Spatial Reasoning

RUTH M. J. BYRNE AND P. N. JOHNSON-LAIRD
MRC Applied Psychology Unit, Cambridge CB2 2EF, England

We carried out two experiments to investigate how people reason about the spatial relations among objects. The experiments were designed to test a theory of spatial inference based on mental models. The theory predicts that problems requiring only one model of the spatial layout to be constructed should be easier than those requiring more than one model to be constructed, even when the multiple-model problems have valid conclusions. A contrasting theory, based on rules of inference, predicts that problems based on fewer inferential steps should be easier than problems based on more steps. The first experiment held constant the number of inferential steps specified by the inference-rule theory, but varied the number of models required to make a valid response: the one-model problems were reliably easier than the problems requiring more than one model. The second experiment contrasted opposing predictions from the two theories, and once again the results supported the model-based theory. © 1989 Academic Press, Inc.

In daily life, many inferences depend on relations. If, for example, a law asserts that dog owners must pay a tax, and Alicia owns a labrador, then one can readily infer that she must pay the tax. Strictly speaking, the inference is based on a missing premise that establishes the relation of class-inclusion between labradors and dogs. Ordinary individuals, unlike logicians, neither notice that the premise is missing nor categorize the inference as invalid on these grounds. They know that labradors are dogs and they automatically use this knowledge in drawing the conclusion. The relation of class inclusion is ubiquitous, perhaps because it gives rise to transitive inferences, and so it plays a central part in theories of semantic memory (e.g., Collins & Quillian, 1969; Rips, Shoben, & Smith, 1973; Miller & Johnson-Laird, 1976).

Comparative adjectives, such as "taller than," are another source of relations that yield transitive inferences. Thus, given the premises

Cathy is taller than Linda
Linda is taller than Mary,

it is easy to deduce that

Cathy is taller than Mary,

it is very much harder, however, to determine exactly how people make these so-called "three term series" inferences. Their introspective reports have little to say about underlying mental processes (Evans, 1989), and there has been considerable controversy about the correct interpretation of the various experimental results (see, e.g., Hunter, 1957; Huttenlocher, 1968; Clif, 1969; Potts, 1972; Griggs & Osterman, 1980; Sternberg, 1981; Richardson, 1983). For example, the difference between "negative-equals" such as "x is not as good as y" and affirmative premises is consistent with various theories (see Evans, 1982 for a review). The problem of identifying the inferential mechanism is made still harder by the fact that people can use different strategies in solving these problems (Sternberg & Weil, 1980; Egan & Grimes-Farrell, 1982). These difficulties have left open the nature of the underlying processes of deductive reasoning, and it has been modelled in two quite distinct ways.

The first sort of theories are based on proof-theoretic methods, and they post-
Spatial reasoning consists of formal rules of inference that are used to derive conclusions in a quasi-syntactic way (see, e.g., Inhelder & Piaget, 1958; Osherson, 1975; Shanks, 1978; Rips, 1983). The first step according to such theories is to translate the premises into a language-like mental representation that makes their logical form explicit. Inference rules can then be applied to these representations in order to derive a conclusion. A rule must correspond to the form to which it is applied (i.e., the premises, or assertions, derived for them). Moreover, the premises fail to match the logical form of the rule, then the rule cannot be applied to them. If the conclusion is to be derived, it is mandatory either to find an alternative rule or to make a preliminary inference to an assertion that does have the required logical form.

The logical properties of relations can be captured by general schemas, such as the following one for transitivity (where “R” denotes a relation):

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

If transitive relations, such as “taller than,” must be tagged in some way to indicate that this schema applies to them (e.g., Shiffrin, 1967). Alternatively, each individual transitive relation can have its own tagging postulate, e.g.,

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

Experiments have yet to determine whether such postulates are represented in the mind (but see, e.g., Cheng, Holyoak, Nobeck, & Oliver, 1986; Pobod, 1977; Kanch, 1974).

