclusions; every subject went in the predicted direction, except for two ties ($p = .5^{18}$). The difference between the one-model problems and the multiple-model problems with no valid conclusions (50% correct) was in the predicted direction, but was not reliable (Wilcoxon's $T = 42.5$, $n = 16$, $p > .05$). As Table 3 suggests, the first figural arrangement was slightly easier than the second (Wilcoxon's $T = 21.5$, $n = 16$, $p < .02$).

For the one-model problems, those with the same quantifiers in the middle term yielded 80% correct conclusions, and those with conflicting quantifiers yielded 63% correct conclusions. This predicted difference was significant (Wilcoxon's $T = 9$, $n = 12$, $p < .01$). Although the trend was in the predicted direction for the multiple-model problems with valid conclusions, there

Table 3
Percentages of Correct Responses to the Three Sorts of Problems in Experiment 1

Figure	One model	Multiple model with no valid conclusion	Multiple model with valid conclusion	Overall
1	73	50	25	49
2	63	50	0	38

were too few correct responses overall for the difference to be significant. The reason for the difference, we believe, is that when one premise asserts that a set is in the same place as all of the members of a second set and another premise locates only some of this second set, there is an initial problem in construing the two premises. The way to construct a unified model is, in effect, to overrule the premise containing "some" and to interpret the quantifier as referring to the whole set.

The results corroborate the model theory. One-model problems were easier than multiple-model problems with valid conclusions. Although the one-model problems were not reliably easier than the multiple-model problems with no valid conclusions, the trend was in the right direction. The failure to obtain a significant difference might be because of the qualitatively distinct responses required to the two sorts of problem: in one case, a conclusion, and in the other case, the response that there is no valid conclusion. The finding that problems in the first figural arrangement are slightly easier than those in the second figural arrangement is entirely analogous to a figural effect for syllogisms. Figure principally influences the order in which terms are mentioned in syllogistic conclusions, but it sometimes has an effect on difficulty (see, e.g., Johnson-Laird & Bara, 1984a, who explain the phenomenon in terms of the order in which information enters working memory).

One incidental but important observation is that the results cannot be explained by any simple response-priming effect. The correct conclusion to the one-model problems always matches the logical form of one of the premises; for example,

All X are in the same place as All Y.

Some Y are in the same place as All Z.

Therefore, All X are in the same place as All Z.

In at least two cases, however, this condition also applied to valid multiple-model problems (i.e., the two problems, in the first and second figural arrangements, based on Item 8 in Table 2):

None of the X are in the same place as some of the Y.

All of the Y are in the same place as some of the Z.

Therefore, None of the X are in the same place as some of the Z.

The match here between the conclusion and the first premise in no way facilitated performance. The subjects were correct on only 13% of trials—identical to their performance on the other valid multiple-model problems in which there was no such match.

Although our results appear to confirm the model theory, there is an alternative possibility to be taken into account. All of the one-model problems are based on two affirmative premises, but all of the multiple-model problems are based on one affirmative and one negative premise. It is well-known that negative assertions are harder to understand (see, e.g., Clark & Clark, 1977; Wason, 1959), and so this factor could account for at least part of the difference between the two sorts of problems. Experiment 2 examined this alternative explanation.

Experiment 2

The aim of Experiment 2 was twofold. First, we needed to control for the effects of negative premises, which were previously confounded with the number of models. Second, we wanted to extend our findings, if possible, to a variety of different relations with the same appropriate logical properties.

Method

The subjects acted as their own controls and were asked to state in their own words what, if anything, followed from 18 different doubly-quantified pairs of premises. All of the problems, including those with no valid conclusion, were based on the combination of one affirmative premise and one negative premise. Hence, any difference in difficulty could not be attributed to negation. The problems were either one-model or multiple-model problems. They derived from three basic patterns of premises, which we state here, together with valid conclusions, although these were not presented to the subjects.

One-model problem:

None of the A is in the same place as any of the B.

All of the B are in the same place as all of the C.

[Therefore, None of the A is in the same place as any of the C.]

Multiple-model problem with a valid conclusion:

None of the A is in the same place as any of the B.

All of the B are in the same place as some of the C.

[Therefore, None of the A are in the same place as some of the C.]

Multiple-model problem with no valid conclusion:

None of the A is in the same place as some of the B.

Some of the B are in the same place as all of the C.

[No valid conclusion interrelating the end terms, A and C.]

Each of these basic patterns is in the first figural arrangement, and a further three basic patterns were constructed in the second figural arrangement by stating each pair of premises in the opposite order. We used three different transitive and symmetric relations to make a set of 18 problems out of the six basic patterns: "in the same place as," "equal in height to," and "related to" in the simple consanguineal sense that subjects naturally treat as being transitive and symmetric. The nouns in the premises referred to hobbies and professions selected so as to obviate any a priori connections.

