Psychological Review
2005, Vol. 112, No. 2, 468–493

0033-295X/05/$12.00 DOI: 10.1037/0033-295X.112.2.468

Reasoning About Relations

Geoffrey P. Goodwin and P. N. Johnson-Laird
Princeton University

Inferences about spatial, temporal, and other relations are ubiquitous. This article presents a novel model-based theory of such reasoning. The theory depends on 5 principles. (a) The structure of mental models is iconic as far as possible. (b) The logical consequences of relations emerge from models constructed from the meanings of the relations and from knowledge. (c) Individuals tend to construct only a single, typical model. (d) They spontaneously develop their own strategies for relational reasoning. (e) Regardless of strategy, the difficulty of an inference depends on the process of integration of the information from separate premises, the number of entities that have to be integrated to form a model, and the depth of the relation. The article describes computer implementations of the theory and presents experimental results corroborating its main principles.

Consider the following problem:

> Pat stood in the last borough elections.
>
> The borough is in the state of New Jersey.
>
> New Jersey had its last borough elections on Thursday.
>
> When did Pat stand in the borough elections?

You should have no difficulty in inferring that the answer is Thursday. You infer from the first two assertions that Pat stood in the last borough elections in New Jersey. You infer from this intermediate conclusion and the third premise that Pat stood in the borough elections on Thursday. Your inference is an example of reasoning from relations—both spatial relations (the borough is in New Jersey) and temporal relations (its last local elections were on Thursday). Your reasoning is often about relations, and they underlie many of your inferences in daily life. Psychologists have studied reasoning about relations for many years, but they disagree about the process (see Evans, Newstead, & Byrne, 1993, chap. 6, for a comprehensive review). Logicians, however, have analyzed the logical implications of relations and have shown how they can be captured in the *predicate calculus* (see, e.g., Jeffrey, 1981).

In this article, we present a general theory of relational reasoning. It aims to answer three questions. First, how are relations and their logical properties mentally represented? Second, what are individuals computing when they reason about relations? Third, what are the mental processes that carry out this reasoning? The first part of the article outlines the nature of relations and their logical properties. The second part reviews previous accounts of relational reasoning. The third part presents a model-based theory of what individuals do when they reason about relations and of how they carry out the process. The fourth part assesses this theory and other alternative accounts in the light of empirical evidence, including three new experiments from our laboratory. Finally, the article draws some general conclusions about relational reasoning.

What Are Relations?

Relations and Functions

The declarative sentence, "The boy is taller than the dog," asserts a relation, *taller than*, between two arguments, the boy and the dog, respectively. Hence, the sentence can be used to express a proposition that is either true or false. A relation is *satisfied* by those arguments that yield a true proposition. In logical terminology, the *extension* of a relation is the set of ordered entities that satisfy it, such as the ordered pair, the boy and the dog, which satisfy the relation *taller than* if the sentence above is true. In contrast, the *intension* of a relation is what it means, for example, what *taller than* means.

Relations can have any finite number of arguments. In an extensional analysis, an *n*-place relation is nothing more than a subset of the set of all possible orders of *n* entities from the *n* relevant sets. Hence, if the *n* relevant sets are

$$S_1 \quad S_2 \quad S_3 \ldots S_n,$$

then an *n*-place relation is a subset of the set of all possible orders in which the first member comes from S_1, the second member comes from S_2, . . . and the *n*th member comes from S_n (i.e., the Cartesian product of the sets). The sets, of course, need not be distinct: A relation such as *x loves y*, for example, can take both its arguments from the set of human beings. Two relations have the same extension if they both are satisfied by the same set of ordered entities. Yet they could have different intensions, for example, in a particular domain, *taller than* and *heavier than* could have the same extension, yet they have different intensions.

Predicates that have just a single argument, such as *happy* (as in "Pat is happy"), can denote properties, though some predicates

Geoffrey P. Goodwin and P. N. Johnson-Laird, Department of Psychology, Princeton University.

This research was supported by National Science Foundation Grant BCS-0076287, to study strategies in reasoning. We thank Nick Chater, Jonathan Evans, Graeme Halford, and Sun-Joo Shin for their criticisms of a draft of this article and Georg Jahn for his advice on preferred mental models. And for their helpful comments, we also thank Ruth Byrne, Sam Glucksberg, Uri Hasson, Cathy Haught, Louis Lee, Tom Ormerod, and Sanna Reynolds.

Correspondence concerning this article should be addressed to Geoffrey P. Goodwin, Department of Psychology, Princeton University, Princeton, NJ 08544. E-mail: ggoodwin@princeton.edu

with a single argument denote relations, such as *heavy* (as in "The book is heavy," which really means that it is heavy *for a book*). A logic of properties (monadic predicates) was first formulated in Aristotle's account of syllogisms, but this logic does not include inferences that hinge on relations. Likewise, a semantics of relations that is based solely on properties is not feasible. Such an account tries to treat, say, *father of* as a set of properties, such as human, adult, male, having-children. But it is unworkable because the composition of relations is not necessarily commutative; for example, your *father's mother* is not the same as your *mother's father* (see, e.g., Winkelmann, 1980). There is no way to capture the distinction if the underlying semantics is based solely on sets of properties, because operations on sets, such as union, complement, and intersection, cannot capture relations. Relations are irreducibly relational. When their arguments are quantified, as in "All the women are taller than some of the men," their logical analysis calls for the full power of the predicate calculus (see, e.g., Jeffrey, 1981).

A particular event can enter into many relations (see Davidson, 1967):

> I did it for my spouse in the bathroom at midnight on Christmas eve with a piece of buttered toast and a knife and fork for half an hour in a vigorous way . . .

Hence, relations are normally taken to hold between two or more arguments. But, in daily life, they seldom take more than three arguments, for example, *Pat gave Viv a book*. And most of our discussion focuses on binary relations, that is, relations between two entities.

A *function*, which is a special case of a relation, takes a finite number of arguments and relates them to a unique result. For example, addition takes a certain number of arguments and relates them to their sum, which is a unique result. In daily life, the expression *father of* can be interpreted as a function. (To refer properly to relations and functions calls for a technical language, such as the lambda calculus, but for simplicity we use informal usage.) If, say, the argument of *father of* is George W. Bush, then the function yields George H. W. Bush as its unique value. The values of functions can be anything, including truth values. For example, given a relation, we can construct a characteristic function that returns the result *true* or *false* depending on whether its arguments belong to the relation. In general, a function yields a unique result for a given argument, though it may not yield any value in certain cases, for example, father of God. In such cases, the function is *partial*. Functions may also be *total*, that is, they may return a result for all possible arguments in the domain. A relation that is not a function does not yield a unique result, for example, many pairs of individuals satisfy the relation *taller than*.

This account of relations may strike psychologists as peculiar. They are likely to think that the key feature of a relation is, not its extension, but its intension, that is, what the relation means. We are sympathetic to this view, and we return to it below, but first we must outline the logical properties of relations.

The Logical Properties of Relations

Logicians have studied relations since classical times, but the logic of relations came to fruition in the 19th century as a result of the work of De Morgan, Schröder, and especially the American logician C. S. Peirce (see, e.g., Kneale & Kneale, 1962). Peirce was concerned with the composition of relations, that is, the way in which they can be combined to form a single new relation, as in *x is the father of the mother of y*, and, conversely, with the reduction of complex relations into simpler components (see, e.g., 1.66,[1] Peirce, 1931–1958). He devised diagrams of relations that anticipated modern semantic networks (see Sowa, 1984), and he argued that an *n*-place relation can always be reduced to a set of relations in which no relation has more than three arguments.

Relations have various logical properties, which give rise to *valid* inferences, that is, inferences for which the conclusion must be true if the premises are true. We outline three sets of properties of binary relations that are pertinent to psychological investigations.

The first set of properties of binary relations concerns transitivity. A relation such as *in the same place as* is *transitive*, because the following sort of inference is valid:

> a is in the same place as b
>
> b is in the same place as c
>
> therefore, a is in the same place as c.

In the above inference, a, b, and c denote entities. A relation such as *next in line to* is *intransitive*, because the following sort of inference with a negative conclusion is valid:

> a is next in line to b
>
> b is next in line to c
>
> therefore, a is *not* next in line to c.

And a relation such as *next to* is neither transitive nor intransitive. It is *nontransitive*, because given premises of the form

> a is next to b
>
> b is next to c

no definite conclusion about the relation between a and c follows validly from the premises. They may or may not be next to one another depending on their spatial arrangement.

The second set of properties concerns symmetry. A relation such as *next to* is *symmetric*, because the following sort of inference is valid:

> a is next to b
>
> therefore, b is next to a.

A relation such as *taller than* is *asymmetric*, because the following sort of inference with a negative conclusion is valid:

> a is taller than b
>
> therefore, b is *not* taller than a.

And a relation such as *nearest to* is neither symmetric nor asymmetric. It is *nonsymmetric*.

[1] This is the standard notation for citing articles in Peirce's (1931–1958) *Collected Works*. For example, 1.66 refers to Vol. 1, paragraph 66, which is the first paragraph of an article, "The Logic of Relatives," originally published in 1883.

The third set of properties of binary relations concerns reflexivity. A relation such as *in the same place as* is *reflexive*, because it follows validly for any entity, a, that

a is in the same place as a.

A relation such as *next to* is *irreflexive*, because it follows validly that

a is not next to a.

And a relation such as *loves* is *nonreflexive*. There are other logical properties of relations, but the preceding ones are the most important. Transitivity and symmetry are independent properties of relations, and Table 1 presents a set of spatial relations to illustrate each of their combinations.

If you sample a dictionary at random for words that can express binary relations, then the most frequent instances are transitive verbs, where "transitive" refers, not to the logical property, but to verbs that take a subject and a direct object. Some of these verbs can take two arguments from the same set, for example, the subject and object of *consults* can both be human beings. Other verbs, however, normally take their arguments from distinct sets, for example, the relation *poured*. In this latter case, linguists say that there are different "selection restrictions" on the subject and object of the verb. From *The Longman Dictionary of Contemporary English* (Proctor, 1978), we took a random sample of 25 verbs that in certain meanings can select subject and object from the same set. The set included, for example: *consult*, *tell*, *synthesize*, *honor*, *ingest*, *popularize*, *delouse*, *hose*, and *confuse*. Most of the verbs have relational meanings that are nontransitive, nonsymmetric, and nonreflexive. The one exception in our sample is *ingest*, which appears to be transitive, asymmetric, and irreflexive. For example, if the fish ingests the minnow, and the shark ingests the fish, then the shark ingests the minnow; if the fish ingests the minnow, then the minnow does not ingest the fish; and nothing—we hope—ingests itself.

A lack of these general logical properties does not imply that a relation yields no inferences. Consider the following examples of valid inferences:

Abe forced Beth to stop smoking.

∴ Beth stopped smoking.

∴ Beth smoked at one time and then no longer smoked.

Table 1
Spatial Relations Showing That the Transitivity and Symmetry of Relations Are Independent Logical Properties

Relation	Transitivity	Symmetry
In the same place as	Transitive	Symmetric
Beyond	Transitive	Asymmetric
Not beyond	Transitive	Nonsymmetric
Next in line to	Intransitive	Symmetric
Directly on top of	Intransitive	Asymmetric
Nearest to	Intransitive	Nonsymmetric
Next to	Nontransitive	Symmetric
On the right of	Nontransitive	Asymmetric
At	Nontransitive	Nonsymmetric

Cath managed to prevent Dean from pretending to be a priest.

∴ Cath prevented Dean from pretending to be a priest.

∴ Dean did not pretend to be a priest.

The verbs *force* and *prevent* are each nontransitive, nonsymmetric, and nonreflexive, yet the inferences are valid. Their validity plainly depends on the meaning of the particular verbs in the context of the sentences (Miller & Johnson-Laird, 1976). We argue in due course that the same analysis applies to general logical properties. This latter claim, however, is controversial, and before we defend it, we need to consider how general logical properties can be used to make inferences.

In logic, the standard treatment of the logical properties of relations is to capture them in axioms, that is, propositions that are assumed to be universally true. Because these axioms concern the meanings of terms, they are often referred to as *meaning postulates* (Bar-Hillel, 1967). As an example, consider the treatment of the inference

a is in the same place as b

b is in the same place as c

therefore, a is in the same place as c.

In logic, this inference is an enthymeme because it lacks a premise, and the missing premise is a meaning postulate capturing the transitivity of the relation:

For any x, y, and z, 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.

In the statement above, x, y, and z are variables ranging over the entities in the domain of discourse. The proof of the inference now proceeds as follows. The first step is to instantiate each of the quantified variables in the axiom with the relevant name from the premises. This process calls for three such instantiations, which eliminate the quantifiers and yield the following sentence:

If a is in the same place as b, and b is in the same place as c, then a is in the same place as c.

The conjunction of the two premises can be made using a formal rule of inference (X, Y; therefore, X and Y):

a is in the same place as b and b is in the same place as c.

This proposition corresponds to the antecedent of the conditional above, and so the formal rule of modus ponens (If X then Y; X; therefore, Y) yields the required conclusion:

a is in the same place as c.

Analogous axioms can capture intransitivity, symmetry and asymmetry, and various other properties, including the implication from a relation to its converse, for example,

For any x and y, if x is above y then y is below x.

Tense is important in all of these examples; for example, no transitive conclusion follows from the premises

a was in the same place as b.

b is in the same place as c.

Theories of Relational Reasoning

Psychological studies of relations have mainly concerned transitive inferences. These inferences are based on pairs of premises, such as

Ann is taller than Beth.

Cath is shorter than Beth.

The participants' task is to draw a conclusion of their own, to evaluate a given conclusion, or to answer a question such as, Who is the shortest? The problems are sometimes known as *linear syllogisms*, but we refer to them as *three-term series problems*. Störring seems to have been the first to study such inferences in the laboratory (see Woodworth, 1938), and Piaget (1928) also carried out some early studies of them. Accounts of transitive reasoning are accordingly of long standing, and we outline the six main theories.

The Theory of Mental Operations

The first brief conjecture about transitive inferences is due to William James, and he referred to it as the fundamental principle of inference. Given a linear series of the form $a > b > c > \ldots >$ z, as James (1890) wrote: "any number of intermediaries may be expunged without obliging us to alter anything in what remains written" (p. 646). This idea lies at the heart of Hunter's (1957) theory. He proposed that transitive reasoning depends on two mental operations designed to transform the premises so that they describe a linear order. One operation *converts* a premise; for example, *a is shorter than b* becomes *b is taller than a*. The other operation switches the order of the two premises, for example,

b is better than a

c is better than b

becomes

c is better than b

b is better than a.

Once the premises are in a linear order, then the term common to both of them can be expunged to leave the appropriate relation between the two remaining terms.

Some inferences require neither operation, some require one operation, and some require both operations. Hunter (1957) reported that the response times to answer questions such as, "Who is shortest?" provided some corroboration for his theory. The theory deals solely with transitive inferences, but it provides no account of how individuals know which relations the operations can be applied to. Their application to intransitive relations would yield invalid conclusions.