The second sort of theories are based on model-theoretic methods, and they postulate semantic procedures for generating valid conclusions. Such theories applied to three-term series problems propose that people use their understanding of the premises to imagine the state of affairs they describe—they construct a scenario, image, or spatial array of the situation, which they then use as a basis for reasoning (see, e.g., De Soto, London, & Handel, 1965; Huttenlocher, 1968; Erickson, 1974; Sternberg, 1985).

An unfortunate fact about the study of three-term series problems is that it has not yielded any clear answer to the question of which of these two sorts of theory—the proof-theoretic or the model-theoretic—is correct. Moreover, no such answer seems likely to be forthcoming from studying these problems. They are so simple that the two theories appear to mimic one another in their predictions, and so it may be impossible to design a crucial experiment. A more promising domain concerns a richer sort of relational reasoning that has been hitherto neglected: spatial inference.

Previous studies of how people understand spatial descriptions, such as

The knife is in front of the vase.
The vase is on the left of the glass.
The glass is behind the dish.

have established that subjects tend to envisage such layouts as symmetric with approximately equal distances between adjacent objects (Ehrlich & Johnson-Laird, 1982) and that they can represent the content of such descriptions in two distinct ways, one closer in structure to the linguistic form of the description itself and the other to the structure of the state of affairs that is described (Mani & Johnson-Laird, 1982). In this paper, we will consider how people make inferences from spatial descriptions.

Spatial reasoning by Rules or Models

Spatial reasoning is ubiquitous in our everyday interactions with the world: it underlies our ability to plan a route, to locate entities, and to envisage objects from de-
scriptions of their arrangement. A simple sort of spatial inference depends on premises that describe a one-dimensional layout of objects, such as:

A is on the right of B
C is on the left of B
Hence, A is on the right of C.

A meaning postulate for making this inference is:

If x is on the right of y, and z is on the left of y, then x is on the right of z

and inference-rule systems containing such postulates have been proposed for human reasoning (e.g., Hagerty, 1983; Ohlsson, 1981, 1984, 1988). A second sort of one-dimensional problem depends on premises such as:

B is on the right of A
C is on the left of B

which do not yield any valid conclusion about the relation between A and C. These premises fail to match any of the formal rules in the system and hence the response of no valid conclusion can be made (although, strictly speaking, a conjunction of the two premises could be inferred). An inference-rule theory might thus predict that people should find it easier to make an inference from the first problem than from the second problem.

The model-based theory makes the same prediction for these simple problems, but for different reasons. According to this theory, in order to understand the premise:

A is on the right of B,

it is necessary to grasp the meaning of the relational predicate, "on the right of." This knowledge, together with compositional principles that combine meanings according to syntactic relations, can be used to construct a representation of a specific situation that satisfies the premise.

A computer program implementing the theory builds up two-dimensional spatial arrays containing tokens in the positions described by such premises (see Johnson-Laird, 1983, Chap. 11). Given an initial premise of the form:

A is on the right of B,

it constructs an initial minimal array that satisfies the truth conditions of the premise:

B A

assuming a viewpoint such as an observer would have standing in front of the object. The meaning of "on the right of" specifies an appropriate increment to one Cartesian coordinate while holding the other coordinate constant. The program checks each subsequent premise to determine whether it makes reference to any items already in the array. Given the second premise:

C is on the left of B,

it accordingly finds B in the model, and using its representation of the meaning of "on the left of," it inserts C into the array in an appropriate place:

C B A

A further assertion:

A is on the right of C

contains only referents that are already in the array, and so a procedure is called to verify the premise—again, using the meaning of "on the right of." The verification procedure returns the value "true" for the assertion because its truth conditions are satisfied by the items in the array. This evaluation, in turn, elicits a falsification procedure that searches for an alternative model of the previous premises that falsifies the current assertion. The procedure does not use inference rules sensitive to the logical form of the premises, but seeks to rearrange the model without violating the meaning of the premises. If there is such an alternative, then the current assertion provides new information by ruling out this possibility. In the present case, however,
Hagert (1983) has proposed an ingenious inference-rule system for two-dimensional reasoning that uses the following sorts of one- and two-dimensional rules:

1. Left (x, y) & Front (z, x) → Left (front (z, x), y), where the right-hand side signifies "z is in front of x, all of which is on the left of y."
2. Left (x, y) & Front (z, y) → Left (x, front (z, y)), where the right-hand side signifies "x is on the left of z, which is in front of y."
3. Left (x, y) & Left (y, z) → Left (x, left (y, z)).
4. Left (x, y) → Right (y, x).
5. Left (front (x, y), z) → Left (x, z) & Left (y, z) & Front (x, y).
6. Left (x, front (y, z)) → Left (x, y) & Left (x, y) & Left (x, y) & Left (y, z).
7. Left (x, y) → − Right (x, y).
8. Right (x, y) → − Left (x, y).

The first premise in the problem is irrelevant, and the inference rules must be used to derive the relations between D, C, and B, and then those between D, E, and B, before there is sufficient information to use a rule to infer the relation between D and E. Thus, the proof proceeds by deriving the relation between D, C, and B:

5. C is on the left of B and D is in front of C [conjunction of premises 2 and 3].
6. D is in front of C, which is on the left of B [rule 5] which is equivalent to:
7. D is on the left of B, and C is on the
left of $B$, and $D$ is in front of $C$ [rule $e$ to 6].

Now, the relation between $D$, $E$, and $B$ can be derived:

8. $D$ is on the left of $B$ [conjunction elimination to 7].
9. $D$ is on the left of $B$ and $E$ is in front of $B$ [conjunction of 8 and premise 4].
10. $D$ is on the left of $E$ which is in front of $B$ [rule $b$ to 9] which is equivalent to:
11. $D$ is on the left of $E$, and $D$ is on the left of $B$, and $E$ is in front of $B$ [rule $s$ to 10].
12. $D$ is on the left of $E$ [conjunction elimination to 11], which is the required conclusion.

The following problem (Problem II):

$B$ is on the right of $A$
$C$ is on the left of $B$
$D$ is in front of $C$
$E$ is in front of $B$.

Hence, $D$ is on the left of $E$.

requires an identical derivation. The first premise is again irrelevant, and the remaining three are the same as those of Problem I. Hence, if people are using such rules, there should be no reliable difference in difficulty between Problem I and Problem II.

In contrast, the model-based theory predicts a difference in the difficulty of the two problems. The premises of Problem I yield the model:

$$CBA$$

$$DE$$

The relation between $D$ and $E$ can be established in a way similar to the evaluation of the assertion:

$$D$$ is on the left of $E$.

The procedure that revises models will not succeed in producing an alternative model that refutes this assertion, and so $D$ must be on the left of $E$. Thus, this set of premises can be classified as a single-model problem that supports a valid conclusion.

The premises of Problem II, however, are consistent with at least two distinct models. One model is

$$CAB$$

$$DE$$

which supports the conclusion

$$D$$ is on the left of $E$.

The falsification procedure succeeds in constructing an alternative model of the premises

$$ACB$$

$$DE$$

but both of these models support the same conclusion, and there is no alternative model that falsifies it. Hence, Problem II is a multiple-model problem that supports a valid conclusion. The model-based theory predicts that a problem which requires more than one model will be harder than one which requires only one model, and so it predicts that Problem II should be harder than Problem I.

Some two-dimensional problems do not yield a valid answer, e.g., Problem III:

$B$ is on the right of $A$
$C$ is on the left of $B$
$D$ is in front of $C$
$E$ is in front of $A$.

What is the relation between $D$ and $E$?