We devised one set of lexical contents and assigned it twice at random to the 18 problems; one half of the subjects were tested with one of the assignments, and the other half of the subjects were tested with the other

Table 4
Percentages of Correct Responses in Experiment 2 as a Function of Models and the Nature of the Relation

Relation	One model	Multiple model with valid conclusion	Multiple model with no valid conclusion
Equal in height to	69	14	42
In the same place as	58	17	33
Related to	72	17	44

assignment. Each subject was presented with the inferences in a different random order. The procedure was the same as in Experiment 1.

We tested 20 members (17 women and 3 men) of the Applied Psychology Unit subject panel, whose ages ranged from 22 to 60 years. These subjects come from a variety of occupations and are more generally representative of the population at large than are university students. We eliminated 2 subjects because they failed to understand the instructions.

Results and Discussion

Table 4 presents the percentages of correct responses to each of the different sorts of problems. Although the data are binary, we carried out an analysis of variance on them (see Scheffé, 1960). There was only one significant result: As the model theory predicts, the number of models had a striking effect on the accuracy of subjects' responses; there were 67% correct conclusions for one-model problems, 16% correct conclusions for multiple-model problems with valid conclusions, and 40% correct responses to multiple-model problems with no valid conclusions, $F(2, 34) = 17.05$, $p < .0005$. Subsequent comparisons with Newman-Keuls tests revealed that the one-model problems were reliably easier than both the valid multiple-model problems, $q(3, 34) = 8.26$, $p < .001$, and the invalid multiple-model problems, $q(2, 34) = 4.36$, $p < .001$. Within the multiple-model problems, the invalid ones were reliably easier than the valid ones, $q(2, 34) = 3.91$, $p < .05$. No other factor, or interaction between factors, approached significance. In particular, there was no effect of the different sorts of relations.

One incidental observation provided a striking corroboration of the model theory. We observed a tendency for subjects to draw an odd sort of conclusion. Given the following multiple-model problem:

None of the tourists are in the same place as any of the artists.

All of the artists are in the same place as some of the doctors,

subjects typically erred by concluding

None of the tourists are in the same place as any of the doctors,

which suggests that they were constructing an initial model of the form

| t t t | a a a d d | d d |

and were failing to check for alternatives. One subject, however, drew exactly the opposite conclusion,

Some doctors and tourists are in the same place,

and another concluded

Some tourists and doctors could be in the same place.

The reason for such conclusions, we believe, is that subjects, perhaps after they have constructed the aforementioned initial model, realize that the following alternative model is also possible:

| t t t d d | a a a d d |

Their attention is then caught by the fact that some of the doctors could indeed be in the same place as the tourists—as the model illustrates—and they overlook the need for a conclusion to be true in all possible models of the premises. If subjects are constructing alternative models in this way, then they should tend to use modal qualifications of their conclusions. In fact, we observed that 21% of the responses to the valid multiple-model problems contained a modal auxilliary such as "could" or "may." Only 2% of the one-model problems elicited such conclusions, and this difference is reliable because not a single subject went in the opposite direction ($p = .5^8$). Of the conclusions erroneously drawn to the invalid multiple-model problems, 20% were also modal.

Critics sometimes suggest that one-model problems might be easier because there are more potentially correct, although logically weaker, conclusions to them than to valid multiple-model problems. Hence, subjects might do better with one-model problems merely because they are guessing among the possible conclusions that relate the two end terms. In fact, the results of this experiment clearly rebut the argument. The correct conclusions to the one-model problems were without exception of the sort

None of the A is in the same place as any of the C,

or its logical equivalents. In theory, there could be logically weaker conclusions, such as "Some of the A are not in the same place as some of the C." The subjects never drew any such conclusions to the one-model problems.

The results of Experiment 2 again confirm the model theory: There was a marked effect of the number of models on the difficulty of drawing valid conclusions. Because all of the problems are based on affirmative and negative premises, there is no way to account for the results purely in terms of the difficulty of understanding negative assertions. There is, however, one other possible alternative explanation. The one-model problems are based on premises that have no scope ambiguities, whereas the valid multiple-model problems include an affirmative premise that is potentially ambiguous in scope (e.g., "All the artists are related to some of the beekeepers"). In fact, such sentences are normally interpreted with the scope of the quantifiers following their surface order and cause no interpretative difficulty (see Johnson-Laird, 1970). Nevertheless, our final experiment was designed to examine this alternative explanation.

Experiment 3

We constructed a set of 16 doubly-quantified pairs of premises: one half of them were in the first figural arrangement, and the other half were in the second figural arrangement; likewise, one half of them contained two affirmative premises, and the other half contained one affirmative and one negative premise. Table 5 summarizes the eight problems in the first figural arrangement. Each problem was based on the relation "is related to." There were 10 one-model problems, 2 multiple-model problems with valid conclusions, and 4 multiple-model problems with no valid conclusions. The problems included premises with scope ambiguities and contained such premises in both the one-model and multiple-model problems; for example,

None of the A are related to any of the B

Some of the B are related to all of the C (one-model)

and

None of the A are related to any of the B

All of the B are related to some of the C (multiple-model).