Imagery Theories

De Soto, London, and Handel (1965) proposed that individuals carry out transitive inferences by constructing a unitary mental representation of the situation described in the premises. This representation takes the form of a visual image of the three terms on a horizontal or vertical axis. The nature of the relation matters. For example, the relation *a is better than b* refers to items toward the "good" end of the scale, whereas *b is worse than a* refers to entities toward the "bad" end of the scale. These evaluations are represented on a vertical scale with the "good" items at the top. The construction of the unitary representation is based on two principles. First, individuals prefer to construct vertical arrays working from the top downward, and to construct horizontal arrays working from left to right. Second, it is easier to represent a premise if its first term is an *end-anchor*, that is, a term that is at an end of the final array rather than in the middle. De Soto and his colleagues obtained some evidence that corroborated their account rather than the operational theory. For example, premises of the form

a is better than b

b is better than c

yielded easier inferences than premises of the form

a is worse than b

b is worse than c.

Huttenlocher (1968) proposed an important variant of the imagery theory. She argued that there was no obvious reason why a premise should be easier to deal with because its subject term was an end-anchor. What mattered was instead that individuals find it easier to move an item that is referred to in the subject of a sentence rather than in its object. This claim held for children who had to move real blocks and, she argued, for adults who had to move an object into an imaginary array.

The Linguistic Theory

Clark (1969) argued that deductive reasoning is almost identical to the process of comprehension, and that the difficulties in making transitive inferences can be explained by three psycholinguistic principles. First, certain relational terms are *lexically marked* and harder to understand and to remember. According to linguistic theory, *unmarked* comparatives, such as *taller than*, can be used in a neutral way to convey the relative degrees of two items on a scale. In contrast, *marked* comparatives, such as *shorter than*, can be used to refer only to items toward the shorter end of the scale. Unmarked terms likewise give their names to the scale, for example, the dimension is called *length* rather than *shortness*. Some dimensions yield marked terms at both ends, for example, *fatter* and *thinner*, but many have an unmarked relational term and a converse marked relational term. Clark proposed that transitive inferences should be easier with unmarked relational terms than with marked relational terms. This principle provides an alternative explanation for De Soto's notion of a preferred direction for constructing an array. But, their respective predictions diverge for premises such as *a is not as good as b*, which uses an unmarked term but calls for the construction of an array working upward. The evidence supported Clark's principle.

The second principle in the theory is the primacy of *functional relations*. It postulates that given a premise, such as

a is worse than b

individuals understand that both a and b are bad faster than they understand their relative degrees of badness. Clark proposed that the sentence has an underlying representation of the form

(a is bad) more than (b is bad)

and that individuals grasp the two parenthesized clauses faster than the relation between them.

The third principle is that individuals search for information congruent with the question posed after the premises. Hence, if the premises both concern the relation *better than*, but the question asks, "Who is worst?", there is an incongruity and individuals should take longer to respond than to the question, "Who is best?".

The image and linguistic theories often run in parallel, but where they diverge, the evidence tended to support the linguistic theory. For example, the image theory predicts that *shallower than* should be easier to work with than *deeper than*, because *shallower than* calls for individuals to construct the array by working downward, whereas *deeper than* calls for them to construct the array by working upward. In contrast, the linguistic theory predicts the opposite difficulty because *shallower than* is the marked term and *deeper than* is unmarked. The experimental results supported the linguistic theory (Clark, 1969). Yet, it is not easy to refute the image theory. On the one hand, the linguistic theory concerns the causes of difficulty rather than the mental processes necessary to make transitive inferences. On the other hand, evidence shows that individuals do construct arrays (e.g., Barclay, 1973; Breslow, 1981; Newstead, Pollard, & Griggs, 1986; Potts, 1978; Potts & Scholtz, 1975; Riley & Trabasso, 1974). In fact, it is feasible to reconcile the two theories. Some authors have argued that individuals use both linguistic principles and arrays at different times during the process (Johnson-Laird, 1972; Sternberg & Weil, 1980), as result of different experimental procedures (Ormrod, 1979; Verweij, Sijtsma, & Koops, 1999), or as alternative strategies (Egan & Grimes-Farrow, 1982; Ohlsson, 1984; Roberts, 2000).

A Connectionist Implementation

Hummel and Holyoak (2001, 2003) have implemented a connectionist system that makes three-term series inferences and that seeks to combine the linguistic theory with the spatial array theory (see above). The program carries out a mapping of entities to locations in an array, relying on principles from an earlier model of analogical reasoning (Hummel & Holyoak, 1997), and then interprets the array in order to draw a conclusion. Entities that occur in a lexically marked relation are mapped to one end of the array and are treated as more similar than those in a lexically unmarked relation. Throughout an inference, the array accumulates information across multiple iterations concerning the greatest and least positions until these positions stabilize. The number of such iterations correlated with human reaction times (from Sternberg, 1980), and the model made other correct predictions. Its principal importance is as a working demonstration of how to implement a high-level symbolic account in a lower level connectionist system. The system does not handle negative relations, and it remains an open question whether any connectionist system can cope with the composition of noncommutative relations, or with relations between relations, which we discuss below.

Formal Rule Theories

One of the drawbacks of the theories that we have discussed so far is that they concern solely transitive inferences. But, psychologists have proposed general theories of reasoning, and a major class of such theories is based on formal rules of inference (e.g., Braine & O'Brien, 1998; Rips, 1994). These theories can incorporate axioms for transitivity, that is, meaning postulates, and account for transitive inferences in a similar way to formal proofs for these inferences—a procedure that we illustrated in the What Are Relations? section above. The great advantage of these theories is that they can deal with any aspect of relational reasoning that can be captured in meaning postulates. And, at first sight, it is not obvious whether any aspects of relational reasoning lie outside the purview of meaning postulates. But the disadvantage of these theories is that they need to posit an implausibly large number of meaning postulates to handle quite straightforward inferences. We return to this point later.

The Theory of Relational Complexity

An inference that depends on a relation with two arguments seems likely to be easier than one that depends on a relation with three arguments. This intuition underlies Halford's (1993) seminal account of relational complexity. He adopted a metric, which was devised by Leeuwenberg (1969) for the complexity of patterns, in order to analyze the complexity of concepts and relations. The metric is based on the number of arguments that a relation takes. Each argument can have many possible values, and so it is a *dimension* of the relation. Halford, Wilson, and Phillips (1998b) proposed to represent binary relations as a tensor product of three vectors within a system of parallel distributed processing, in which one vector represents the relation and the other two represent its arguments. Transitivity, they argued, is a ternary relation: R(a, b, c), because the smallest structure that can instantiate transitivity is an ordered set of three elements.

The difficulty of transitive inferences according to complexity theory depends on the difficulty of forming an integrated representation of the premises. For example, if a person is given a series of premises of the following sort

$a > b$

$c > d$

$d > a$

it takes time and effort to use the third premise to integrate the previous premises (see, e.g., Ehrlich & Johnson-Laird, 1982; Foos, Smith, Sabol, & Mynatt, 1976; Halford, 1984). It is necessary to hold in mind representations of the first two premises in order to integrate them according to the third premise. Hence, a major determinant of the processing difficulty of any task is its *relational complexity*, which is "the number of interacting variables (i.e., dimensions or arguments) that must be represented in parallel to perform the most complex process in the task" (Halford et al., 1998b, p. 805).

Like a chunk (Miller, 1956), such as a number, letter, or word, dimensions can represent different amounts of information. Humans are limited in the number of dimensions that they can process in parallel, and Halford estimates the boundary of competence to be four dimensions (see Halford et al., 1998b; Halford, Baker, McCredden, & Bain, 2005). However, individuals can recode representations into fewer dimensions—in a process of *conceptual chunking*—with concomitant gains in efficiency. An example is the concept of velocity. Its full representation is three-dimensional (the function itself, and its arguments: distance and time), but when we assess it in terms of a needle's position on a speedometer, we reduce it to a one-dimensional concept. The cost of such chunking is that the components of a relation are no longer accessible for independent processing. An alternative strategy to reduce the peak demands of a task is to segment the task into a different sequence of steps that reduce the maximum number of dimensions that have to be processed simultaneously.

Chunking and segmentation can reduce the number of arguments that have to be processed, and so effective complexity depends on the minimal number of dimensions, or arguments, to which a relation can be reduced without the loss of any information necessary to solve the current problem (Birney & Halford, 2002; Halford et al., 1998b). Hence, determining relational complexity is not simply a matter of the number of arguments in a relation, but also depends on whether the relation can be decomposed without loss of critical information. Transitive inferences for ordering a series of *n* entities can be carried out in a series of steps, the most complex of which call for integrating three entities into an ordered triple (Halford, Wilson, & Phillips, 1998a). On Halford's account, the peak difficulty should be the integration of two premises into a unitary representation. The latency of response to a probe tone in a secondary task corroborated this prediction (Maybery, Bain, & Halford, 1986). Likewise, the retention of the premises in memory is independent from the process of integrating them (see Brainerd & Kingma, 1984).

Halford and his colleagues have devised a method for representing relational structure within a connectionist framework (see, e.g., Halford et al., 1994, 1998b). They represent relations with an explicit symbol for the relation. The representation of a relation is composed of the outer product of vectors representing relation symbols and each of the arguments. Any relational instance is represented by a binding of the relation symbol to the fillers for each argument role.

We have described the main theories of relational reasoning. Most of these theories offer accounts of what the mind computes rather than of how it carries out the computations. For example, many of the theories do not explain how individuals represent the transitivity of a relation (but cf. Halford et al., 1995), or how they use it to construct a representation of the premises. Despite intense debate, no one knows which of the competing theories is best, though only the formal rule theory and relational complexity theory go beyond three-term series problems to offer a general account of reasoning with relations. The fundamental difference between the theories is whether conclusions derive from an integrated representation of the premises (the mental operation, imagery, connectionist, and complexity theories) or from representations of separate premises in some linguistic format (the linguistic theory and the formal rule theory). Our goal in what follows is to formulate a theory that resolves these issues.

The Model Theory of Relational Reasoning

The precursor of the present account is the general theory of mental models (see, e.g., Johnson-Laird & Byrne, 1991) and a pioneering treatment of this theory combined with relational complexity (English, 1998). The general theory postulates that reasoners use the meaning of assertions (their intensions) and general knowledge to construct models of the possibilities compatible with the assertions (their extensions). The model or models representing the extension of each premise are integrated into a single set of models. This set is used either to formulate a conclusion or to evaluate a given conclusion. Reasoners infer that a conclusion is necessary if it holds in all the models of the premises, that it is probable if it holds in most of the models granted that they are equipossible, and that it is possible if it holds in at least one of the models. Individuals can likewise refute a conclusion as invalid if they can construct a counterexample to it, that is, a model that satisfies the premises but that is not consistent with the conclusion.

This theory applies to reasoning with sentential connectives, such as *if*, *or*, and *and*, and reasoning with quantifiers, such as *any* and *some* (Johnson-Laird & Byrne, 1991, 2002). It also applies to probabilistic reasoning (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999), to modal reasoning (Bell & Johnson-Laird, 1998), to causal reasoning (Goldvarg & Johnson-Laird, 2001), to deontic reasoning (Bucciarelli & Johnson-Laird, in press), and to the detection and resolution of inconsistencies (Johnson-Laird, Girotto, & Legrenzi, 2004; Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 2000). Experiments have revealed a number of tell-tale phenomena that reflect the use of mental models. First, reasoners cope much better with inferences that depend on only a single mental model than those that depend on multiple models: they appear to focus on one model at a time (e.g., Bauer & Johnson-Laird, 1993; Evans, Handley, Harper, & Johnson-Laird, 1999). Second, individuals can use counterexamples to refute invalid inferences, both those that depend on sentential connectives (Johnson-Laird & Hasson, 2003) and those that depend on quantifiers (Bucciarelli & Johnson-Laird, 1999). Third, erroneous conclusions correspond to some of the models of the premises, again typically just a single model (e.g., Bara, Bucciarelli, & Johnson-Laird, 1995).

In the past, the expansion of the model theory to new domains has called for additional assumptions, especially about the meanings of expressions in the domain. In order to explain reasoning with relations, we also need to make some additional assumptions. The remainder of this part of the article describes the five principal assumptions of the new theory.

The Iconic Structure of Models

Peirce (1931–1958) analyzed the properties of diagrams, representations, and thoughts, which he referred to collectively as *signs*. He distinguished three properties of signs (e.g., 4.447). First, they could be *iconic*, such as visual images, representing entities in virtue of structural similarity. Second, they could be *indexical*, such as an act of pointing, representing entities by a direct physical connection. Third, they could be *symbolic*, such as a verbal description, representing entities by way of a conventional rule or habit. These properties can coexist, and so a photograph with verbal labels has all three properties. Peirce argued that diagrams should be as iconic as possible (4.433), that is, there should be a structural analogy between a diagram and what it represents: The parts of the diagram should be interrelated in the same way that the entities that it represents are interrelated (3.362, 4.418, 5.73). He developed two diagrammatic systems for logic (see, e.g., Shin, 1994; Johnson-Laird, 2002), which exploited their iconic structure. Iconicity is also central to the theory of mental models: "A *natural* model of discourse has a structure that corresponds directly to the structure of the state of affairs that the discourse describes" (Johnson-Laird, 1983, p. 125). A model is essentially an analog of the situation it represents. Other formulations of iconicity include Maxwell's (1910) analysis of diagrams and Wittgenstein's (1922) picture theory of meaning, with its key proposition 2.15: "That the elements of the picture are combined with one another in a definite way, represents that the things [in the world] are so combined with one another." Peirce, of course, anticipated all these accounts, and his concept of an iconic representation contrasts, as he recognized, with the syntactical symbols of language.

A major advantage of iconic representations is that they can yield novel conclusions that do not correspond to any of the propositions used in their construction. When individuals make relational inferences, they seek such a novel relation in a model and formulate a conclusion about it. This advantage of iconicity may explain why the human inferential system relies on models rather than the syntax of sentences or their intensional representations. Consider, for example, the following assertion about a temporal relation:

> The last bomb struck the ground before the siren had begun sounding.

One sort of model of this sequence could itself unfold in time kinematically, though not necessarily at the same speed as the original events themselves. This sort of representation uses time itself to represent time (Johnson-Laird, 1983, p. 10). Another sort of model represents temporal relations statically as a sequence of events akin to a spatial model, except that the main axis represents time. The system using such a model needs to keep track of how to interpret the axes; for example, a program can use one array to represent time, and cells in this array can contain pointers to three-dimensional arrays representing spatial relations among entities (see below). The various sorts of temporal relation—at least as expressed in English (see, e.g., Allen, 1983)—can all be represented spatially. Thus, according to this account, temporal reasoning depends on mapping expressions into static models such as

```
b      s -------
```

in which the time axis runs from left to right, b denotes a model of the bomb striking the ground, and s ------- denotes a model of the interval for which the siren sounded. Events can be described as momentary or as having durations, definite or indefinite (see, e.g., Miller & Johnson-Laird, 1976; Steedman, 1982). Hence, the further assertion

> Viv realized that there was an air-raid while the siren was sounding

means that Viv's realization, r, occurred at some time between the start and end of the siren sounding:

```
b      s -------
            r
```

This model represents indefinitely many different situations that have in common only the truth of the two premises. For example, the model contains no explicit representation of the duration for which the siren sounded, or of the precise point at which Viv's realization occurred. Yet, the conclusion

> Viv realized that there was an air-raid after the last bomb struck the ground

is true in this model, and no model of the premises falsifies this conclusion. Hence, the iconicity of the model with respect to temporal relations yields a novel but valid conclusion.