Both theories predict that these problems should be hardest of all. According to the inference-rule theory, they should be difficult because the response of "no valid conclusion" is forthcoming only after all possible derivations have been tried. According to the model-based theory, the premises of Problem III support two models:

$$CAB$$

$$ACB$$

$$DE$$

in which there is no common relation between $D$ and $E$. The theory predicts that this problem will be harder than the one-model Problem I. It also predicts that it will be harder than the multiple-model Problem II.
that supports a valid conclusion. Where
couple models support a valid conclu-
sion, a reasoner who in fact constructs just
e one will nevertheless draw the cor-
correlation. However, where the mul-
tiple models do not support a valid conclu-
sion, a reasoner who constructs just one
model will make an erroneous inference,
prove the only way to appreciate that
there is no valid conclusion is to consider
more than one model of the premises.

**EXPERIMENT I**

Because there had been no previous
study of two-dimensional spatial infer-
er, our principal aim was to find out
whether people were capable of making
such inferences. Evidence from three-term
problems suggested that subjects
would have no difficulty with one-
dimensional problems, and so we designed
an experiment to compare the ease of such
inferences with those based on two-
dimensional layouts. We naturally ex-
pected that the two-dimensional problems
yield be harder than the one-dimensional
ones.

Our second aim was to make a prelimi-
nary comparison between the inference-
rules and the model-based theories. We
designed three sorts of two-
dimensional problems: those that yield a
valid conclusion based on one model (such
as Problem I above), those that yield a valid
conclusion based on more than one model
(such as Problem II), and those that do not
yield a valid conclusion from their multiple
models (such as Problem III). The model-
based theory predicts that the one-model
problems should be easier than the multi-
model problems with a valid conclu-
sion; the inference-rule theory makes no
prediction. Both theories predict that
problems lacking a valid conclusion
would be hardest of all.

**Materials and methods.** The subjects acted
under their own controls and each received five
different sorts of problems: one-dimensional
problems with valid answers, one-dimensional
problems with no valid answers, two-
dimensional problems with valid answers,
and which, according to the model-based
theory, have one model, two-dimensional
problems with valid answers and which, ac-
according to the model-based theory, have
more than one model, and two-dimensional
problems with no valid answers. The one-
dimensional problems consisted of two pre-
mises and the two-dimensional problems
consisted of four premises. The two-
dimensional problems were in the following
form of orientation:

\[ A \quad B \quad C \quad D \quad E \]

Each subject received five, six, or seven
instances of each of the five types of prob-
lem, and each problem was presented with
different lexical materials about various do-
 mains concerning familiar objects, such as
cups and plates, drawn from the same se-
matic domain. These sets of objects were
assigned to the problems at random in order
to produce two sets of materials. Half the
subjects received one set and half the sub-
jects received the other set.

**Procedure.** The subjects were given the
instructions by way of an example. They
were told that the experimenter would read
aloud a description of the layout of some
objects, which they could imagine arranged
on the tabletop in front of them. They were
courted to listen attentively since the
description would be read only once, al-
though at a reasonable pace. They would
then be asked about the location of two of
the objects, e.g., “What is the relation be-
tween the cup and the plate?” In fact, the
two items in the question were always the
cardinal for the two-premise problems,
and the pair of items (the \(D\) and the \(E\) in
the examples above) that were not explicitly in-
terrelated in any premise of the four-
premise problems. Finally, the subjects
were told that for some of the problems
they might think there was not enough in-
formation to determine the relation between the two objects; in which case, they were to say so. Their responses were recorded on cassette recorder.

Subjects. The 15 subjects (11 women and 4 men) ranged in age from 19 to 55 years old. They were members of the APU subject panel, and they were paid £3 per hour for participating in the experiment, which lasted just under half an hour.

Results

The percentages of correct responses are presented in Table 1. Contrary to our expectations, the subjects coped as readily with the two-dimensional inferences (39% correct) as with the one-dimensional inferences (44% correct), and the difference between them was not reliable (Wilcoxon’s $T = 45.5, n = 15, p > 0.05$). However, as Table 1 shows, the results corroborated the model-based prediction: there were more correct responses to the valid problems that required one model (61% correct) than to the valid problems that required more than one model (50% correct) and this difference was reliable (Wilcoxon’s $T = 30.5, N = 15, p < 0.05$).