Method

The subjects acted as their own controls and were asked to state in their own words what, if anything, followed from 16 different doubly-quantified pairs of premises. Of the problems 8 were in the first figural arrangement (see Table 5), and the remaining 8 were constructed from them by reversing the order of the premises to form problems in the second figural arrangement. All of the problems were based on the relation, "is related to." The nouns in the premises referred to hobbies and professions selected so as to obviate any a priori connections. Of the 8 problems, 4 consisted of two affirmative premises, and 4 consisted of one affirmative and one negative premise.

We devised one set of lexical contents and assigned it twice at random to the 16 problems; one half of the subjects were tested with one of the assignments, and the other half of the subjects were tested with the other assignment. There were four filler items that were based on a different relation ("taller than"), and they were presented at random intervals within the series of test trials.

The procedure was the same as before. We tested 14 members (11 women and 3 men) of the Applied Psychology Unit subject panel, whose ages ranged from 23 to 68 years. We eliminated 3 subjects—2 because they had failed to understand the instructions, and 1 because he had studied logic.

Table 5
Eight Problems Used in Experiment 3 and Their Status as One-Model or Multiple-Model Problems

Affirmative problems

1. All of the X are related to all of the Y
   All of the Y are related to all of the Z
   (Therefore, All of the X are related to all of the Z.)
   One-model problem.
2. All of the X are related to some of the Y
   All of the Y are related to all of the Z
   (Therefore, All of the X are related to all of the Z.)
   One-model problem.
3. All of the X are related to some of the Y
   All of the Y are related to some of the Z
   (Therefore, All of the X are related to some of the Z.)
   One-model problem.
4. All of the X are related to some of the Y
   Some of the Y are related to all of the Z
   (No valid conclusion interrelating X and Z.)
   Multiple-model problem.

Negative problems

5. None of the X are related to any of the Y
   All of the Y are related to all of the Z
   (Therefore, None of the X are related to any of the Z.)
   One model problem.
6. None of the X are related to any of the Y
   Some of the Y are related to all of the Z
   (Therefore, None of the X are related to any of the Z.)
   One model problem.
7. None of the X are related to any of the Y
   All of the Y are related to some of the Z
   (Therefore, None of the X are related to some of the Z.)
   Multiple-model problem.
8. None of the X are related to some of the Y
   Some of the Y are related to all of the Z
   (No valid conclusion interrelating X and Z.)
   Multiple-model problem.

Note. A further eight problems were constructed by presenting the premises in the opposite order, that is, the second figural arrangement. The correct conclusions (in parentheses) were not presented to the subjects.

Results and Discussion

Table 6 presents the percentages of correct conclusions for each of the problems. There were 72% correct solutions to the one-model problems, and only 23% correct conclusions to the valid multiple-model problems (Wilcoxon's $T = 1$, $n = 10$, $p < .005$, one tailed). There were also 23% correct responses to the invalid multiple-model problems, which were reliably harder than the one-model problems (Wilcoxon's $T = 4$, $n = 10$, $p < .01$, one tailed).

It is instructive to compare the pair of one-model problems based on Item 6 (see Table 6) with the pair of valid multiple-model problems based on Item 7. The only difference between the premises of these two pairs of problems is in the order of the two quantifiers in one premise; both of these premises are, in principle, ambiguous in scope. Yet, there were 64% correct conclusions to the one-model problems, but only 23% correct conclusions to the valid multiple-model problems (Wilcoxon's $T = 2$, $n = 7$, $p < .025$, one tailed). This difference cannot be attributed to the polarity of the problems (both pairs contain a negative premise); it cannot be attributed to differences in the difficulty of coping with one sort of quantifier in comparison to another (the quantifiers are identical); and it cannot be attributed to differences in scope ambiguity (both pairs contain one scope ambiguous premise). The only plausible explanation is that the easy problems call for just one model to be constructed, but the difficult problems call for multiple models to be constructed.

Experiment 3 once again supported the model-based theory. The results failed to show that the polarity of premises—whether they are affirmative or negative—has a reliable effect on the difficulty of inferences. The subjects made 58% correct responses to the affirmative problems and 49% correct re-

Table 6
Percentages of Correct Responses to the Problems in Experiment 3

Variable	Problem			
	Affirmative			
	All–all All–all	All–some All–all	All–some All–some	All–some Some–all
No. of models	One	One	One	Multiple[a]
Item no.	1	2	3	4
Figure				
First	82	82	55	19
Second	82	73	64	9
	Negative			
	None–any All–all	None–any Some–all	None–any All–some	None–some Some–all
No. of models	One	One	Multiple	Multiple[a]
Item no.	5	6	7	8
Figure				
First	91	55	27	36
Second	64	73	19	27

[a] No valid conclusion.

sponses to the negative problems; this difference was not statistically significant (Wilcoxon's $T = 11.5$, $n = 9$, $p > .05$).