There is an important caveat: Diagrams or mental representations in themselves do not have a specific meaning. You cannot answer the question of what they mean unless you know how the system interprets them. We can illustrate this point with Peirce's (1931–1958) own diagrammatic systems for reasoning (see, e.g., Johnson-Laird, 2002). In his first system, the so-called *entitative graphs*, separate assertions written in the same diagram are interpreted as disjunctive alternatives. For example, the following diagram

> Ann is here   Beth is here

represents the inclusive disjunction: Ann is here or Beth is here, or both. To represent a conjunction in this system it is necessary to use circles that function as negations. We use parentheses instead as a typographical convenience:

> ((Ann is here)   (Beth is here))

This diagram represents *It is not the case that Ann is not here or that Beth is not here*, that is, *Ann is here and Beth is here*. In Peirce's (1931–1958) second system, the so-called *existential graphs*, separate assertions written in the same diagram are treated as a conjunction of the two. And to represent a disjunction, it is necessary to use negating circles. Hence, the preceding diagram represents: *It is not the case both that Ann is not here and that Beth is not here*, that is, *Ann is here or Beth is here, or both*. More recent attempts at formalizing graphical reasoning illustrate the same point. Stenning and Oberlander's (1995; Stenning, 1996) system for syllogistic reasoning with Euler circles does not invest all the inferential power in the graphical displays—there are quite specific procedures of interpretation that need to be followed in order for a conclusion to be reached. Mere inspection of a diagram does not tell you how to interpret it. You cannot *perceive* which system of graphs you are looking at. Peirce was aware of this fact: When he addressed the question of how diagrams are capable of any form of representation, he remarked: "Such a figure [a diagram or diagrammatoidal figure] cannot, however, show what it is to which it is intended to be applied" (Peirce, 3.419, cited in Shin, 2002, p. 180).

We draw the same moral for mental models. What a mental model represents depends on the model and on the system for interpreting it. The same model can have different interpretations depending on the system of interpretation. This principle applies to data structures in computer programs too. Hence, models or any other sort of representation cannot themselves be charged with the entire burden of inference. This idea was similarly (and more generally) expressed by Anderson (1978), who argued that to understand any system of representation, one needs to know not only the format of the representations, but also the processes of operation on such representations. Consider a mental model that we represent in the following diagram:

○ Δ

What does it mean? The answer, of course, depends on the system of interpretation. It could mean, for example, that the circle is to the left of the triangle, or that the circle and triangle are both present. The model theory postulates that a key component in the interpretation of models that derive from linguistic descriptions is a separate representation of the intension, or meaning, of the assertion (see Johnson-Laird & Byrne, 1991, chap. 9). The intensional representation is linguistic in form, and it is used together with general knowledge to construct the model or models, which represent the extension of the assertion. Each model represents a distinct set of possibilities compatible with the assertion. Hence, the model above can represent the extension of the following assertion about a spatial relation, such as

> The triangle is to the right of the circle

where the left-to-right axis of the model represents the left-to-right axis in the situation.

Models, which represent extensions, do not normally allow individuals to recover the meaning, let alone the precise wording, of the assertions from which they derive. They capture at best the gist of assertions (for evidence, see, e.g., Barclay, 1973; Garnham & Oakhill, 1996). Hence, when individuals modify a model, say, in a search for a counterexample, they need access to an independent representation of the intension of the assertion (Johnson-Laird & Byrne, 1991, chap. 9). Without such a representation, the system would have no way of determining whether a modification to the model was consistent with the meaning of the assertion, nor of determining what the model is supposed to mean in the first place.

As Peirce (1931–1958) argued, a diagram differs from the objects that it represents in being stripped of *accidents*. Likewise, as the logician Jon Barwise (1993) pointed out, a mental model captures what is common to *all* the different ways in which a possibility might occur. Hence, the model in the diagram above, together with the intension of the sentence, represents what is common to any situation in which a triangle is on the right of a circle, or any other equivalent proposition, such as the circle is on the left of the triangle. However, nothing is represented about the size, color, and distance apart of the shapes, or any other such accidents. Indeed, the model can be updated to take into account any further information about properties and relations. But, no matter how many details are added to a description, it is always consistent with indefinitely many different situations.

In sum, a model of a relation contains a corresponding relation. In this sense, the model is *iconic*: Its parts and the relations among them are interpreted *in the system* to correspond to the parts and the relations in the situation that it represents. To ensure that this account is viable, and that models do not merely mean what we, the authors, stipulate that they mean, we have developed computer programs that implement the theory. Of course, the programs do not use line drawings as models, but instead use arrays, which we describe in detail later along with the main principles of the program.

Iconicity is based on the concept of resemblance: "A sign that is an icon represents a certain object or a certain state of affairs by its likeness to its object or state of affairs" (Shin, 2002, p. 24). But, resemblance is not the whole story of iconicity. It is a symmetrical relation, whereas representation is not. A portrait of Einstein, say, resembles him; it represents him, but he does not represent his portrait (Dipert, 1996; Goodman, 1976; Shin, 2002). According to Peirce (1931–1958), resemblance can occur in at least two distinct ways: resemblance in terms of appearance, and resemblance in terms of structure (Shin, 2002). And it is this second sort of resemblance that underlies the iconicity of mental models.

The two sorts of resemblance may explain the common misconception that mental models are images. In fact, the two sorts of representation should not be confused with one another. Visual images represent how something looks from a particular point of view. They are akin to Marr's (1982) two-and-a-half dimensional sketches, and operations on images are visual. In contrast, mental models are abstract structures akin, for instance, to spatial arrays in a programming language. Just as such an array can be used to construct a two-dimensional image in, say, a graphics program, so too can the brain use an underlying three-dimensional mental model of an object to construct a two-dimensional image of the appearance of the object from a particular point of view. Likewise, when individuals carry out a mental rotation of a depicted object (e.g., Shepard & Metzler, 1971), it is not an image that they rotate but an underlying model. The evidence for this claim is that rotations in depth produce the same pattern of results as rotations in the picture plane. As Metzler and Shepard (1982) remarked,

> These results seem to be consistent with the notion that . . . subjects were performing their mental operations upon internal representations that were more analogous to three-dimensional objects portrayed in the two-dimensional pictures than to the two-dimensional pictures actually presented. (p. 45)

A matter of current controversy is whether there are any *amodal* mental representations, that is, representations that are not in any sensory modality. Markman and Dietrich (2000) wrote, "Theoretical arguments and experimental evidence suggest that cognitive science should eschew amodal representations" (p. 472). This hypothesis has much to recommend it. Earlier theorists had underestimated the power of perceptual representations (see Barsalou, 1999, for a similar argument). But, as we will argue, models are spatial rather than visual, and some aspects of models are amodal. Certainly, you can have a visual (and auditory) image of the bomb striking the ground followed by the siren sounding. The visual images, however, are projections from an underlying three-dimensional model. Moreover, models can represent abstract relations of a sort that are not entirely visualizable, such as

> Viv realized that there was an air-raid.

Similarly, no visual image alone can capture the content of a negative assertion, such as

> The bomb did *not* strike the ground while the siren was sounding.

You might form a visual image, say, of a large red cross superimposed on your image of the bomb striking the ground while the siren was sounding. But, in this case, you would need to know that the red cross symbolizes negation (cf. Wittgenstein, 1953). Alternatively, you might form several visual images to represent the *contrast* class of affirmative relations given the truth of the negation, that is, the bomb striking the ground before the siren sounded, and the bomb striking the ground after the siren sounded (Oaksford, 2002; Oaksford & Chater, 1994; Schroyens, Schaeken, & d'Ydewalle, 2001). In this case, you would have to know that the different images represent *alternative* possibilities as opposed to some other relation between them (e.g., that the two images represent distinct events happening in different places). And in cases where negative assertions allow indefinitely many alternatives, you would have to intend that the alternative images exhaust the appropriate set of relations. A set of representations in itself does not convey this information. Hence, negation cannot be captured in an image alone. It calls for what Peirce called a symbolic element that refers in turn to the meaning of negation: If an assertion is true then its negation is false, and if an assertion is false then its negation is true.

These considerations lead to the following assumption about the structure of mental models:

> 1. The principle of *iconicity*: Models are iconic insofar as possible, that is, their parts and relations correspond to those of the situations that they represent. They underlie visual images, but they also represent abstractions, and so they can represent the extensions of all sorts of relations. They can also be supplemented by symbolic elements to represent, for example, negation. Iconicity arises from a representational system with access to the intensional representations of assertions.

A consequence of this assumption is that relations that evoke vivid images should not necessarily enhance reasoning, and may even impede it. The literature on this topic yields seemingly conflicting results, but, as we show later, the principle of iconicity appears to resolve the conflict.

Emergent Logical Consequences

Earlier we raised a question that most previous accounts of transitive reasoning have not answered: how do individuals represent the logical properties of relations? Formal rule theories have axioms (meaning postulates) for this purpose (e.g., Rips, 1994). The model theory, however, does not. Indeed, it postulates that individuals do *not* represent logical properties at all. Instead, it adopts the following assumption:

> 2. The principle of *emergent consequences*: Individuals use the meanings of relational assertions in intensional representations to construct mental models of the extensions of assertions, and the logical consequences of relations emerge from these models.

The previous example of the bomb and siren illustrated this principle—a novel conclusion was drawn from a model of the premises. We include a further illustration to demonstrate that emergent consequences include not only novel conclusions, but also the basic logical consequences of the relations themselves. Suppose that someone asserts the following spatial relation:

Ann is in the same place as Beth.

The comprehension of this assertion yields a representation of its meaning, which is used to construct a model of the situation:

(Ann Beth)

The parentheses represent a place, and the diagram denotes a model of the two individuals in the same place. A computer program for spatial reasoning inserts a list of the two individuals in the same cell of a spatial array (see below). The symmetry of *in the same place as* emerges at once from the model, because it supports the symmetrical conclusion:

Beth is in the same place as Ann.

No model of the original assertion can falsify this conclusion, and so it follows necessarily from the original assertion. The model also yields the reflexivity of the relation, because both of the following assertions are also true and have no counterexamples:

Ann is in the same place as Ann.

Beth is in the same place as Beth.

These conclusions are so obvious that individuals are unlikely to draw them spontaneously. Yet, they are valid inferences.

A more interesting conclusion emerges given a further assertion in the same discourse:

Cath is in the same place as Beth.

The previous model can now be updated with the new information conveyed by this second assertion:

(Ann Beth Cath)

The transitive conclusion follows of necessity:

Ann is in the same place as Cath.

The same general approach yields transitivity, symmetry, and reflexivity as emergent properties from the construction of models based on the meanings of assertions containing certain relational terms.

In general, logical properties depend on the meanings of relations and the constructions of models based on them. The theory accordingly predicts that individuals do not normally represent logical properties, and that inferences emerge from models. The next section, An Assessment of the Theory, presents evidence that corroborates this prediction.

An Algorithm for Relational Reasoning

The predictions of the relational theory derive from its five main principles— of which we have so far described two. But, we have also implemented the theory in programs for spatial and temporal reasoning. The programs are intended to be working models of how, in principle, conclusions can be emergent properties of representations rather than derived from meaning postulates and formal rules of inference. In this section, we illustrate this point in a description of the program for spatial reasoning. The programs have also allowed us to explore different strategies for reasoning (see the next section).

The first stage of the spatial program's interpretative process yields a representation of the meaning (or intension) of sentences. Each word has a lexical entry specifying the word's contribution to the truth conditions of assertions. Each rule in the grammar has a corresponding semantic rule so that the process of parsing implements a *compositional* semantics (Montague, 1974); that is, the program's parser uses the semantic rules to combine the meanings of words and phrases according to the grammatical relations among them. The particular proposition that a sentence expresses also depends on knowledge, both of general affairs and of the specific context. Indeed, knowledge can even prevent the construction of a model of a possibility (Johnson-Laird & Byrne, 2002). To simplify matters, however, the program treats context as the information already represented in the models of the discourse so far.

Given an assertion that describes the spatial relation between two entities, such as

The triangle is on the right of the circle

the first step in the program is to use the parser to construct an intensional representation of the sentence. The program's spatial models are three-dimensional Cartesian arrays, and the meanings of spatial relations are captured in terms of procedures for scanning these arrays. The meaning of *on the right of* is ambiguous between a deictic sense that depends on the speaker's point of view and a sense that may be elicited by the intrinsic right-hand side of an object, such as a chair or automobile. For simplicity, we assume a deictic sense so that, for example, the following arrangement in a spatial array satisfies the relation that the triangle is on the right of the circle:

○ Δ

Figure 1 shows a three-dimensional array of cells in which the origin of the array, cell 0 0 0, is in the top left-hand corner (following the conventions for representing arrays in programming languages such as LISP). The figure shows the location of the cell containing the circle as specified by the following three coordinate values:

value of the cell on the left–right axis: 1

value of the cell on the back–front axis: 0

value of the cell downward on the vertical axis: 2

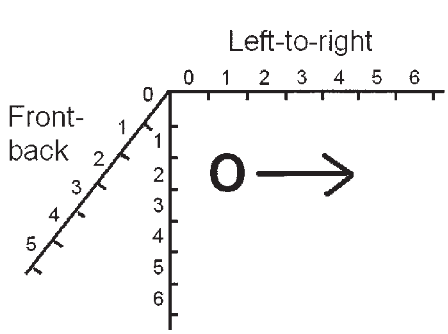

Figure 1. The location of the circle in a three-dimensional array of cells. The arrow shows the direction of the cells to the right of the circle.

Any location in the direction shown by the arrow satisfies the relation *on the right of the circle*. In other words, if you scan the array in that direction, you encounter all the cells in the array that satisfy the relation. To scan in that direction, you need to start at the circle's location and to increment the left-to-right axis while holding constant the values on the other two axes. A general way in which to represent the meaning of *on the right of* is in terms of these changes to the coordinate values of a reference object. Hence, the lexical entry for the deictic meaning of *right* is as follows:

right preposition (1 0 0)

The first item represents the word. The second item represents its syntactic category. And the third item represents its semantics, which specifies that the left–right axis should be incremented (i.e., keep adding 1 to its current value), whereas the other two axes should be kept constant (i.e., keep adding 0 to their current values). The human interpretative system is likely to rely on an analogous system, except that objects themselves, together with the relations among them, are represented within a Cartesian framework. Table 2 summarizes the semantics for a variety of spatial relations relying on the same principles. The result of parsing the sentence, "The triangle is on the right of the circle," is an intensional representation:

((1 0 0) Δ ○)

The second stage of the interpretative process uses the intensional representation of a sentence to build a new model of a discourse, or to update an existing model, or to assess the truth or falsity of the sentence in relation to the current model. There are, in fact, seven principal functions that relate an intensional representation to extensional models. A procedure determines which of them to use. It checks on whether any of the referents in the intensional representation already occur in a model of the discourse.