Finally the results corroborated the prediction common to the two theories: problems with a valid answer were easier than those with no valid answer. There were 59% correct responses overall to the problems with a valid conclusion, but there were only 18% correct responses to the problems without a valid conclusion (Wilcoxon’s $T = 1, N = 15, p < 0.01$). This difference was reliable for both the one-dimensional problems ($W = 1, N = 14, p < 0.01$) and for the two-dimensional problems ($W = 1, N = 15, p < 0.01$).

Discussion

The experiment showed that people can reason about the spatial relations between objects and do not seem to find the two-dimensional spatial layouts any harder than the one-dimensional layouts. Problems with a valid answer are evidently easier than problems with no valid answer. This phenomenon is consistent with both the inference-rule theory and the model-based theory, but it would also have occurred if subjects were loath to respond that there is no definite answer. This explanation seems implausible, especially given so striking a difference in accuracy, because subjects do quite well with syllogisms that do not support a valid conclusion (see, e.g., Johnson-Laird & Bara, 1984; but cf. Rumain, Connell, & Braine, 1983).

The result that people make more correct inferences from one-model problems than from multiple-model problems corroborates the model-based theory. Because these two sorts of problem require identical formal derivations, the result cannot be predicted by the inference-rule theory. The premises used in the derivations for the two sorts of problem are the same. In Problem I, the irrelevant first premise could be used to derive, en passant, an ultimately useless inference about the relation between A and C. The derivation for Problem I would then contain an extra step interrelating them. The derivation for Problem II would be unaltered because the rules cannot interrelate A and C. But, this difference in the derivation lengths would predict incorrectly that Problem I should be harder than Problem II.

We have established that two-dimensional reasoning is a feasible task for examining theories of reasoning, and that the inference-rule theory that we have considered fails to predict an observed difference.
Spatial Reasoning

It is predicted by the model-based theory that a more stringent test, however, needs to be applied to one model-based and model-based theory in direct opposition to each other on problems where the inference-rule theory predicts a difference in one direction, and the model-based theory predicts a difference in the opposite direction. Our second experiment makes such a comparison between the two theories.

**Experiment 2**

In this experiment we examined three sorts of two-dimensional inferences. First, problems of the following sort (Problem IV):

- A is on the right of B
- C is on the left of B
- D is in front of C
- E is in front of A

What is the relation between D and E?

Like Problem I, there is no premise that directly asserts the relation between the items to which D and E are directly related. Hence, the relation between them must be inferred. The first two premises of the problem:

- A is on the right of B
- C is on the left of B

must be used to make the following sort derivation:

- B is on the right of C [rule d to prove this]
- A is on the right of B and B is on the right of C [conjunction of 1 and 5]

A is on the right of C [transitivity rule]

The inferred relation, together with the fifth premise, now permits the same solution as Problems I and II.

Second, multiple-model problems, such as Problem II above, that have a shorter solution than Problem IV.

Third, multiple-model problems, such as Problem III above, that have no valid answers.

The model-based theory predicts that the one-model problems (IV) should be easier than the multiple-model problems with valid answers (II), whereas the inference-rule theory makes the opposite prediction. The one-model problems call for additional steps in their formal derivations over and above the derivations of the multiple-model problems and the inference-rule theory predicts that a conclusion requiring more inferential steps will be harder. Both theories predict, of course, that the problems with no valid answers should be hardest of all.

**Method**

**Design.** The subjects, acting as their own controls, carried out 18 inferences of three sorts: six one-model problems with valid answers, six multiple-model problems with valid answers, and six multiple-model problems without valid answers. Each problem was presented with different lexical materials. These lexical materials were assigned to the 18 problems at random in two different ways, and half the subjects received one assignment and half the subjects received the other assignment. The problems were presented in a different random order to each subject.