A Procedure for Constructing Models of Multiply-Quantified Assertions

Granted that the model-based theory appears to predict performance, we will consider in more detail how the theory actually works. One of its advantages, unlike a theory based on formal rules, is that it gives an account of comprehension—the construction of models—and reasoning itself depends merely on formulating conclusions and searching for counterexamples. Most of the theoretical work is done once there is an analysis of the meanings of expressions (i.e., the mapping in either direction from expressions to models). It is therefore easy to extend the theory to a new domain: Given the meanings of the expressions in the domain, the same general procedures for constructing and manipulating models apply (Johnson-Laird & Byrne, 1989). Indeed, once such a system has access to the meanings of quantifiers, relations, and other terms, it can construct models of premises, formulate conclusions based on them, and search for counterexamples to such conclusions. Moreover, the logical properties of these terms, such as the transitivity and symmetry of a relation, will emerge from the process without any need to use explicit statements of these properties in the form of meaning postulates. We demonstrate first how the logical properties of a relation can emerge from its meaning, and then how the meanings of quantifiers can be represented so that the resulting models yield inferences.

In order to understand an assertion, such as

Alicia is in the same place as Bill,

the interpretative system needs to have a grasp of the meaning of the relation. This meaning will enable it to construct a model of the situation; for example,

| Alicia Bill |

and to verify that the relation holds within such models. What is therefore needed is a representation of the contribution that the relation makes to the truth conditions of an assertion, that is, how the world (or rather a model of it) would have to be for the assertion to be true. Thus, the meaning of the relation is a specification that can enter into the semantic representation of a sentence. It ensures that when this representation is used to construct a model, the two arguments of the relation are represented in a part of the model corresponding to one place. Plainly, certain notions must be taken for granted in this specification of truth conditions: These are the primitives of the system, which are not normally expressible in the language under analysis. In writing computer programs based on this idea, one treats the meaning of the relation as a fragment of code that can be used (when it has been combined with the code corresponding to the relevant noun phrases) to construct and evaluate models (see Johnson-Laird, 1983, chap. 11). Thus, the code for "in the same place as" will lead to the construction of models like the aforementioned one.

The meaning of a relation is *not* the same as its logical properties. The logical properties govern the implications of assertions containing the relation; for example, "Alicia is in the same place as Bill" implies that "Bill is in the same place as Alicia." The meaning, however, governs the truth or falsity of an assertion containing the relation with respect to models of the world. Moreover, there are many relations that, like "in the same place as," are transitive and symmetric (e.g., "equal in height to"). Hence, if the interpretative system knows only that a relation has these logical properties, it will not know which particular relation is referred to or how to interpret it.

There is a strong moral to be drawn: The logical properties of a term emerge from its meaning as soon as it is put to use in constructing and evaluating models (Johnson-Laird, 1983). The two premises, "Alicia is in the same place as Bill" and "Bill is in the same place as Chris" yield the model

| Alicia Bill Chris |

from which it follows that Alicia is in the same place as Chris. No alternative model of the premises refutes this conclusion, and so it is valid. A transitive inference has therefore emerged from the meaning of the relation.

The construction of models also depends, of course, on the meanings of the quantifiers in the premises. They too must contribute to the semantic representation of a sentence that is used to guide the construction or evaluation of a model. Any theory of the interpretation of multiply-quantified sentences must allow for the occurrence of an indefinite number of quantifiers, even within the same noun phrase; for example, Some of the relatives of every employee of all the friends of . . . ,

and so the interpretation of quantifiers is likely to depend on a mechanism that operates compositionally (i.e., building up the interpretation of an expression from the meanings of its constituents and the syntactic relations among them). We assume that the system works on a rule-by-rule basis in which for each syntactic rule, there is a structural semantic principle for assembling the semantic representation that will guide the construction of models. A comparable assumption is standard in grammars based on Montague's work (see Partee, 1975), but the "possible worlds" semantics of such theories is not intended to be psychologically realistic because the meaning of each sentence corresponds to an infinite number of possible worlds.

In the lexicon, each word has a lexical entry that specifies its contribution to the truth conditions of assertions. These meanings are combined by semantic principles that take into account the syntactic relations between the constituents of sentences, and the resulting information is used to assemble a program that constructs a model of the sentence. The program depends on three principal components: (a) a knowledge of the meaning of the relation expressed in the sentence, (b) a knowledge of the meanings of quantifiers, and (c) an ability to use the procedure for each quantified phrase—essentially a "loop"—in constructing an appropriate overall program.

The meaning of a quantifier is, in essence, the raw material for a loop that is used in the model-building (or model-evaluating) program. Thus, the universal quantifier "all" needs a loop that constructs (or evaluates) values for each item in the set, whereas the existential quantifier "some" constructs (or evaluates) values for some arbitrary number of the members of the set. It will be easier to understand these principles by way of simple examples of the construction of models.