If none of the referents in the intensional representation are represented in a model of the discourse, then Function 1 is called to start the construction of a new model. Given an intensional representation of the assertion, "The triangle is on the right of the circle," that is, ((1 0 0) Δ ○), it inserts the token representing the circle into an array, and then calls Function 2 below. The program works with the smallest possible three-dimensional array, which it expands whenever necessary. At this point, the array consists of a single cell containing the circle.

If at least one entity in the intensional representation is in a current model, as it will be after the use of the preceding function, then Function 2 is called to add a new token to the current model according to the relation in the intensional representation. In the example, the function adds the token representing the triangle to the right of the location of the circle. The function uses the representation of a relation, (1 0 0), to establish the direction in which to expand the array to form a new cell for the location of the new entity. As in the example, the result of this function may be merely to update an existing model.

Granted the semantics for spatial relations, it should be straightforward to add the first argument of a spatial relation to a model representing its second argument. It should be harder, however, to add the second argument to a model representing the first argu-

Table 2
The Lexical Semantics for a Set of Prepositions Expressing Spatial Relations

Relation	Left–right	Back–front	Top–down
In the same place as	0	0	0
On the right of	1	0	0
On the left of	−1	0	0
In front of	0	1	0
In back of	0	−1	0
Behind	0	−1	0
Above	0	0	1
On top of	0	0	1
Below	0	0	−1

Note. 1 and −1 signify the directions in which to scan to meet the semantics of the relations. 0 means hold a value constant.

ment. There is accordingly an inherent directionality built into the semantics of spatial relations, and in the way in which the program builds up models from them (see Oberauer & Wilhelm, 2000 for corroboratory evidence).

Consider the following description:

The triangle is on the right of the circle.

The cross is on the left of the line.

Functions 1 and 2 construct two separate models, because the two assertions have no referent in common. If the next assertion is

The line is on the left of the circle.

then Function 3 is called to build an integrated model of the two previous models:

✠ | ○   Δ

The initial separate models are therefore like two glimpses of the same situation, but they cannot be integrated until an assertion establishes a relation between referents in them.

Given the further assertion

the cross is on the left of the triangle

all the referents in the intensional representation are represented in a single model of the discourse (as shown above). Hence, Function 4 is called to check whether or not the model satisfies the relation in the intensional representation. It starts at the location of the object of the intensional relation and scans in the direction specified by the relation in order to determine whether or not the subject of the intensional representation is located in one of the cells it scans. In the present case, the function accordingly returns the value *true*. In principle, a model might not represent any information about the relation in the intensional representation. For example, the model might represent the spatial locations of the referents, but not their relative heights asserted in the intensional representation. Function 5 adds the new relation to the current model.

When Function 4 yields the result *true* in the current model, Function 6 tries to find a new model that satisfies the earlier assertions in the discourse but that is a counterexample to the intensional representation of the current assertion. If the function finds such a model, then the current assertion is contingently true; if it fails to find such a model, then the current assertion follows necessarily from the previous discourse. Function 6 can work only if it has access to the intensional representations of the previous assertions in the discourse. As we remarked earlier, individuals cannot modify models in a search for alternatives unless they have access to an independent representation of the meaning of the assertions in the discourse, or at least can reinterpret them. In the present case, no alternative model of the premises is a counterexample to the assertion that the cross is on the left of the triangle. Its validity is an emergent property from the intensions of the propositions and the model-building system. And the system has no need for meaning postulates specifying the logical properties of relations.

If Function 4 yields the result that the intensional representation is false in the current model, then Function 7 tries to find a new model that satisfies the earlier assertions in the discourse and the intensional representation of the current assertion. If the function finds such a model, then the current assertion can be true, and the model is adjusted to satisfy its intensional representation; if the function fails to find such a model, then the assertion is necessarily false given the previous discourse. That is, the assertion is inconsistent with the previous discourse: Its negation follows from the previous discourse.

The interpretative process captured in the previous functions treats each sentence in a discourse as a putative conclusion, but the theory allows that individuals can also formulate conclusions for themselves. The theory postulates that conclusions are drawn by scanning models for a parsimonious and novel relation between entities, that is, a relation that was not asserted explicitly in the original set of sentences. If such a relation is found, then Function 6 is called in order to check whether the relation is necessarily true. Otherwise, no conclusion is forthcoming.

A major source of difficulty in reasoning is the need to represent multiple possibilities. Individuals tend to work with just a single model (see, e.g., Bauer & Johnson-Laird, 1993). Hence, the program is based on the following assumption in the theory of models of relations:

3. The principle of *parsimony*: Individuals tend to construct only a single mental model of a set of relations, to construct the simplest possible model, and to use their knowledge to yield a typical model.

Other theorists have anticipated the same principle (e.g., Evans, Over, & Handley, in press; Ormerod & Richardson, 2003; Sloutsky & Goldvarg, 1999). It reflects the limited processing capacity of working memory and its constraint on inference (e.g., Baddeley, 1986; Klauer, Stegmaier, & Meiser, 1997). Many relational problems, however, are consistent with multiple possibilities. What individuals do in this case is an issue that we now take up.

Strategies of Relational Reasoning

Some theories of reasoning postulate that individuals rely on a single deterministic strategy (e.g., Rips, 1994). In contrast, the model theory, like other theories (e.g., Roberts, 2000), assumes that individuals use a variety of strategies. In particular, the theory is based on the following assumption:

4. The principle of *strategic assembly*: Naive reasoners assemble reasoning strategies from an exploration of different sequences of tactical steps. The exploration is not deterministic, and so individuals are likely to develop different strategies and to switch from one strategy to another. These strategies should be tuned to the exigencies of the problems with which the reasoners are working.

Consider, for instance, these two premises:

The triangle is below the circle.

The cross is above the triangle.

They are spatially indeterminate, because they are consistent with at least three different possibilities, depending on the relation between the circle and the cross:

(1)	(2)	(3)
✠	○	
o	✠	○ ✠
Δ	Δ	Δ

The spatial reasoning program outlined earlier constructs the first of these possibilities, and, if necessary, searches for alternative possibilities (using Functions 6 and 7); a temporal reasoning program that we have also implemented (see below) constructs models of all possibilities. But what do human reasoners do?

Reasoners could try to keep track of all the possibilities in multiple models (Byrne & Johnson-Laird, 1989). They could try to keep track of them in a single model but represent symbolically that the relation between the cross and the circle is indeterminate (Vandierendonck, De Vooght, Desimpelaere, & Dierckx, 1999). Such models do not reduce the number of possibilities that reasoners need to bear in mind, but they do reduce the processing load on working memory, because each entity is represented only once. Another possibility is embodied in the spatial reasoning program: Reasoners construct a single determinate model—a *preferred* model—and, if necessary, they use intensional representations of the premises to search for alternatives. Studies of diagrams based on the 13 possible relations between two time intervals (Allen, 1983) showed that individuals have definite preferences in combining them (Knauff, Rauh, & Schlieder, 1995). They tend to maintain the same rank order of starting and end points or to assume that the intervals are of approximately the same length (Berendt, 1996). Likewise, Jahn, Johnson-Laird, and Knauff (2004) have shown that individuals have preferred models of spatial descriptions based on relations such as, *between*, *next to*, and *on the right of*. Like the spatial program, they tended to use the order in which entities are referred to in indeterminate descriptions to construct their preferred models. They were also likely, where possible, to envisage a layout in which entities referred to in the same assertion were adjacent to one another in the layout. For example, given a description of the following form

b is between a and d

c is on the right of a

the order of terms in the first premise, and the use of adjacency, yield the following preferred model:

a c b d

Strategies are not merely ways to cope with multiple models. They should be sensitive to the goals of reasoning or to what reasoners envisage as relevant to conclude (Van der Henst, Sperber, & Politzer, 2002). But strategies should also be geared to efficient reasoning. For instance, when a problem is posed with a question about a relation, a useful strategy is to ignore any premises that are not relevant to answering the question—a strategy that can yield as a by-product a reduction in the number and complexity of models. Consider, for example, the following problem concerning temporal relations:

a happens before e

b happens before e

c happens before e

d happens before e

What is the relation between a and d?

You readily infer that there is no definite relation between them. There are only two relevant premises, the first and the last, and they yield two possible temporal orders that show that there is no definite relation between a and d:

(1) a d e

(2) d a e

Hence, if individuals have immediate access to all the premises and the question about them, they can construct models from just those premises that are relevant to its answer.

We have implemented this idea in a program that models temporal reasoning. The program constructs multiple models of indeterminacies, but only to a limited number (by analogy with a limited-capacity working memory). When this number is exceeded, the program searches for a coreferential chain of premises interrelating the two events in the question and constructs models only from these premises. Hence, it ignores all premises that are not part of the chain connecting one event in the question to the other. As a corollary, it deals with the premises in a coreferential order, in which each premise after the first refers to an event already represented in the set of models. Experiments have shown that people can also learn to ignore irrelevant premises when the question to be answered is posed before the presentation of the premises (Schaeken & Johnson-Laird, 2000). If there is no question or given conclusion, then neither the program nor human reasoners can use this strategy.

In general, as individuals reason about a series of problems, the natural variation in their tactical steps usually leads to the development of a strategy for coping with problems (Van der Henst, Yang, & Johnson-Laird, 2002). The model theory therefore makes two predictions: First, different individuals are likely to develop different strategies; and, second, experimental manipulations—especially of the sorts of problems that occur in an experiment—should bias participants toward particular strategies.

Higher Order Relations: Relations Between Relations

Relations can occur between relations. The sources of these higher order relations include sentential connectives. The following assertion, for instance, concerns a relation between a pair of relations:

> If Ann is taller than Beth, then Beth is taller than Cath.

There is a vast psychological literature on how individuals reason with connectives such as *if*. The topic is beyond the scope of the present article, but a model theory of sentential reasoning has been presented elsewhere (see, e.g., Johnson-Laird & Byrne, 1991, 2002).

Functions are another source of relations between relations. For example, the relation *taller than* depends on the function, *height of*, and so the assertion

> Abe is taller than Ben

can be paraphrased as

> The height of Abe is greater than the height of Ben.

There can therefore be the following sort of relation between relations:

> Abe is taller than Ben to a greater extent than Cal is taller than Dan.

The individual clauses express relations, but the sentence also asserts a difference between these relations. That is, it asserts that the difference in height between one pair of individuals is greater than the difference in height between another pair of individuals:

> (height of Abe) > (height of Ben) and
>
> (height of Cal) > (height of Dan) and
>
> (height of Abe) – (height of Ben) > (height of Cal) – (height of Dan).

A statistical interaction is another example of a relation between relations (Halford et al., 2005). In theory, there are no bounds on the order of relations between relations. For example, there can be a relation between relations between relations:

> The degree to which Abe is taller than Ben to a greater extent than Cal is taller than Dan is larger than the degree to which Eve is taller than Faith to a greater extent than Gerd is taller than Hope.

This higher order relation is likely to be near to the boundary of everyday competence. It is of a third order in *depth*, that is, it has three levels in depth: At the highest level is one binary relation, at the next level down are two binary relations, and below them are four binary relations depending on the heights of eight individuals. The assertion is difficult to understand, but not merely because it refers to eight individuals. The following assertion also refers to eight individuals, but it is easy to understand because it can be decomposed into a set of binary relations that can be processed separately:

> Abe is taller than Ben, who is taller than Cal, who is taller than Dan, who is taller than Eve, who is taller than Faith, who is taller than Gerd, who is taller than Hope.

What adds to the complexity of understanding the first assertion is indeed its depth—the fact that it concerns a relation between relations between relations. Halford et al. (1998a, p. 855) have argued that relational complexity depends primarily on the number of arguments that have to be processed in parallel. But, depth does appear to affect perceptual judgments (see Kroger, Holyoak, & Hummel, 2004, e.g., correct comparisons between two binary relations were faster than those between relations between two binary relations).

The model theory accordingly postulates that both the number of arguments and their depth should affect reasoning. The theory is based on the following assumption:

> 5. The principle of *integration*: A major component in the difficulty of relational reasoning is the need to integrate information in models. The more complex the integration, the harder the task should be. This complexity depends on the number of entities that need to be integrated (what Halford et al., 1998b, have termed *processing load*), and on the depth of the relation over these entities. It also depends on the particular process of integration as we described in outlining the algorithm for spatial reasoning.

We describe tests of this principle later.

An Assessment of the Theory

The previous part of the article presented a theory of the mental processes and representations underlying reasoning with relations. Earlier studies have provided some indirect evidence that individuals rely on mental models to perform relational inferences. Trabasso, Riley, and Wilson (1975) demonstrated that transitive inferences are performed by assembling elements into an ordered array. Children (aged 6 and 9 years) and college students were presented with a set of adjacent, binary premises concerning the length of a set of items (e.g., A < B, B < C, C < D, D < E, E < F). They were faster to verify the transitive inference that B < E than to verify the inferences that B < D and that C < E. This *symbolic distance effect* suggests that individuals rely on mental models rather than formal rules of inference. The inference that B < E requires more logical steps, but the items are more clearly separable in a mental model. Similarly, participants were generally faster to verify the inferences that B < D and that C < E than to verify the binary pairs they had been trained on (those that did not involve end elements, e.g., B < C, C < D, D < E), the latter being, again, less discriminable in an ordered array or model. Riley (1976) replicated and generalized these findings to transitive relations other than length (e.g., height, weight, happiness, niceness).

Other recent studies have directly corroborated the model theory's predictions (Byrne & Johnson-Laird, 1989; Carreiras & Santamaría, 1997; Schaeken, Johnson-Laird, & d'Ydewalle, 1996; Vandierendonck & De Vooght, 1996). They have established that in general, individuals make more errors with multiple-model problems than with one-model problems. In contrast, the formal derivations of proofs predict the opposite difference in certain cases. The rationale of these studies depends on problems of the following sort:

> 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 b
>
> What is the relation between d and e?

The lower case letters denote entities, such as cups and forks. The preceding premises are consistent with a single spatial layout. In contrast, if instead the first two premises are

> b is on the right of a
>
> c is on the left of b

there are two distinct spatial layouts. Figure 2 presents these two examples (and their layouts) and two other sorts of spatial problem. Theorists had proposed that individuals use formal rules to prove conclusions (see, e.g., Hagert, 1984; Ohlsson, 1984), and these theories postulated two-dimensional meaning postulates. Hagert, for example, proposed meaning postulates of the following sort:

> For any x, y, and z, if x is on the left of y and z is in front of x, then z is in front of x, which is on the left of y.

These rules imply that a one-model problem (such as the first example above: Problem I) calls for the proof of a transitive relation between a and c, which then yields the derived relation between d and e. No such transitive inference has to be proved for Problems I^{i} and II^{i}, because the analogous relation between b and c is asserted in a premise, and the relation between d and e can be inferred from it. Hence, if reasoners use formal rules and meaning postulates, then Problem I should be harder than Problems I^{i} and II^{i}. In contrast, if reasoners rely on mental models, then both sorts of one-model problems, I and I^{i}, should be easier than the multiple-model Problem II^{i}. The two sorts of theory both predict that Problem II, which has no valid conclusion, should be hardest. Overall, the experiments corroborated the model theory for spatial relations (Byrne & Johnson-Laird, 1989). Analogous problems were constructed in the temporal domain using the relations *before*, *after*, and *while*, and events, such as "Mary reads the newspaper." Experiments corroborated the model theory for temporal relations (Schaeken et al., 1996), for studies combining both sorts of relation (Vandierendonck & De Vooght, 1996), and for abstract relations such as *studying more than*, and *copying from* (Carreiras & Santamaría, 1997).