**Materials and procedure.** Four orientations formed the basis of the problems in each of three conditions:

$$
\begin{array}{ll}
(1) & (2) \\
A & B & C & E & D \\
D & E & A & B & C \\
(3) & (4) \\
D & C & A & E \\
B & B \\
E & A & C & D
\end{array}
$$

The six problems in a condition consisted in one of each of these four orientations, and two extra problems in two of the four orientations, one vertical and one horizontal.
This variation, and the fact that the two items in the question, e.g.,

What is the relation between the D and the E?

did not occur in fixed positions in the premises, made it difficult for subjects to predict which items would occur in the question.

Each of the sets of five objects making up the contents of a problem was based on high frequency one-syllable words drawn from the same semantic domain. No two words within such a set began with the same initial letter. The procedure was identical to the first experiment, except that the subjects were given a four-term series problem as a preliminary practice problem. Furthermore, in order to try to improve performance, each problem was read out twice to the subject before they were asked the question about two of the objects.

**Subjects.** The 18 subjects (16 women and 2 men) ranged in age from 24 to 56 years old. They were members of the APU subject panel, and they were paid £3 per hour for participating in the experiment, which lasted for about half an hour.

**Results and Discussion**

The one-model problems (70% correct responses) were easier than the multiple-model problems with a valid conclusion (46% correct responses), which in turn were easier than the multiple-model problems without a valid conclusion (15% correct responses, Page's \( L = 4.43, p < 0.05 \); Page, 1963). The difference between the one-model problems and the multiple-model problems with valid answers was highly reliable (Wilcoxon's \( T = 7, N = 14, p < 0.005 \)). This difference corroborates the model-based theory and runs counter to the inference-rule theory. Problems which require more models were harder than those which require only one model, but problems which require more inferential steps in their derivation were not harder than those which required fewer steps.

The subjects did not perform more accurately overall than those in the previous experiment, despite the repeated reading of the problems, and so their difficulties do not seem to be attributable to remembering the descriptions. Their performance did not improve with practice: they made as many correct inferences to their first six problems (47%) as they did to their next six problems (40%) and to their final six problems (45%; Page's \( L = 3.82, n.s. \)). Because the inferences depend on a two-dimensional layout, the chances of guessing the correct answer are approximately one in five: there are four obvious candidate spatial relations between the two objects, plus the response that there is no relation between them. Yet, there were only 15% correct responses to the multiple-model problems without a valid conclusion. These problems yielded 42% of responses in which the subjects interrelated the two objects on the dimension that was consistent with the premises (although the premises fail to support any definite relation on that dimension). Such responses are precisely what is to be expected if subjects are constructing only one of the possible models of the premises.

The errors on the other problems mainly consisted of conclusions that established the wrong relation. Where the subjects' conclusion is a relation orthogonal to the correct dimension, they have probably failed to build the correct model. Such errors should thus reflect the increasing difficulty of representing the premises; they occurred with 13% of one-model problems, but with 23% of multiple-model problem with a valid conclusion and 23% of multiple-model problems with no valid conclusion.

**General Discussion**

The results of both experiments show that it is easier to draw a valid spatial inference when a description corresponds to just a single layout as opposed to two or more distinct layouts. This phenomenon is
ously explained if people naturally reason about the state of affairs described by the premises, drawing a conclusion from a mental model, and searching for alternative models that might refute that conclusion.

Could our instructions to imagine the objects have caused our subjects to use an imagery-based strategy otherwise not to them? We used this instruction—only mentioned during the subjects’ instruction to the task—to avoid any problems in the interpretation of “on the right” or the left of,” and the other spatial terms. These expressions have two distinct senses: the “deictic” sense which depends on the speaker’s point of view, and an “indicative” sense which depends on the context. These objects were being described from a referent point of view. It seems unlikely that this instruction could be powerful enough to lead the subjects to adopt a purely unnatural reasoning strategy. Several authors lament the considerable difficulty of inducing reasoning strategies by explicit instructions (e.g., Dickson, 1978), while others deny a significant effect of imagery in reasoning (e.g., Richard-son, 1982). The critical feature of the mental model theory concerns the structure of representations used in reasoning—rather than that they should be represented as images.