Given the assertion,

Every Avon letter is in the same place as some Bury letter,

the program enters a loop corresponding to the interpretation of the first noun phrase (which, for convenience, we refer to as Loop A). On first entering this loop, it chooses an arbitrary size for the universally quantified set (e.g., three Avon letters). Because the loop is based on a universal quantifier, it is going to ensure that each member of the set satisfies the basic relation expressed by the predicate of the sentence, x is in the same place as y. Sometimes, as we shall see, there is an existing value for x or y, which places constraints on the process of selection. Here, at the start of the process, there is no such constraint, and so Loop A merely ensures that an Avon letter is put into a particular place:

| a |

The program then enters the loop for the second quantified phrase (Loop B). Because the loop is based on the quantifier "some", it will ensure that at least some members of the set satisfy the basic relation, and so it selects an arbitrarily sized subset (e.g., two Bury letters) that are going to satisfy the basic relation. The remaining items in the set will ultimately be represented as *not* satisfying the basic relation, but, if one follows the logical principle that "some" signifies "at least some and possibly all," they will be represented as optional items. Loop B selects a member of the subset, y, and puts it into the model in a way that satisfies the basic relation, which now has a value for its first variable, *a* is in the same place as y:

| a b |

The process continues to cycle through Loop B until all of the members of the subset have been inserted into the model:

| a b b |

Next, the program returns to Loop A to select the next item of the set to satisfy the basic equation, which has been reset to, x is in the same place as y:

| a b b | a |

The program reenters Loop B to assign an arbitrarily sized subset (in this case, one Bury letter) to the same place:

| a b b | a b |

The process continues in the same manner until all of the Avon letters have been inserted into the model. When this task has been completed, the optional Bury letters that do not satisfy the basic relation are added to the model. The final result will be the following sort of model:

| a b b | a b | a b b | ob |

The interpretation of the negative quantifier, "none," is equivalent to that of the universal quantifier, except that the relation is reset to be negative. Thus, for example,

None of the Avon letters is in the same place as any of the Bury letters,

calls for two universal loops (the quantifier "any" can here be interpreted as a universal) and a negated basic relation, x is not in the same place as y.

The order of the loops depends on the scope of the quantifiers: The first loop corresponds to the quantified phrase with the largest scope, the second loop corresponds to the quantified phrase with the next largest scope, and so on. Thus, consider the interpretation of

Some Bury letters are in the same place as every Avon letter,

in which the scope is taken to correspond to the surface order of the quantifiers. The program consists in the same two loops as the first example, but they are in the opposite order. Hence, the program begins with Loop B and selects an arbitrarily sized subset of the Bury letters (e.g., two letters). It inserts the first one into the model,

| b |

and enters Loop A. On first entering this loop, it chooses an arbitrary size for the universally quantified set (e.g., four Avon letters). The loop iteratively ensures that each of these y's satisfies the basic relation, *b* is in the same place as y:

| b a a a a |

Next, the program returns to Loop B to select the next (and in this case, final) item in the subset. Because Loop A is for a universal quantifier, the basic equation has not been reset, and so the new Bury letter, x, has to satisfy the relation, x is in the same place as *a*:

| b b a a a a |

Finally, the optional Bury letters that do not satisfy the relation are added to yield the final model:

| b b a a a a | ob ob |

Strictly speaking, both this sentence and the first example are ambiguous in scope, but we have now illustrated both of their possible interpretations. The program must allow for scope ambiguities, and we assume that they exist unless they are explicitly ruled out by the choice of quantifiers. There is also another sort of ambiguity that arises with certain universal quantifiers, particularly "all." They can be used both distributively (as in the case of "every" in the preceding examples) or collectively (i.e., in the sense of "all together"). This interpretation is highly salient for

> All the Avon letters are in the same place as some of the Bury letters,

and yields the following sort of model:

| a a a b b | ob ob |

The collective interpretation requires a modification in the loop for the universal quantifier that ensures that it is exhaustively iterated before control passes to the loop below.

In general, the construction program can be assembled as the sentence is parsed because it is necessary only to assemble loops that correspond to the appropriate quantifiers. The overall principle is that when a loop is satisfied, control passes to the loop below. But, when all the loops below have been satisfied (or the loop is the bottommost one), control passes to the loop above. And, if there is no loop above (the loop is the topmost one), then the program halts with the final model. Hence, once the three main components are specified—the interpretation of the quantifiers (the core of the loop), the interpretation of the relation (a component of each loop), and the principle by which loops are put together—it is possible to assemble a program for constructing a model of any multiply-quantified assertion.

The procedure for formulating conclusions takes as its input one or more models and looks for the appropriate quantifiers to characterize a relation between terms that are not explicitly related in any premise. Given a relevant pair of terms, A and C, and a model, the procedure establishes whether the relation between them is affirmative or negative. It then works down through the hierarchy of the six possible interrelationships in terms of their semantic strength. The strongest possible affirmative relationship is the following:

1. $(\forall A)(\forall B)ARB$: All the A are in relation, R, to all the B.

Two slightly weaker ones are

2. $(EA)(\forall B)ARB$:
   Some of the A are in relation, R, to all the B.
3. $(EB)(\forall A)BRA$:
   Some of the B are in relation, R, to all the A.