Defenders of formal rules have made three counterarguments. First, spatial relations (and instructions, say, to imagine objects on top of a table) may encourage participants to rely on visual images (cf. Rips, 1994, p. 415). But, as more recent studies have shown (e.g., Carreiras & Santamaria, 1997; Schaeken et al., 1996), the results also apply to contents that are not readily visualized. Second, as Figure 2 shows, Problems I^{i} and II^{i} have irrelevant premises, which could lead participants astray as they search for formal proofs (Rips, 1994, p. 415). Subsequent research, however, has shown that one-model problems remain easier than multiple-model problems even when all the premises are relevant (Schaeken, Girotto, & Johnson-Laird, 1998). Third, Van der Henst (2002) has argued that meaning postulates could be framed for indeterminate relations, such as

> For any x, y, and z, if x is on the left of y and z is on the left of y, then x is on the left of z or z is on the left of x.

If such rules are added to Hagert's (1994) set, then proofs for I^{i}, that is, the one-model problems with irrelevant premises, are of the same length as proofs for II^{i}, that is, the multiple-model problems. Van der Henst (2002) argued that the II^{i} problem imposes a greater memory load or that the accessibility or ease of use of its rules yields its additional difficulty. There are still other possibilities to save the formal rule theory (see Van der Henst, 2002, p. 199). The moral, as philosophers of science have long argued, is that post hoc auxiliary assumptions can save any theory. Whether they are plausible, however, is another matter.

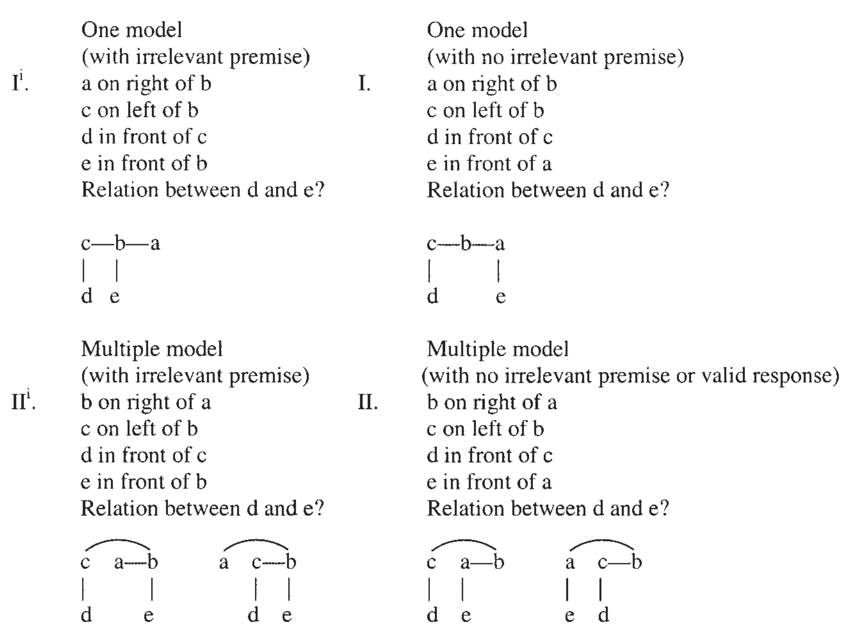

Figure 2. Examples of four sorts of relational problem. Lower case letters denote entities, and the final question is of the following form: What is the relation between d and e?

A major difficulty for Van der Henst's (2002) proposal arises from another way of describing temporal relations (see, e.g., McTaggart, 1927; Miller & Johnson-Laird, 1976, section 6.2). Descriptions can locate events in the past, present, or future, using tense and aspect, rather than relational terms such as *before* and *after*, for example,

> John has cleaned the house.
>
> John is taking a shower.
>
> John is going to read the paper.
>
> Mary always does the dishes when John cleans the house.
>
> Mary always drinks her coffee when John reads the paper.
>
> What is the relation between Mary doing the dishes and drinking coffee?

This way of establishing temporal relations poses a difficulty even for indeterminate meaning postulates. Meaning postulates capture the logical properties of relational terms. It is not easy to see how to frame them for the example above, which relies only on tense and aspect. Yet, even with these materials, one-model problems are easier than multiple-model problems (Schaeken et al., 1996). We now turn to some new predictions of the model theory.

Iconic Models Rather Than Images Underlie Reasoning

Does reasoning depend on visual images or mental models? According to the principle of iconicity, models underlie images, and reasoning depends on models. Hence, problems with contents that elicit imagery should not necessarily enhance reasoning and may even impede it. At first sight, this claim is discrepant with those findings in the literature that appear to show that imagery improves reasoning. In fact, the principle of iconicity resolves an apparent conflict between two lines of evidence in the literature.

One line of evidence suggests that imagery enhances reasoning. However, these studies have tended to use materials that differ in the ease of constructing spatial representations. For example, as we mentioned earlier, De Soto et al. (1965) examined the extent to which different relations evoked a representation with a spatial axis. Spatial relations such as *above* and *below* elicited such representations, as did control relations such as *better* and *worse*, whereas visual relations such as *lighter* and *darker* elicited no directional preference. The spatial and control relations facilitated reasoning in comparison with the visual relations. Shaver, Pierson, and Lang (1976) investigated transitive inferences with the three relations: *above*, *better than,* and *has lighter hair than*. The participants made fewer errors with the spatial relation of *above* than with the control relation *better than*, which in turn yielded fewer errors than the visual relation *lighter hair than*. Hence, rather than a general benefit being conferred by imagery (which was the implication of these studies), it seems instead that visual relations impede reasoning in comparison with other sorts of relation.

The second line of evidence is that imagery has no effect, or indeed has an impeding effect, on reasoning. These studies, however, tended to use materials that evoked visual imagery rather than spatial relations. For example, Sternberg (1980) found no difference between the accuracy of solving problems that were easy or hard to visualize, and he did not find any reliable correlation between scores on the imageability items of IQ tests and reasoning ability. Richardson (1987) found that manipulating the imageability of faces had no effect on reasoning. Clement and Falmagne (1986) studied conditional reasoning and varied the imageability and availability of pertinent knowledge. Statements that were easy to visualize included such conditionals as, "If a man walks his golden retriever, then he gets upset about his insect bite." Statements that were hard to visualize included, "If the man takes an economic perspective, then he uses the new memory technique." These two sorts of contents had no reliable effect on reasoning, perhaps because the materials may have confounded imageability with other factors (Evans, 1980). Egan and Grimes-Farraw (1982) gathered retrospective reports from participants who had drawn three-term series inferences. Those who reported using visual imagery performed worse than those who reported using other, more abstract strategies. Knauff, Jola, and Strube (2001) examined the disruption of three-term series inferences by secondary tasks that were either visual or auditory, and either spatial or nonspatial. Only the spatial tasks disrupted inference, regardless of whether they were auditory or visual.

In collaboration with Knauff, P. N. Johnson-Laird carried out a recent study to test the principle of iconicity (Knauff & Johnson-Laird, 2002). The first step was to find materials that dissociated ease of forming a visual image from ease of forming a spatial representation. Relations that evoke spatial representations, such as *above* and *below*, also tend to be easy to visualize. However, a panel of judges did rate some relations, such as *cleaner* and *dirtier*, as easier to visualize than to envisage spatially. The second step was to determine the effect of these various relations on reasoning. The visual relations slowed down transitive reasoning in comparison with the other sorts of relation.

A subsequent study used functional magnetic resonance imaging in the absence of any correlated visual input (problems were presented acoustically via headphones). Reasoning with all sorts of relations evoked activity in the left middle temporal gyrus, in the right superior parietal cortex, and bilaterally in the precuneus (Knauff, Fangmeier, Ruff, & Johnson-Laird, 2003). In the prefrontal cortex, increased activity was found in the middle and inferior frontal gyri. However, only problems based on visual relations activated area V2 in the visual cortex. Hence, cortical activity during reasoning depends on whether the contents evoke spatial representations or, in addition, visual images. These results support the principle of iconicity: Reasoning depends on spatial models, not visual images, which can be a distraction. This account accordingly differs both from those theories that postulate that imagery is central to higher cognitive processes and from those theories that postulate that it is epiphenomenal (Pylyshyn, 1981, 2002). Imagery is not central, because reasoning does not depend on it; it is not epiphenomenal, because it can impede thinking.

Logical Properties as Emergent From Models

The model theory makes the major claim that individuals do not possess meaning postulates stipulating the logical properties of relations, but that instead these properties emerge from the construction of models. Our computer programs demonstrate how, in principle, such a system could work. Transitive inferences are so simple that it is hard to see how to test empirically the difference between formal rules and mental models. But, the principle of parsimony together with the principle of emergent properties does yield an unexpected and testable prediction.

Some relations have clear-cut logical consequences within the family of transitive, intransitive, and nontransitive properties. For instance, a relation such as *in the same place* is obviously transitive, a relation such as *father of* is obviously intransitive, and a relation such as *loves* is obviously nontransitive. If mental models follow the principle of parsimony, however, then they should be simple and typical of the domain, and individuals should tend to overlook alternatives to them in any sort of reasoning. One consequence is that individuals should infer certain transitive relations when, in fact, the inference is unwarranted. On this account, there should be a class of relations that are not obviously intransitive, that are in fact nontransitive, but that elicit simple models yielding transitive conclusions. We refer to relations of this sort as *pseudotransitive*. Here is an example of an inference based on a pseudotransitive relation:

Ann is a blood relative of Beth.

Beth is a blood relative of Chris.

Therefore, Ann is a blood relative of Chris.

According to the theory, individuals should tend to construct a simple model of collateral relatives such as siblings or linear descendants:

Ann
|
Beth
|
Chris

They should accordingly infer the transitive conclusion:

Ann is a blood relative of Chris.

A counterexample shows that the inference is invalid: Your mother is related to you, and you are related to your father, but your mother and father are not necessarily blood-relatives. Naive individuals abiding by the principle of parsimony should be likely to overlook such counterexamples. Yet, a simple manipulation should enhance the likelihood that they consider them. When individuals are given a context that elicits alternatives to oversimplified models, they should be less likely to draw pseudotransitive conclusions. For example, the previous fallacy should be reduced by an instruction to consider the consequences of marriage on kinship.

We have carried out two experiments on pseudotransitive inferences (Goodwin & Johnson-Laird, 2004a). The first experiment tested whether individuals spontaneously draw their own pseudotransitive conclusions from the premises of three-term series problems. It examined five pseudotransitive relations: *blood relative of*; *in front of* and *behind*, which are nontransitive if they refer to the intrinsic parts of objects; and *went faster than* and *overtook*, which are nontransitive if they refer to events at different times, say, in a race. The experiment also examined five relations that are obviously transitive, for example, *taller than*, and three relations that are obviously not transitive, for example, *loves*. The problems were presented in a different random order to each of 24 Princeton University students, whose task was to state what must follow given that the premises were true. The participants drew transitive conclusions on 98% of the problems with transitive relations, on 72% of the problems with pseudotransitive relations, and on 8% of the problems with relations that were not transitive. This predicted trend was highly reliable.

The second experiment showed that when the context drew attention to less typical models, there was a reliable decrease in the frequency of pseudotransitive inferences. The participants stated what followed from 20 sets of premises: 12 sets were pseudotransitive and included contents akin to those of the first experiment, 4 sets were obviously transitive, and 4 sets were obviously not transitive. Half of the problems were presented with a short linguistic context, which in the case of the pseudotransitives was designed to elicit alternative models, whereas the other half of the problems were presented without a context. Likewise, half of the problems were based on affirmative relations, whereas the other half of the problems were based on negated relations (e.g., *not taller than*). The problems were presented in a different random order to each of 30 Princeton University students, whose task was the same as in the previous experiment. They drew transitive conclusions on 91% of the problems with transitive relations, on 52% of the problems with pseudotransitive relations, and on 3% of the problems with relations that were not transitive. The trend was again highly reliable. But, in addition, the contexts reliably reduced the tendency to draw pseudotransitive conclusions (only 41% of them in comparison with 62% conclusions without the contexts). The contexts had no reliable effects on the inferences from relations that were obviously transitive or obviously not transitive. Negation is known to cause difficulties in reasoning, and negative relations were less likely to yield transitive conclusions.

The results of the study corroborate the model theory, but they are difficult to reconcile with the view that logical properties are represented in meaning postulates. You either have, or do not have, a meaning postulate for the transitivity of a particular relation. If you have the postulate, which is axiomatic, then you should draw the transitive conclusion, barring momentary lapses in performance. If you do not have the meaning postulate, then you should not draw the transitive conclusion. Our results, however, put this account on the horns of a dilemma. On the one hand, if individuals draw a pseudotransitive conclusion, then they presumably have the relevant postulate. But, in this case, why does the context reduce this propensity to draw the transitive conclusion? On the other hand, if they do not draw a pseudotransitive conclusion from premises presented with the context, then they presumably do not have the relevant postulate. But, in this case, why do they draw the transitive conclusion in the absence of the context? Meaning postulates are not entities that switch on and off like a light. They are axioms that have universal application, because they capture the logical properties of words (Bar-Hillel, 1967). And although they may not be always accessible, a linguistic context priming alternative models should not inhibit them.

One conjecture to save meaning postulates is that the pseudotransitive relations are ambiguous and have one sense that is transitive and one sense that is not. This conjecture seems plausible for *in front of*, *behind*, and their cognates. Their *deictic* sense is transitive, but the sense referring to intrinsic parts is not. But, even for these relations, there are difficulties for the conjecture. Our participants still drew pseudotransitive conclusions for relations that refer to intrinsic parts, such as *on X's right*. Likewise, the conjecture cannot explain the pseudotransitive inferences with unambiguous relations, such as *blood relative of*.

A particularly thorny problem for meaning postulates arises from the vagaries in spatial relations based on intrinsic parts. Given a set of individuals seated down one side of a table, a relation such as *on X's right-hand side* is treated as transitive. When the individuals are seated round a small circular table, however, it is intransitive. If the circular table were larger, then transitivity would extend over a small number of individuals, but gradually break down with a larger number as they get further round the table. To capture these phenomena with meaning postulates seems to call for an indefinite number of postulates for the same relation. At one end of the spectrum the relation is intransitive (a small round table), then it is transitive over three individuals but not four (a slightly larger round table), and so on, up to a relation that is unboundedly transitive (a long rectangular table). In fact, what matters in these examples is the actual seating arrangement, and whether or not Y can be truly described as on X's right-hand side. A single meaning of *on X's right-hand side* captures all these vagaries given a mental model of the seating arrangement. Transitivity is an emergent property from models.