Experimental evidence really decide between model-based and inference-rule based theories? There is little doubt that our two experiments fit the model-based theory better. But, could there be an alternative to the rule theory that does account for the results? One approach is worth de-

scribing because it illuminates the particular difficulties of such theories. It postulates such principles as:

If \( x \) is on the left of \( y \),
and \( w \) is in front of \( x \),
and \( z \) is in front of \( y \),
then \( w \) is on the left of \( z \).

The first premise in Problems I and II is irrelevant, but the application of this rule to the remaining premises yields the conclusion directly:

The \( D \) is on the left of the \( E \).

Because a large number of such rules would be necessary to deal with the full set of spatial relations, we can invoke abstract postulates of the same structure:

If \( x \) is related to \( y \) on one dimension,
and \( w \) is related to \( x \) on an orthogonal dimension,
and \( z \) has the same orthogonal relation to \( y \),
then \( w \) is related to \( z \) in the same way as \( x \) is related to \( y \).

This abstract rule has the advantage that it captures the content of a large number of specific rules, including the rule above because it applies to any configuration of the form

|          |
|          |
| \( x \)   |
| \( y \)   |
|          |
| \( w \)   |
| \( z \)   |

regardless of its rotation, or reflection. A system of such rules still makes the wrong predictions about relative difficulty. For example, the rule applies immediately to multiple-model problems, such as II, but fails to apply to one-model problems such as IV until the relation between \( x \) and \( y \) has been established. However, the rule might be revised so that it makes the required predictions (cf. Ohlsson, 1981, 1984, 1988). The resulting rule is more complex because it has to have the principle of transitivity built into it, i.e., the transitivity concerning \( x, v, \) and \( y \), in the following sort of layout:
The revised rule has the form:

If \( x \) is related to \( v \) on one dimension and \( v \) is related to \( y \) in the same way, and \( w \) is related to \( x \) on an orthogonal dimension, and \( z \) has the same orthogonal relation to \( y \), then \( w \) is related to \( z \) in the same way as \( x \) is related to \( y \).

The cost of such a revision is severe, since the rule fails altogether to match the premises of the multiple-model Problem II, or even the premises of the one-model Problem I. The theory does not predict that such inferences will be harder; it predicts that they will be impossible. Yet, if both the original and the revised form of the rule are retained, then the theory ceases to predict the differences between one-model and multiple-model problems.

Inference-rule theories almost certainly have the computational power of universal Turing machines, and so there is unlikely to be any empirical phenomenon that cannot in principle be described within their framework. Our results do not refute inference-rule theories but rather present a challenge to them. To save the approach, it is necessary to devise a system of formal rules that explains why one-model problems are easier than multiple-model problems. The strategy behind this study was indeed to examine inferences where the number of mental models does not correlate with the number of inferential steps according to a particular inference-rule theory. Where the two theories diverge in this way, we have shown that performance follows the predictions of the model-based theory. We have used the same strategy with similar success in all the main domains of deductive inference, including propositional reasoning (Byrne, 1989; Johnson-Laird, Byrne, & Schaecken, 1989), and reasoning with single and multiple quantifiers (Johnson-Laird & Byrne, 1989; Johnson-Laird, Byrne, & Tabbott, in press). In each of these domains, the model-based theory relies, not on formal rules of inference, but on processes that construct models, formulate conclusions from them, and search for alternative models to serve as counterexamples. The logical consequences of words, like those denoting spatial relations, are nothing more than the emergent properties of their meanings within the system.

ACKNOWLEDGMENTS

We are grateful to Ian Nimmo-Smith for statistical advice and to Sam Glucksberg and Mark Keen for critical readings of an earlier draft of this paper.

REFERENCES