Still weaker, respectively, are the following:

4. $(\forall B)(EA)BRA$:
   All the B are in relation, R, to some of the A.
5. $(\forall A)(EB)ARB$:
   All the A are in relation, R, to some of the B.

And, finally, the weakest possible relationship is

6. $(EA)(EB)ARB$:
   Some of the A are in relation, R, to some of the B.

There is a similar hierarchy for the six possible negative interrelationships. The procedure seeks to establish the strongest conclusion that is true in all of the models of the premises; it will respond "no valid conclusion" if no description holds across all the models.

The search for alternative models of the premises to refute a putative conclusion could, in theory, be carried out in a purely random way, provided that the same models are not sampled more than once. A random alteration can be made to a model and, if the result is still consistent with the premises, the conclusion can be evaluated with respect to it. A model is finite, and there are only a finite number of alterations that can be made to it. Hence, sooner or later, they will have been exhausted. Although ordinary individuals seem not to possess any simple deterministic search procedure, they perhaps do not search entirely at random, either. At present, however, we have too little information to characterize the details of the search, which in any case may differ from one individual to another.

In summary, once a compositional semantics has been specified for the quantifiers and relations, then it can be used to construct models on the basis of the meaning of assertions, to formulate conclusions that hold in models, and to evaluate alternative models that may be counterexamples to putative conclusions. The same meanings can be used to control all of these processes.

General Discussion

There are three main views about the human inferential mechanism: It uses abstract formal rules, content-specific rules, or model-building procedures based on the meanings of premises. In reasoning, one could use any or all of these means, although it is difficult to see how experimental observations could falsify the logical union of the three sorts of theories. Other studies have shown, however, that the model-based theory suffices to account for the phenomena of syllogistic reasoning (Johnson-Laird & Bara, 1984a), two-dimensional spatial reasoning (Byrne & Johnson-Laird, in press-c), and reasoning with propositional connectives (Johnson-Laird, Byrne, & Schaeken, 1989). We have shown here that the theory can be extended naturally to reasoning with multiply-quantified premises. Content-specific rules are highly implausible as a basis for such inferences because people can make deductions that depend solely on the basis of quantifiers and connectives. The theory would merely pass on the explanatory burden to some (yet to be formulated) account of how such inferences could be encoded in content-specific rules. The real issue raised by multiple quantification is therefore between theories based on models and those based on formal rules.

Could one devise a rule-based theory to account for our results? In principle, rule-based theories have universal Turing machine power (see Jeffrey, 1981), and hence they are unlikely to be refuted once and for all by empirical observations. At present, there is no rule-based psychological theory that is powerful enough to cope with the multiply-quantified inferences that we used in our experiments (see, e.g., Braine, 1978; Braine, Reiser, & Rumain, 1984; Johnson-Laird, 1975; Rips, 1983), and we anticipate severe difficulties in extending existing theories so that they will explain the phenomena. We have explored this problem in two different ways.

Our first approach was to prove the conclusions of *all* of the valid inferences used in our experiments, and we used a predicate calculus based on the rules for connectives proposed by current psychological theories (e.g., Braine, 1978; Rips, 1983). As Table 1 shows, the derivation for a one-model problem of the form

None of the P is in the same place as any of the C.

All of the C are in the same place as all of the D.

Therefore, None of the P is in the same place as any of the D.

is 19 lines in length. The derivation for the valid multiple-model problem, which differs in only one quantifier:

None of the P is in the same place as any of the C.

All of the C are in the same place as some of the D.

Therefore, None of the P is in the same place as some of the D.

is also 19 lines in length, and 17 of the lines are identical. The entire propositional stage of reasoning is identical for the two deductions. All that differs is that the existential quantifier in the second problem is instantiated at the start of the derivation and therefore is re-introduced at the end of the derivation. Yet, as Experiment 2 showed, problems of the first sort yielded 67% correct conclusions, whereas problems of the second sort yielded only 16% correct conclusions. Even if one adopts a postulate of negative transivity,

If x is not in the same place as y, and y is in the same place as z,

then x is not in the same place as z,

there is no difference between the two derivations though they are both shorter. Conversely, the one-model problem (1) in Table 6 has a 14-line derivation, whereas the one-model problem (6) has a 19-line derivation, but there was no reliable difference in difficulty between them. In short, the lengths of derivations predict differences in difficulty that are not observed, and fail to predict differences that are observed (see also Byrne & Johnson-Laird, in press-a).

Our second approach to developing a formal theory was based on the following consideration. A formal theory has to manipulate a representation of the logical form of premises, which is essentially a syntactic representation. Hence, what is needed is a linguistic difference between the two sorts of problem (one model vs. valid multiple model) that somehow leads to a much longer or more complex derivation for the multiple-model problems. The problem shown in Table 1 differs from its multiple-model variant only by one quantifier. The second premise of the multiple-model problem contains an existential quantifier:

All of the C are in the same place as some of the D.