Strategies in Relational Reasoning

The pioneering studies of strategies in reasoning were based on series problems. When participants carried out five-term series problems based on the same relation, such as *taller than*, they rapidly developed a short-cut strategy (Wood, 1969; Wood, Shotter & Godden, 1974). They looked for a noun phrase that occurred solely on the left-hand side of a single premise. It denoted the tallest entity. One result of this strategy was that the participants could rapidly answer the question posed in the problem but were unable to answer unexpected questions about other entities in the series. Quinton and Fellows (1975) asked their participants to talk about the strategies that they had developed. After repeated experience with problems sharing the same formal properties, the participants tended to identify invariants (e.g., an extreme term is mentioned only once, and the middle term is never the answer of the question), and to use them to solve the problems with minimal effort. These investigators described five different perceptual strategies, such as one in which the participants try to answer a question about a three-term series problem solely from the information in the first premise, for example, Pat is taller than Viv. If they obtain an answer, for example, Pat is the taller, and this term does not occur in the second premise, then they do not need to consider the second premise: The answer is correct. The strategy works only for problems that yield a determinate order for the three individuals.

According to the principle of strategic assembly, a strategy is assembled from a sequence of tactical steps. Individuals try out various tactics until they succeed in finding a sequence yielding an answer to the problem. The principle predicts that different individuals should develop different strategies for relational reasoning, and may switch from one strategy to another. We have tested these predictions in a recent study of inferences based on relations between relations (Goodwin & Johnson-Laird, 2004b). A typical problem from the study was based on the following premises:

Abe is taller than Ben to a greater extent than Cal is taller than Dan.

Ben is taller than Dan to a greater extent than Cal is taller than Dan.

The task was to put the four individuals into their rank order in height, or to state that the task was impossible because there was no definite order. In the problem, the second premise yields a model of the height of the three individuals to which it refers. Both Ben and Cal are taller than Dan, but Ben is taller than Dan by a greater amount than Cal is. The resulting model is accordingly

Ben
Cal
Dan

where the vertical axis represents relative height. The first binary relation in the first premise is that Abe is taller than Ben, and so the model can be updated to represent the complete order:

Abe
Ben
Cal
Dan

The first premise, of course, contains further information, but it is not needed for this problem.

In contrast, a second sort of higher order problem is based on premises, such as

Abe is taller than Cal to a greater extent than Ben is taller than Dan.

Ben is taller than Dan to a greater extent than Cal is taller than Dan.

Once again, the second premise yields the model

Ben
Cal
Dan

The first premise states that Abe is taller than Cal, but is Abe taller than Ben? To answer this question, it is necessary to use all the information in the first premise. Suppose that the model-building system adds Abe in a preferred way:

Abe
Ben
Cal
Dan

The first premise holds for this model: Abe is taller than Cal, and Ben is taller than Dan, and the model is consistent with the second order relation. In contrast, suppose that the system adds Abe to the model in the following way, that is, it assumes that he and Ben are of the same height:

Ben Abe
Cal
Dan

It is still the case that Abe is taller than Cal and that Ben is taller than Dan. But, it is impossible for the difference in heights between Abe and Cal to be greater than the difference between Ben and Dan. Hence, the first model is the correct one. Its construction, however, should be harder than the construction of the model for the first problem, because its construction for the right reasons depends on all the information in the first premise.

Because a strategy is a sequence of tactical steps, the space of all conceivable strategies depends on the set of possible tactical steps. There are seven main tactical steps for dealing with binary relations: starting a model, updating it, integrating two models, verifying a relation in a model, searching for an alternative model that satisfies a relation, and searching for an alternative model that refutes a relation (see the program described in The Model Theory of Relational Reasoning section above). For the present problems, versions of these tactics also use second-order premises to construct and to verify ternary and quaternary orders. A separate tactical step relevant to some strategies is to search for an end item in the order. Given Problem 1 above, for example, the first premise yields two candidates for the shortest individual: Ben and Dan. The second premise reveals that Dan is unequivocally the shortest.

The participants in the experiment were allowed to use paper and pencil, but to help us to elucidate their strategies, they had to think aloud as they tackled the problems. We video-recorded what they wrote, drew, and said. We were able to make sense of over 90% of the protocols, and all the main tactical steps outlined above occurred in them. At first, the participants tended to flail around, trying out various tactics, but they soon settled down and were able to produce the correct answer on the vast majority of trials. In some cases, they abandoned one strategy and took up another in order to solve a problem. The possible strategies could have been based on formal rules or on mental models, but the protocols show that the participants relied on models. One participant made an initial use of algebraic expressions, but he subsequently abandoned this procedure. There are three main strategies based on models, and we observed each of them:

1. The ternary strategy (49% of observed strategies): Reasoners used the second premise to construct a ternary order, and then used the first premise to add the fourth individual to the order.

2. The transitive strategy (24% of observed strategies): Reasoners mentally conjoined a diagram (or model) of one binary relation with one of another binary relation to yield an integrated transitive order of three individuals. Each binary relation came from a separate premise. They then imported the fourth individual into the order from a premise.

3. The frame strategy (17% of observed strategies): Reasoners constructed an initial frame, typically by inferring, or guessing, the tallest and shortest individuals. On occasion, reasoners would instead use a model of one clause in the premises as the frame. They would then insert the remaining individuals within the frame by using an educated guess and check the results against the premises.

As the principle of strategic assembly predicted, individuals differed in the strategies they developed, and almost all of them tried more than one strategy. They used a mean of 2.27 out of the three main strategies. Moreover, they shifted from one strategy to another reliably more often than necessary. Regardless of strategy, as the theory predicted, the participants solved the problems that depended on a single relation from the first clause (M = 59.2 s) faster than the problems that depended on all the information in the first premise (M = 87.9 s).

The principle of strategic assembly postulates that individuals experiment with various tactical steps and assemble strategies "bottom up" from sequences of these steps. It follows that the characteristics of problems should influence the particular strategies that individuals develop. To test this prediction, we carried out a further study in which we manipulated the order of premises. When the premise yielding the ternary order occurred first, it should have inculcated the ternary strategy. To inculcate the transitive strategy, we made the two relevant binary relations in the separate premises more salient by presenting them in a different color than the rest of the premises.

The results corroborated the predictions. The participants relied on the same three strategies that we observed in the first study. But, their use was reliably biased by the experimental manipulations. The occurrence of the ternary premise first yielded the ternary strategy on 54% of trials, whereas the color coding of the binary relations yielded this strategy on only 26% of trials, and instead the participants used either the transitive or the frame strategy. Evidently, the emphasis on the two binary relations also encouraged the participants to identify the end items in the order. Once the participants had developed a strategy in the first block of problems, they tended to continue to use it in the second block. But, the ternary strategy transferred to a greater degree than the other strategies. Notwithstanding the differences in strategies, the complexity of integrating information predicts that problems calling only for one binary relation from a premise should remain easier than those calling for all the information from the premise. As in the previous study, the results corroborated this prediction.

A special domain of relations concerns kinship. The particular relationships that are lexicalized vary from one culture to another, though they are based on a small number of relations such as *parent of*, and properties such as *female* (Miller & Johnson-Laird, 1976). Studies of kinship inferences have focused on how participants infer that two different descriptions, such as *father's brother* and *uncle*, can refer to the same relative (e.g., Wood & Shotter, 1973). Some equivalences, including the previous example, are likely to be stored in long-term memory, but others can be grasped only by using the intensions of the terms to construct models of lineages. A major result is that individuals do develop strategies to deal with such inferences; for example, they learn short cuts, such as using a mismatch in the sex of the two relatives in a putative equivalence (Cech & Shoben, 1980; Oden & Lopes, 1980). The fundamental concepts are biological, and children appear to grasp them by the age of 4 (see Springer, 1992, 1995), contrary to earlier theories that children have a social theory of the family (Carey, 1985; Haviland & Clark, 1974).

The Integration of Information

According to the principle of *integration*, a major component of the difficulty of relational reasoning is the need to integrate information in models. This difficulty depends on the number of arguments that have to be integrated, as Halford and his colleagues have argued, and on the process of integration (see the description of the algorithm in The Model Theory of Relational Reasoning section above). However, according to the principle of integration, difficulty also depends on the depth of the relation holding over the arguments. No previous studies had demonstrated a clear-cut effect of either number of arguments or depth on reasoning, and so we carried out three experiments to test the predictions of the principle of integration.

Experiment 1: The Number of Unique Arguments

Does the number of unique arguments to be integrated in an inference affect its difficulty? Halford and his colleagues have examined this question indirectly. They analyzed the underlying processes in various inferences and comprehension tasks, and argued that such effects do occur (see, e.g., Andrews & Halford, 1994, 2002; Birney & Halford, 2002; Halford et al., 1998b). What has yet to be shown, however, is that a simple manipulation of the number of arguments affects the difficulty of inferring the *same* conclusion (cf. Maybery et al., 1986, who kept the premises constant, and investigated the difficulty of evaluating conclusions of different complexities).

The first of our studies examined spatial inferences about the relative starting positions of runners in a race. Each problem was based on three premises, and the task was to answer a question about the starting position of two of the four runners who were allocated to five lanes (hence, there was one empty lane). In one sort of problem, the third premise referred to three different runners, for example,

a is left of c and b is left of a

b is left of c and d is left of b

a is further away from c than b is from a

Who is closer to the empty lane, b or a?

The third premise has two noun phrases referring to the same runner. Coreference of this sort should improve reasoning, because it reduces the complexity of the integration process (Walsh & Johnson-Laird, 2004). The answer to the question depends on inferring the allocation of the four runners to the five lanes. The first premise yields the order

b a c

The second relation in the second premise updates the order

d b a c

The third premise yields the allocation to the lanes. Reasoners can use it to represent that the distance from a to c is greater than the distance from b to a:

| d | b | a | | c |

Hence, the answer to the question is that a is closer to the empty lane than b is.

In a second sort of problem, the first two premises were the same, but the third premise referred to four different runners:

b is further away from c than d is from a.

This premise yields the allocation to the lanes. But, in this case, reasoners need to represent that the distance from b to c is greater than the distance from d to a; that is, they need to integrate a relation concerning four distinct individuals:

| d | b | a | | c |

As before, the answer to the question is that a is closer to the empty lane than b is. The principle of integration predicts that the problem with three referents in the crucial premise should be easier than the problem with four referents in the crucial premise.

Method

The participants acted as their own controls and carried out five problems of the two sorts in separate blocks of trials. There were two separate groups of participants to counterbalance the order of the two blocks. Within each block, each participant carried out the problems in a different random order. To prevent the participants from learning the answers, the first two premises were presented in two different linguistic versions, the premises described two different spatial arrangements, there were five distinct questions concerning five pairs of individuals, and the order of each pair in a question was also counterbalanced. As in the preceding examples, the runners were denoted by single letters (in capitals). Ten different sets of four letters were constructed and were allocated at random to the problems.

The 21 paid participants, who were students at Princeton University, were tested individually and carried out the experiment on an IBM computer running the SuperLab Pro (Cedrus Corporation, 1999) program. They were not allowed to write anything down, although a diagram of five empty lanes was presented at the bottom of the computer screen. The participants had to make their responses by typing the letter corresponding to the correct person. They followed a self-paced procedure in which the premises and the question were presented by a separate key press. The first two premises were presented simultaneously: These pairs of premises were identical in both sorts of problem. All the sentences in a problem, premises and question, remained on the screen until the completion of the problem.

Results

The percentages of correct conclusions for the two sorts of problem and the mean reading latencies for the third premise were as follows:

three-argument problems: 90%, 15 s

four-argument problems: 77%, 23 s.

The three-argument problems yielded a greater proportion of correct responses and faster reading times of the third premise than the four-argument problems (Wilcoxon's tests, $z = 2.45$, $p < .007$, one-tailed, and $z = 3.22$, $p < .001$, one-tailed, respectively; the difference in latencies was also reliable in an analysis for the correct responses only). No reliable difference occurred in the times to respond to the question.

The experiment corroborated the principle of complexity and Halford's theory of relational complexity. Inferences based on a relation between two binary relations are easier when a common referent occurs in both binary relations than when there are four distinct referents. They yield a greater percentage of correct responses, and the crucial premise takes less time to read. The number of unique arguments in a relation does affect adult reasoning.

Experiment 2: The Effects of Depth

The principle of integration predicts that the depth of a relation should increase the difficulty of reasoning. In order to test this prediction, it is crucial to hold the number of arguments constant, because otherwise it could explain the phenomenon. Indeed, Halford and his colleagues argued that number of variables is the critical factor. They wrote, ". . . higher-order relations are important, but dimensionality is the more general criterion of complexity, and can be applied to structures of any depth" (Halford et al., 1998a, p. 855). We therefore used two sorts of problem in which the participants had to infer the rank order of four individuals in terms of a property such as height. Both sorts of problem contained the same number of arguments, but one sort was lower in depth than the other.

The first sort of problem was based on premises, such as

Ann is taller than Beth to a greater extent than Beth is taller than Cath.

Dot is taller than Ann to a greater extent than Beth is taller than Cath.

The two clauses in the first premise regardless of the relation between them yield a transitive inference about the ternary order:

Ann
Beth
Cath

All that is needed to complete the order is a single binary relation in the second premise, Dot is taller than Ann:

Dot
Ann
Beth
Cath

We refer to problems of this sort as *shallow*, because they are only of a first-order depth. Although the premises are of a second order, the conclusion depends only on three binary relations.

The second sort of problem was based on premises, such as

Abe is taller than Ben to a greater extent than Cal is taller than Ben.

Dave is taller than Abe to a greater extent than Cal is taller than Ben.

All the information in the first premise, including the higher-order relation between the relations, is needed to infer the ternary order:

Abe
Cal
Ben

But, as in the first problem, only a single binary relation from the second premise is needed to put Dave at the top of the order. We refer to problems of this sort as *deep*, because they depend on all the information in a second-order premise. The principle of complexity predicts that the shallow problems (first order) should be easier than the deep problems (second order).

Method

The participants acted as their own controls and carried out 12 test problems, which included 4 shallow problems, 4 deep problems, and 4 problems that did not yield a consistent order. The problems were presented in a different random order to each participant. They were based on two different transitive relations: *taller than* and *heavier than*, and on six sets of female names and six sets of male names. We made two different assignments of the relations to the problems, and each participant was tested with one of the resulting sets selected at random.

The 27 participants from the same population as before were tested individually and carried out the experiment on an IBM computer running the E-Prime (Psychology Software Tools, 2002) program. They were told to decide on a possible order of the four individuals referred to in each set of premises, and to respond "inconsistent" if they thought no such order was possible. They were not allowed to write anything down, and they pressed the "Enter" key to receive each new premise. The instructions emphasized that they were to read and to understand each premise fully before proceeding to the next one.

Results

The problems were easy to solve: The shallow problems yielded 91% correct orders, the deep problems yielded 93% correct responses, and the difference was not reliable (Wilcoxon's test, $z = 0.88$, $p > .3$, one-tailed). Table 3 presents the mean reading times for the first premise and the mean times both to read the second premise and to complete the problem. As the theory predicted, the participants understood the shallow first premises faster than the deep first premises (Wilcoxon's test, $z = 4.25$, $p < .001$, one-tailed). This result occurred both for the consistent problems (Wilcoxon's test, $z = 2.76$, $p < .01$, one-tailed) and for the inconsistent problems (Wilcoxon's test, $z = 3.29$, $p < .001$, one-tailed). Once they had read these premises, the completion times did not differ reliably (Wilcoxon's test, $z = 1.44$, $p = .15$, two-tailed). The principal difference between the two problems was in the depth of the relation that established a ternary order from the first premise. The results accordingly show that depth, not just number of arguments, has a reliable effect on reasoning.

Experiment 3: Second-Order Versus Third-Order Depth

The previous experiment showed that depth contributed to the difficulty of relational reasoning in a contrast between first-order and second-order problems. The present experiment aimed to extend these results to a comparison between second-order and third-order problems. Both sorts of problem had a third-order premise, that is, referring to a relation between relations between relations. The third-order problems called for all of this information to be taken into account, whereas the second-order problems could be solved by taking into account only the second-order information. Because both problems require the same number of

Table 3
Mean Latencies (in Seconds) and Standard Errors for Times to Read the First Premise and to Complete the Problem in Experiment 2

Type of problem	Reading times: First premise		Completion times	
	M	*SE*	*M*	*SE*
Shallow	7.68	1.01	44.09	3.66
Deep	10.60	1.16	39.97	3.07

arguments to be considered in parallel, a difference in difficulty would show that depth does make reasoning harder.

An example of a second-order problem is

> Abe is taller than Ben to a greater extent than Cal is taller than Dave,
> Cal is taller than Abe to a greater extent than Abe is taller than Dave,
> and the first of these differences is larger than the second.

The second premise yields a ternary order from its two separate clauses alone, that is, without taking into account the relation between them:

Cal

Abe

Dave

The first premise as a whole then yields the complete order:

Cal

Abe

Dave

Ben

where the space in the order satisfies the first premise. Because the problem does not require the relation between the two premises to be taken into account, it is only second-order in depth.

An example of a third-order problem is

> Abe is taller than Ben to a greater extent than Abe is taller than Cal.
>
> Abe is taller than Ben to a greater extent than Dave is taller than Cal,
> and the first of these differences is larger than the second.

The first premise yields a ternary order:

Abe

Cal

Ben

But, the correct order of the four individuals depends on the second premise and the third-order relation (stated in the third line of the problem):

Dave

Abe

Cal

Ben

Method

The participants acted as their own controls and carried out two instances of the two sorts of problem, which were presented in a different random order to each of them. The problems were based on three different transitive relations: *taller than*, *heavier than*, and *bigger than*. We made three different assignments of these relations to the four forms of problem. The 16 Princeton University participants were tested with one of the resulting sets selected at random. They were told that their task was to write down an order of the four individuals that was consistent with the premises in a box at the bottom of the page, and to write "none" if they thought no such order existed.

Results

The second-order problems yielded 84% correct orders, whereas the third-order problems yielded only 56% correct orders (Wilcoxon's test, $z = 3.00$, $p < .005$, one-tailed). As the model theory predicted, an increase in depth makes a reasoning problem more difficult. The phenomenon cannot be explained solely in terms of the number of arguments that need to be taken into account. Both sorts of problem call for the processing of quaternary relations.

General Discussion

Our goal has been to advance a comprehensive theory of reasoning about relations. The theory is based on mental models, and on five principles:

1. Iconicity: The structure of models is iconic as far as possible; that is, their parts and relations correspond to those of the situations that they represent, but they should be distinguished from images because they can represent any sort of content. Iconicity results from models and a system that accesses an independent representation of the meanings of propositions.

2. Emergent consequences: The logical consequences of relations emerge from models satisfying their premises.

3. Parsimony: Individuals tend to construct only a single, simple, and typical model of a situation satisfying the premises.

4. Strategic assembly: Individuals develop different strategies for reasoning with relations, but these strategies reflect the problems on which they are working.

5. Complexity of integration: The difficulty of relational reasoning depends on the number of entities that have to be integrated to form a model, on the process of integration, and on the depth of the relation holding over the entities.

We have devised programs for spatial and temporal reasoning that implement the main assumptions of the theory. The programs demonstrate how, in principle, conclusions can be emergent properties of models. They also illustrate different strategies for reasoning.

Experimental studies have corroborated each of the principles of the theory. In particular, behavioral studies show that models should not be confused with images: Materials that elicit vivid imagery rather than a spatial representation impede reasoning (Knauff & Johnson-Laird, 2002), presumably because the images that they elicit are irrelevant to inference (Knauff et al., 2003).

Further experiments have shown that the logical consequences of relations appear to emerge from mental models rather than to depend on meaning postulates that represent them explicitly. Hence, adult reasoners tend to succumb to pseudotransitive fallacies, such as

Ann is a blood relative of Beth.

Beth is a blood relative of Chris.

Therefore, Ann is a blood relative of Chris.

In accordance with the principle of parsimony, they construct a simple model of a typical case, such as siblings or linear descendants, which yields the erroneous transitive conclusion. Yet, a context that invites them to consider less typical models—relationships that arise from marriage—reliably inhibits them from making the fallacy (Goodwin & Johnson-Laird, 2004a). If logical properties of relations were represented in axiomatic meaning postulates, then they should apply regardless of context, because they capture fundamental and universally true logical properties. Hence, theories that rely on such postulates face some difficulty in accounting for our results.

Various theorists, notably Evans and his colleagues, have posited two distinct reasoning systems (see, e.g., Evans, 2003; Sloman, 1996; Stanovich, 1999). System 1 is rapid, automatic, and responsible for the influence of beliefs on reasoning. System 2 is slow, voluntary, and responsible for deductive reasoning. Schroyens, Schaeken, and Handley (2003) have argued that the model theory is also a dual-process theory. System 1 is responsible for constructing an initial model of the premises on the basis of language and beliefs; System 2 is responsible for searching for alternative models. The principle of parsimony concerns the initial construction of models. Hence, the phenomena to which the principle gives rise, such as pseudotransitivity, reflect the operation of System 1 rather than System 2.

When individuals think aloud as they work their way through a set of relational inferences, their protocols show that they develop different strategies to cope with the problems, and that they switch from one strategy to another (Goodwin & Johnson-Laird, 2004b). The nature of the problems, however, can bias them reliably in favor of one sort of strategy as opposed to another. These results corroborate the principle of strategic assembly. That is, individuals develop their strategies not by laying out some grand abstract design for reasoning, but rather by trying out various inferential tactics on problems, and in this way they discover a strategy for coping with them.

The main components of the principle of complexity have been corroborated empirically in the studies reported in the An Assessment of the Theory section above. Experiment 1 showed that the number of arguments that reasoners need to take into account affects their reasoning. They accordingly find it easier to reason from a relation holding over three different entities than from a relation holding over four different entities. Further experiments showed that the depth of a relation, even when it holds over the same number of entities, also affects reasoning. Thus, Experiment 2 established that it is easier to reason from two binary relations (a first-order problem) than from a relation between two binary relations (a second-order problem). Likewise, Experiment 3 established that it is easier to reason from a relation between relations (a second-order problem) than from a relation between relations between relations (a third-order problem). We have established that increasing depth affects both accuracy and solution times. As depth increases, we would also expect increasing disruptions on a concurrent task (see, e.g., Halford, 1993; Hunt & Lansman, 1982; Lansman & Hunt, 1982; Maybery et al., 1986) and increasing activation of the prefrontal cortex (see, e.g., Waltz et al., 1999). In a review of the present article, Halford has suggested that depth can be translated into additional arguments. Suppose, he suggests, an individual has to take into account all the information in a second-order premise such as

A is taller than B to a greater extent than C is taller than B.

In this case, the difference between A and B, and the difference between C and B, can themselves count as additional arguments. The idea is ingenious, but it seems to be merely a notational variant on depth. Hence, we leave as an open question whether there are empirical differences between depth as a higher order relation and depth as the introduction of new arguments.

Which principles of the theory are robust and unlikely to be modified as a result of future research? In our view, iconicity, emergent properties, and parsimony are secure, and the evidence for them seems compelling. Likewise, there seems little doubt that individuals do spontaneously develop different strategies for reasoning. What the principle does not specify, however, is *how* tactical experience with inferential problems leads to the development of strategies. We have yet to formulate an algorithm that can develop strategies from elementary tactical steps in reasoning about relations (cf. Dierckx, Vandierendonck, & Pandelaere, 2003). For this problem, the theory is far from comprehensive. It is even conceivable that some gifted individuals develop strategies not by working "bottom up" from the results of their tactical explorations, but by devising high-level principles. Analogous problems arise for the principle governing the complexity of integration. Halford's studies (see, e.g., Halford et al., 1998b) and our own show that the number of arguments in a relation affects the difficulty of an inference. The depth of a relation appears to exert additional effects (see our Experiments 2 and 3). However, other variables affect the complexity of integrating the information in premises. When using a premise such as *in front of*, it is harder to add the second referent to a model representing the first referent than vice versa (Oberauer & Wilhelm, 2000). The difficulty probably derives from the representation of the meaning of the relation, which scans from the second referent to the first. Another source of difficulty is the need to use the same quantified premise more than once in the *same* inference (Cherubini & Johnson-Laird, 2004). It does not seem possible to explain this phenomenon either in terms of number of arguments or depth. If so, then the principle of the complexity of integration also fails to be comprehensive.

For many years, psychologists have tried to explain reasoning in terms of formal rules of inference akin to those of a logical calculus. This approach attempts to deal with relations by positing meaning postulates to capture their logical properties and a single deterministic strategy (see, e.g., Rips, 1994). Our results suggest that such an account does not correspond to the way in which logically naive individuals make relational inferences. It fails to predict pseudotransitive inferences, the difficulty of inferences based on multiple models, or the development of different strategies for relational inference. It would also be a gargantuan task to formulate the set of meaning postulates needed to deal with quite straightforward relations. As an example of this problem, consider the effects of negation on a relation. The negation of a transitive but asymmetric relation remains transitive. The following inference is valid, for instance:

Ann is not as tall as Beth.

Beth is not as tall as Cath.

Therefore, Ann is not as tall as Cath.

Yet, the negation of a transitive but symmetric relation is nontransitive. Hence, no definite conclusion about the spatial relation between Ann and Cath follows validly from the following premises:

Ann is not in the same place as Beth.

Beth is not in the same place as Cath.

In a system based on meaning postulates, it would be necessary to specify the consequences of negation in separate meaning postulates. But, according to the model theory, individuals merely construct a model of the premises based on the meaning of the premises. For the preceding premises, they can construct a model in which Ann and Cath are in the same place, and a model in which they are not in the same place. Nevertheless, mental models are not necessarily incompatible with formal rules. As certain outstanding reasoners develop, they may learn to construct formal rules for themselves in certain domains. But, when we as psychologists theorize about relations, we find it quite difficult to work out the consequences of negation. The difficulty would be odd if our reasoning were based on a system of meaning postulates, because we should by now have acquired a complete and correct system of postulates.

The moral is plain. Your mental lexicon represents the meaning of relational terms, but not their logical properties. What you compute when you reason about relations is a novel relation that is not explicitly stated in the premises, but that follows from them. You do so using a system that constructs iconic models of the situation under description from independent representations of the meanings of the premises. This process may yield an emergent conclusion about, say, a transitive relation. Unless you have succumbed to a fallacy based on an oversimplified model, the inference will be easier than one that requires you to consider a set of alternative possibilities. As you tackle more problems of the same sort (especially in the psychological laboratory), you develop a strategy for dealing with them. This strategy reflects the particular nature of the problems. Yet, regardless of your strategy, the difficulty of integrating the information in the premises—a matter in part of the number of arguments and their depth—continues to exert its effects on your ability to reason correctly.

References

Allen, J. (1983). Maintaining knowledge about temporal intervals. *Communications of the Association for Computing Machinery, 26,* 832–843.

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review, 85,* 249–277.

Andrews, G., & Halford, G. S. (1994). Relational complexity and sentence processing. *Australian Journal of Psychology, 46*(Suppl. 1), 1.

Andrews, G., & Halford, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology, 45,* 153–219.

Baddeley, A. D. (1986). *Working memory.* Oxford, England: Clarendon Press.

Bara, B., Bucciarelli, M., & Johnson-Laird, P. N. (1995). The development of syllogistic reasoning. *American Journal of Psychology, 108,* 157–193.

Barclay, J. R. (1973). The role of comprehension in remembering sentences. *Cognitive Psychology, 4,* 229–254.

Bar-Hillel, Y. (1967). Dictionaries and meaning rules. *Foundations of Language, 3,* 409–414.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–660.

Barwise, J. (1993). Everyday reasoning and logical inference. *Behavioral and Brain Sciences, 16,* 337–338.

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

Bell, V., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science, 22,* 25–51.

Berendt, B. (1996). Explaining preferred mental models in Allen inferences with a metrical model of imagery. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 489–494). Mahwah, NJ: Erlbaum.

Birney, D. P., & Halford, G. S. (2002). Cognitive complexity of suppositional reasoning: An application of the relational complexity metric to the knight-knave task. *Thinking & Reasoning, 8,* 109–134.

Braine, M. D. S., & O'Brien, D. P. (Eds.). (1998). *Mental logic*. Mahwah, NJ: Erlbaum.

Brainerd, C. J., & Kingma, J. (1984). Do children have to remember to reason: A fuzzy-trace theory of transitivity development. *Developmental Review, 4,* 311–377.

Breslow, L. (1981). Reevaluation of the literature on the development of transitive inferences. *Psychological Bulletin, 89,* 325–351.

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science, 23,* 247–303.

Bucciarelli, M., & Johnson-Laird, P. N. (in press). Naïve deontics: A theory of meaning, representation, and reasoning. *Cognitive Psychology*.

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

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

Carreiras, M., & Santamaría, C. (1997). Reasoning about relations: Spatial and non-spatial problems. *Thinking & Reasoning, 3,* 191–208.

Cech, C. G., & Shoben, E. J. (1980). Componential reasoning in kinship. *Journal of Experimental Psychology: General, 109,* 393–421.

Cedrus Corporation. (1999). SuperLab Pro (Version 2.0) [Computer software]. San Pedro, CA: Author.

Cherubini, P., & Johnson-Laird, P. N. (2004). Does everyone love everyone? The psychology of iterative inferences. *Thinking & Reasoning, 10,* 31–53.

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

Clement, C. A., & Falmagne, R. J. (1986). Logical reasoning, world knowledge, and mental imagery: Interconnections in cognitive processes. *Memory & Cognition, 14,* 299–307.

Davidson, D. (1967). The logical form of action sentences. In N. Rescher (Ed.), *The logic of decision and action* (pp. 81–95). Pittsburgh, PA: The University Press.

De Soto, C. B., London, M., & Handel, S. (1965). Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology, 2,* 513–521.

Dierckx, V., Vandierendonck, A., & Pandelaere, M. (2003). Is model construction open to strategic decisions? An exploration of the field of linear reasoning. *Thinking & Reasoning, 9,* 97–131.

Dipert, R. (1996). Reflections on iconicity, representation, and resemblance: Peirce's theory of signs, Goodman on resemblance, and modern philosophies of language and mind. *Synthese, 106,* 373–397.

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

Ehrlich, K., & Johnson-Laird, P. N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behavior, 21,* 296–306.

English, L. D. (1998). Children's reasoning in solving relational problems of deduction. *Thinking & Reasoning, 4,* 249–281.

Evans, J. St. B. T. (1980). Thinking: Experimental and information processing approaches. In G. Claxton (Ed.), *Cognitive psychology: New directions* (pp. 275–299). London: Routledge.

Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7,* 454–459.

Evans, J. St. B. T., Handley, S. J., Harper, C. N. J., & Johnson-Laird, P. N. (1999). Reasoning about necessity and possibility: A test of the mental model theory of deduction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1495–1513.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction.* Hillsdale, NJ: Erlbaum.

Evans, J. St. B. T., Over, D. E., & Handley, S. H. (in press). Rethinking the model theory of conditionals. In W. Schaeken, A. Vandierendonck, W. Schroyens, & G. d'Ydewalle (Eds.), *The mental models theory of reasoning: Refinements and extensions*. Mahwah, NJ: Erlbaum.

Foos, P. W., Smith, K. H., Sabol, M. A., & Mynatt, B. T. (1976). Constructive processes in simple linear order problems. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 759–766.

Garnham, A., & Oakhill, J. V. (1996). The mental models theory of language comprehension. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 313–339). Hillsdale, NJ: Erlbaum.

Goldvarg, Y., & Johnson-Laird, P. N. (2001). Naïve causality: A mental model theory of causal meaning and reasoning. *Cognitive Science, 25,* 565–610.

Goodman, N. (1976). *Languages of art* (2nd ed.). Indianapolis, IN: Hackett.

Goodwin, G. P., & Johnson-Laird, P. N. (2004a). *Fallacies in reasoning about relations.* Manuscript submitted for publication.

Goodwin, G. P., & Johnson-Laird, P. N. (2004b). *Reasoning about relations between relations.* Manuscript submitted for publication.

Hagert, G. (1984). Modeling mental models: Experiments in cognitive modeling of spatial reasoning. In T. O'Shea (Ed.), *Advances in artificial intelligence* (pp. 389–398). Amsterdam: North-Holland.

Halford, G. S. (1984). Can young children integrate premises in transitivity and serial order tasks? *Cognitive Psychology, 16,* 65–93.

Halford, G. S. (1993). *Children's understanding: The development of mental models.* Hillsdale, NJ: Erlbaum.

Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can humans process? *Psychological Science, 16,* 70–76.

Halford, G. S., Smith, S. B., Dickson, J. C., Maybery, M. T., Kelly, M. E., Bain, J. D., & Stewart, J. E. M. (1995). Modeling the development of reasoning strategies: The roles of analogy, knowledge, and capacity. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 77–156). Hillsdale, NJ: Erlbaum.

Halford, G. S., Wilson, W. H., Guo, J., Gayler, R. W., Wiles, J., & Stewart, J. E. M. (1994). Connectionist implications for processing capacity limitations in analogies. In K. J. Holyoak & J. Barnden (Eds.), *Advances in connectionist and neural computation theory, Vol. 2: Analogical connections* (pp. 363–415). Norwood, NJ: Ablex.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998a). Authors' response: Relational complexity metric is effective when assessments are based on actual cognitive processes. *Behavioral and Brain Sciences, 21,* 848–864.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998b). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21,* 803–831.

Haviland, S., & Clark, E. (1974). This man's father is my father's son: A study of the acquisition of kin terms. *Journal of Child Language, 1,* 23–47.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104,* 427–466.

Hummel, J. E., & Holyoak, K. J. (2001). A process model of human transitive inference. In M. Gattis (Ed.), *Spatial schemas and abstract thought* (pp. 279–305). Cambridge, MA: MIT Press.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110,* 220–264.

Hunt, E., & Lansman, M. (1982). Individual differences in attention. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 207–254). Hillsdale, NJ: Erlbaum.

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

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

Jahn, G., Johnson-Laird, P. N., & Knauff, M. (2004). Reasoning about consistency with spatial mental models: Hidden and obvious indeterminacy in spatial descriptions. In C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, & T. Barkowsky (Eds.), *Spatial cognition IV—Reasoning, action, interaction* (pp. 165–180). Berlin, Germany: Springer.

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

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

Johnson-Laird, P. N. (1972). The three-term series problem. *Cognition, 1,* 57–82.

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

Johnson-Laird, P. N. (2002). Peirce, logic diagrams, and the elementary operations of reasoning. *Thinking & Reasoning, 8,* 69–95.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* Hillsdale, NJ: Erlbaum.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109,* 646–678.

Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review, 111,* 640–661.

Johnson-Laird, P. N., & Hasson, U. (2003). Counterexamples in sentential reasoning. *Memory & Cognition, 31,* 1105–1113.

Johnson-Laird, P. N., Legrenzi, P., Girotto, P., & Legrenzi, M. S. (2000, April 21). Illusions in reasoning about consistency. *Science, 288,* 531–532.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J.-P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review, 106,* 62–88.

Klauer, K. C., Stegmaier, R., & Meiser, T. (1997). Working memory involvement in propositional and spatial reasoning. *Thinking & Reasoning, 3,* 9–47.

Knauff, M., Fangmeier, T., Ruff, C. C., & Johnson-Laird, P. N. (2003). Reasoning, models, and images: Behavioral measures and cortical activity. *Journal of Cognitive Neuroscience, 4,* 559–573.

Knauff, M., & Johnson-Laird, P. N. (2002). Imagery can impede inference. *Memory & Cognition, 30,* 363–371.

Knauff, M., Jola, C., & Strube, G. (2001). Spatial reasoning: No need for visual information. In D. R. Montello (Ed.), *Spatial information theory: Foundations of geographic information science. Proceedings of the Fifth International Conference on Spatial Information Theory* (pp. 447–457). New York: Springer.

Knauff, M., Rauh, R., & Schlieder, C. (1995). Preferred mental models in qualitative spatial reasoning: A cognitive assessment of Allen's calculus. In *Proceedings of the Fifth Annual Conference of the Cognitive Science Society* (pp. 200–205). Mahwah, NJ: Erlbaum.

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

Kroger, J. K., Holyoak, K. J., & Hummel, J. E. (2004). Varieties of sameness: The impact of relational complexity on perceptual comparisons. *Cognitive Science, 28,* 335–358.

Lansman, M., & Hunt, E. B. (1982). Individual differences in secondary task performance. *Memory & Cognition, 10,* 10–24.

Leeuwenberg, E. L. L. (1969). Quantitative specification of information in sequential patterns. *Psychological Review, 76,* 216–220.

Markman, A. B., & Dietrich, E. (2000). Extending the classical view of representation. *Trends in Cognitive Science, 4,* 470–475.

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

Maybery, M. T., Bain, J. D., & Halford, G. S. (1986). Information processing demands of transitive inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 600–613.

McTaggart, J. M. E. (1927). *The nature of existence.* Cambridge, England: Cambridge University Press.

Metzler, J., & Shepard, R. N. (1982). Transformational studies of the internal representations of three-dimensional objects. In R. N. Shepard & L. A. Cooper (Eds.), *Mental images and their transformations* (pp. 25–71). Cambridge, MA: MIT Press. (Reprinted from *Theories in cognitive psychology: The Loyola Symposium,* by R. L. Solso, Ed., 1974, Hillsdale, NJ: Erlbaum)

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

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

Montague, R. (1974). *Formal philosophy: Selected papers.* New Haven: Yale University Press.

Newstead, S. E., Pollard, P., & Griggs, R. A. (1986). Response bias in relational reasoning. *Bulletin of the Psychonomic Society, 24,* 95–98.

Oaksford, M. (2002). Contrast classes and matching bias as explanations of the effects of negation on conditional reasoning. *Thinking & Reasoning, 8,* 135–151.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101,* 608–631.

Oberauer, K., & Wilhelm, O. (2000). Effects of directionality in deductive reasoning: I. The comprehension of single relational premises. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1702–1712.

Oden, G. C., & Lopes, L. L. (1980). Kin search: Answering questions about relations among relatives. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 479–491.

Ohlsson, S. (1984). Induced strategy shifts in spatial reasoning. *Acta Psychologica, 57,* 46–67.

Ormerod, T. C., & Richardson, J. (2003). On the generation and evaluation of inferences from single premises. *Memory & Cognition, 31,* 467–478.

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

Peirce, C. S. (1931–1958). *Collected papers of Charles Sanders Peirce.* (8 Vols.; C. Hartshorne, P. Weiss, & A. Burks, Eds.). Cambridge, MA: Harvard University Press.

Piaget, J. (1928). *Judgment and reasoning in the child.* London: Routledge.

Potts, G. R. (1978). The role of inference in memory for real and artificial information. In R. Revlin & R. E. Mayer (Eds.), *Human reasoning* (pp. 139–161). New York: Wiley.

Potts, G. R., & Scholtz, K. W. (1975). The internal representation of a three-term series problem. *Journal of Verbal Learning and Verbal Behavior, 14,* 439–452.

Proctor, P. (Ed.). (1978). *The Longman dictionary of contemporary English* (LCODE). London: Longman.

Psychology Software Tools. (2002). E-Prime (Version 1.1) [Computer software]. Pittsburgh, PA: Author.

Pylyshyn, Z. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review, 88,* 16–45.

Pylyshyn, Z. (2002). Mental imagery: In search of a theory. *Behavioural and Brain Sciences, 25,* 157–238.

Quinton, G., & Fellows, B. J. (1975). "Perceptual" strategies in the solving of three-term series problems. *British Journal of Psychology, 66,* 69–78.

Richardson, J. T. E. (1987). The role of mental imagery in models of transitive inference. *British Journal of Psychology, 78,* 189–203.

Riley, C. A. (1976). The representation of comparative relations and the transitive inference task. *Journal of Experimental Child Psychology, 22,* 1–22.

Riley, C. A., & Trabasso, T. (1974). Comparatives, logical structures and encoding in a transitive inference task. *Journal of Experimental Child Psychology, 17,* 187–203.

Rips, L. J. (1994). *The psychology of proof.* Cambridge, MA: MIT Press.

Roberts, M. J. (2000). Strategies in relational inference. *Thinking & Reasoning, 6,* 1–26.

Schaeken, W. S., Girotto, V., & Johnson-Laird, P. N. (1998). The effect of an irrelevant premise on temporal and spatial reasoning. *Kognitionswisschenschaft, 7,* 27–32.

Schaeken, W. S., & Johnson-Laird, P. N. (2000). Strategies in temporal reasoning. *Thinking & Reasoning, 6,* 193–219.

Schaeken, W. S., Johnson-Laird, P. N., & d'Ydewalle, G. (1996). Mental models and temporal reasoning. *Cognition, 60,* 205–234.

Schroyens, W., Schaeken, W., & d'Ydewalle, G. (2001). The processing of negations in conditional reasoning: A meta-analytic case study in mental model and/or mental logic theory. *Thinking & Reasoning, 7,* 121–172.

Schroyens, W., Schaeken, W., & Handley, S. J. (2003). In search of counter-examples: Deductive rationality in human reasoning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56*(A), 1129–1145.

Shaver, P., Pierson, L., & Lang, S. (1975). Converging evidence for the functional significance of imagery in problem solving. *Cognition, 3,* 359–375.

Shepard, R. N., & Metzler, J. (1971, February 19). Mental rotation of three-dimensional objects. *Science, 171,* 701–703.

Shin, S.-J. (1994). Peirce and the logical status of diagrams. *History and Philosophy of Logic, 15,* 45–68.

Shin, S.-J. (2002). *The iconic logic of Peirce's graphs.* Cambridge, MA: MIT Press.

Sloman, S. A. (1996). The empirical case for two systems of reasoning, *Psychological Bulletin, 119,* 3–22.

Sloutsky, V. M., & Goldvarg, Y. (1999). Effects of externalization on representation of indeterminate problems. In M. Hahn & S. Stones (Eds.), *Proceeding of the Twenty First Annual Conference of the Cognitive Science Society* (pp. 695–700). Mahwah, NJ: Erlbaum.

Sowa, J. F. (1984). *Conceptual structures: Information processing in mind and machine.* Reading, MA: Addison Wesley.

Springer, K. (1992). Children's awareness of the biological implications of kinship. *Child Development, 63,* 950–959.

Springer, K. (1995). Acquiring a naïve theory of kinship through inference. *Child Development, 66,* 547–558.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning.* Mahwah, NJ: Erlbaum.

Steedman, M. J. (1982). Reference to past time. In R. J. Jarvella & W. Klein (Eds.), *Speech, place, and action* (pp. 125–157). London: Wiley.

Stenning, K. (1996). The cognitive impact of diagrams. In A. Clark, J. Ezquerro, & J. M. Larrazabal (Eds.), *Philosophy and cognitive science: Categories, consciousness, and reasoning* (pp. 181–196). Dordrecht, the Netherlands: Kluwer.

Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science, 19,* 97–140.

Sternberg, R. J. (1980). Representation and process in linear syllogistic reasoning. *Journal of Experimental Psychology: General, 109,* 119–159.

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

Trabasso, T., Riley, C. A., & Wilson, E. G. (1975). The representation of linear order and spatial strategies in reasoning: A developmental study. In R. Falmagne (Ed.), *Reasoning: Representation and process* (pp. 201–229). Hillsdale, NJ: Erlbaum.

Van der Henst, J.-B. (2002). Mental model theory versus the inference rule approach in relational reasoning. *Thinking & Reasoning, 8,* 193–203.

Van der Henst, J.-B., Sperber, D., & Politzer, G. (2002). When is a conclusion worth deriving? A relevance-based analysis of indeterminate relational problems. *Thinking & Reasoning, 8,* 1–20.

Van der Henst, J.-B., Yang, Y., & Johnson-Laird, P. N. (2002). Strategies in sentential reasoning. *Cognitive Science, 26,* 425–468.

Vandierendonck, A., & De Vooght, G. (1996). Evidence for mental-model-based reasoning: A comparison of reasoning with time and space. *Thinking & Reasoning, 2,* 249–272.

Vandierendonck, A., De Vooght, G., Desimpelaere, C., & Dierckx, V. (1999). Model construction and elaboration in spatial linear syllogisms. In W. S. Schaeken, G. De Vooght, A. Vandierendonck, & G. d'Ydewalle (Eds.), *Deductive reasoning and strategies* (pp. 191–207). Mahwah, NJ: Erlbaum.

Verweij, A. C., Sijtsma, K., & Koops, W. (1999). An ordinal scale for transitive reasoning by means of a deductive strategy. *International Journal of Behavioral Development, 23,* 241–264.

Walsh, C., & Johnson-Laird, P. N. (2004). Co-reference and reasoning. *Memory & Cognition, 32,* 96–106.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10,* 119–125.

Winkelmann, J. H. (1980). Semantic representation of kinship systems. *Journal of Psycholinguistic Research, 4,* 133–140.

Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). New York: Macmillan.

Wood, D. J. (1969, July 5). Approach to the study of human reasoning. *Nature, 223,* 102–103.

Wood, D. J., & Shotter, J. D. (1973). A preliminary study of distinctive features in problem solving. *Quarterly Journal of Experimental Psychology, 25,* 504–510.

Wood, D. J., Shotter, J. D., & Godden, D. (1974). An investigation of the relationships between problem solving strategies, representation and memory. *Quarterly Journal of Experimental Psychology, 26,* 252–257.

Woodworth, R. S. (1938). *Experimental psychology.* New York: Holt.

Received February 4, 2004
Revision received August 4, 2004
Accepted September 1, 2004 ■