One might argue that this premise calls for the use of different and more complicated rules. Unfortunately, any attempt to explain the phenomena in this way is doomed because premises of exactly this form occur in several one-model problems that caused no great difficulty to our subjects (see, e.g., Problem 2 in Table 2, and Problem 3 in Table 5).

We have searched for some other relevant feature that might account for the difference between the two sorts of problems. Our search has failed. The difference in difficulty between one-model and multiple-model problems seems unlikely to be a result of any of the following factors:

1. A response-priming effect, that is, a match between the form of the correct conclusion and the form of one of the premises (ruled out by Experiment 1, in which there were valid multiple-model problems that had such a match).

2. The affirmative or negative polarity of the premises (ruled out by Experiment 2, in which all of the problems contained one affirmative and one negative premise).

3. The greater number of logically distinct conclusions counting as valid for the one-model problems than for the valid multiple-model problems (ruled out by Experiment 2, in which the subjects drew only one sort of conclusion—the logically strongest—to the one-model problems).

4. The difficulty of "some," which might require longer derivations or rules that are more difficult to use (ruled out by Experiment 3, which used "some" in both sorts of problems).

5. Ambiguity in the scope of quantifiers (ruled out by Experiment 3).

6. An "atmosphere" effect, that is, a correspondence between the quantifiers in the correct conclusion and the quantifiers in the premises (ruled out by Experiments 1 and 3).

7. The difficulty of certain premises as a whole (ruled out by Experiment 3, in which a one-model problem and a multiple-model problem were identical apart from the order of the quantifiers in the second premise, but these two premises occurred in other problems, in which they give rise to opposite effects).

Hence, the likelihood of a syntactic factor accounting for our results is remote, and rule-based theories are, by definition, insensitive to other factors. Such arguments are always open to the criticism that they have ignored alternatives that would survive refutation by the data. We believe that we have considered all of the alternatives that meet the psychological considerations adduced in the construction of existing rule-based theories.

Those who doubt this claim have, of course, a simple way to refute it: They have only to construct a rule-based theory that does account for our results.

In contrast, the phenomena are predicted by the theory based on mental models. The subjects found it reliably easier to draw inferences from premises that were consistent with only one model than from those that were consistent with more than one model. Likewise, they were slightly impeded, as the theory predicts, if there was a clash between the quantifiers of the two occurrences of the middle term. Where more than one model could be constructed, they made a small but reliable use of modally qualified conclusions, and their errors tended to be conclusions consistent with a proper subset of possible models. Both of these phenomena are singularly difficult to explain on the basis of formal rules.

Contrary to what many theorists have supposed, including ourselves at one time, there is a fundamental peculiarity about theories of reasoning based on formal rules. They propose that ordinary individuals, having gone to the trouble of understanding the premises, base their inferences on something other than a full, rich, semantic interpretation. They are supposed instead to reason from an abstract logical skeleton that would require extra pains to derive. Moreover, so much reasoning in daily life is not even intended to be valid, that theories that can account only for attempts to think validly are handicapped from the start. Yet any semantic account of reasoning, such as the theory of mental models, must contain components that carry out the equivalent work of rules of inference. A central assumption of our theory is that models never contain variables, and indeed the work of instantiation is merely part of the normal process of comprehension. Universal quantifiers are instantiated by sets of mental tokens that are treated as exhausting the relevant set; existential quantifiers are similarly instantiated by sets of mental tokens, except that there are optional items that fail to satisfy the conditions of the assertion. One of the consequences of this distinction is that the choice of quantifier in one premise can affect the number of models that can be constructed for the set of premises as a whole. Another consequence is that the procedures that formulate putative conclusions will generalize any relation using the appropriate quantifier. The model-based theory has separate procedures that operate, in effect, to instantiate and to generalize, but because these components operate differently from the rules used in the predicate calculus, they are able to predict the relative difficulty of inferences.

The model-based theory is readily extendible to deal with nonstandard quantification and to account for such inferences as

All A are related to most B.

All B are related to most C.

Therefore, All A are related to most C.

Once one has an account of the semantics of "most," which is relatively straightforward, the same apparatus of constructing and evaluating models can be applied directly. The theory based on mental models is relatively simple to refute: It predicts that whenever the meaning of premises supports more than one model and the models need to be considered to reach a valid conclusion, the inferential task will be harder. This prediction has now withstood empirical testing for reasoning with a variety of multiply-quantified premises.

References

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Boolos, G. (1984). On 'syllogistic inference'. *Cognition, 17,* 181–182.

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

Braine, M. D. S., & Rumain, B. (1983). Logical reasoning. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology; Vol. 3, Cognitive development* (4th ed, pp. 263–340). New York: Wiley.

Braine, M. D. S., Reiser, B. J., & Rumain, B. (1984). Some empirical justification for a theory of natural propositional logic. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 18, pp. 313–371). New York: Academic Press.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition, 31,* 61–83.

Byrne, R. M. J., & Johnson-Laird, P. N. (1989). *On the spontaneous use of propositional connectives.* Manuscript submitted for publication.

Byrne, R. M. J., & Johnson-Laird, P. N. (in press-a). Models and deduction. In K. Gilhooly, M. T. G. Keane, R. Logie, & G. Erdos (Eds.), *Lines of thought: Reflections on the psychology of thinking.* London: Wiley.

Byrne, R. M. J., & Johnson-Laird, P. N. (in press-b). Remembering conclusions we have inferred: What biases reveal. In J.-P. Caverni, J.-M. Fabre, & M. Gonzalez (Eds.), *Cognitive biases.* Amsterdam: North-Holland.

Byrne, R. M. J., & Johnson-Laird, P. N. (in press-c). Spatial reasoning. *Journal of Memory and Language.*

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

Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics.* New York: Harcourt Brace Jovanovich.

Cooper, R. (1983). *Quantification and syntactic theory.* Dordrecht, Holland: Reidel.

Erickson, J. R. (1974). A set analysis theory of behavior in formal syllogistic reasoning tasks. In R. Solso (Ed.), *Loyola Symposium on Cognition* (Vol. 2, pp. 305–330). Hillsdale, NJ: Erlbaum.

Evans, J. St. B. T. (1982). *The psychology of deductive reasoning.* London: Routledge & Kegan Paul.

Gardner, M. (1958). *Logic, machines and diagrams.* New York, Dover.

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

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

Harman, G. (1972). Deep structure as logical form. In D. Davidson & G. Harman (Eds.), *Semantics of natural language* (pp. 25–47). Dordrecht, Holland: Reidel.

Hewitt, C. (1971). Description and theoretical analysis (using schemas) of PLANNER: A language for proving theorems and manipulating models in a robot. Unpublished doctoral thesis, Massachusetts Institute of Technology.

Hornstein, N. (1984). *Logic as grammar: An approach to meaning in language.* Cambridge, MA: Bradford Books, MIT Press.

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

Jeffrey, R. (1981) *Formal logic, its scope and limits* (2nd ed). New York: McGraw-Hill.

Johnson-Laird, P. N. (1970). The interpretation of quantified sentences. In G. B. Flores d'Arcais & W. J. M. Levelt (Eds.), *Advances in psycholinguistics* (pp. 347–372). Amsterdam: North-Holland.

Johnson-Laird, P. N. (1975). Models of deduction. In R. J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults* (pp. 7–54). Hillsdale, NJ: Erlbaum.

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

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

Johnson-Laird, P. N., & Bara, B. G. (1984b). Logical expertise as a cause of error: A reply to Boolos. *Cognition, 17,* 183–184.

Johnson-Laird, P. N., & Byrne, R. M. J. (1989). *Only* reasoning. *Journal of Memory and Language, 28,* 313–330.

Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1989). Reasoning by rule or model: The case of propositional inference. Mimeo, MRC Applied Psychology Unit, Cambridge.

Johnson-Laird, P. N., & Wason, P. C. (1977). *Thinking: Readings in cognitive science.* Cambridge, England: Cambridge University Press.

Keenan, E. L. (1971). Quantifier structures in English. *Foundations of Language, 7,* 255–284.

Lakoff, G. (1972). Linguistics and natural logic. In D. Davidson & G. Harman (Eds.), *Semantics of natural language* (pp. 545–665). Dordrecht, Holland: Reidel.

May, R. (1985). *Logical form: Its structure and derivation.* Cambridge, MA: MIT Press.

Newell, A. (1973). Production systems: Models of control structures. In W. G. Chase (Ed.), *Visual information processing* (pp. 463–526). New York: Academic Press.

Newell, A. (1981). Reasoning, problem solving and decision processes: The problem space as a fundamental category. In R. Nickerson (Ed.), *Attention and performance.* (Vol. 8, pp. 693–718). Hillsdale, NJ: Erlbaum.

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

Osherson, D. (1975). Logic and models of logical thinking. In R. J. Falmagne (Ed.), *Reasoning: representation and process in children and adults* (pp. 81–92). Hillsdale, NJ: Erlbaum.

Partee, B. H. (1975). Montague grammar and transformational grammar. *Linguistic Inquiry, 6,* 203–300.

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

Scheffé, H. A. (1960). *The analysis of variance.* New York: Wiley.

Wason, P. C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology, 11,* 92–107.

Wason, P. C., & Johnson-Laird, P. N. (1972). *The psychology of reasoning: Structure and content.* London: Batsford.

Received August 12, 1988
Revision received February 6, 1989
Accepted March 3, 1989 ■

1990 APA Convention "Call for Programs"

The "Call for Programs" for the 1990 APA annual convention will be included in the October issue of the *APA Monitor.* The 1990 convention will be held in Boston, Massachusetts, from August 10 through August 14. Deadline for submission of program and presentation proposals is December 15, 1989. This earlier deadline is required because many university and college campuses will close for the holidays in mid-December and because the convention is in mid-August. Additional copies of the "Call" will be available from the APA Convention Office in October.